

MAURICE ALLAIS

Fréquence, probabilité et hasard. Appendice II. Le théorème (T). La simulation du hasard par des fonctions presque périodiques

Journal de la société statistique de Paris, tome 124, n° 3 (1983), p. 144-221

http://www.numdam.org/item?id=JSFS_1983__124_3_144_0

© Société de statistique de Paris, 1983, tous droits réservés.

L'accès aux archives de la revue « Journal de la société statistique de Paris » (<http://publications-sfds.math.cnrs.fr/index.php/J-SFdS>) implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques
<http://www.numdam.org/>

COMMUNICATION

FRÉQUENCE, PROBABILITÉ ET HASARD*

APPENDICE II

LE THÉORÈME (T)

LA SIMULATION DU HASARD PAR DES FONCTIONS PRESQUE PÉRIODIQUES**

Maurice ALLAIS

Directeur de Recherche au C.N.R.S.

Professeur à l'École Nationale Supérieure des Mines de Paris

Des séries temporelles que l'on considère généralement comme résultant de processus purement aléatoires en tant qu'obéissant à la loi normale peuvent être interprétées comme représentant des phénomènes totalement déterministes.

Un nouveau théorème, le Théorème (T), est présenté suivant lequel sous des conditions très générales la distribution discrète des valeurs $X(n)$ ($1 \leq n \leq N$) d'une somme de l sinusoides $x_i(n)$ de périodes incommensurables T_i considérées à des instants n régulièrement espacés tend asymptotiquement vers une distribution normale lorsque le nombre N des valeurs considérées et le nombre l des composantes sinusoidales augmentent indéfiniment.

Les calculs numériques illustrant le Théorème (T) montrent que la convergence vers la loi normale est généralement très rapide.

En outre il est également possible de démontrer que, sous quelques conditions additionnelles très générales, la série des valeurs $X(n)$, considérée dans un intervalle donné ($1 \leq n \leq N$), peut avoir, au regard de tous les tests habituellement considérés dans la littérature, toutes les apparences d'une suite de termes aléatoires indépendants.

Le Théorème (T) n'est qu'une illustration d'une propriété très générale; l'applicabilité de tous les théorèmes relatifs aux « variables aléatoires » à des fonctions déterministes de variables réelles, lorsque ces fonctions ont les mêmes distributions que ces variables aléatoires et qu'elles satisfont aux mêmes conditions d'indépendance.

Ce théorème peut être généralisé au cas continu, au cas où les composantes $x_i(n)$ de $X(n)$ sont des fonctions périodiques, et aux cas de fonctions presque périodiques au sens de H. Bohr et de fonctions du temps indépendantes.

Il reste approximativement valable lorsque les périodes T_i étant commensurables, leur plus petit commun multiple est très grand par rapport au nombre N d'observations, l'unité de temps étant l'intervalle séparant deux valeurs successives.

Trois séries S , T et U sont présentées en illustration du Théorème (T) et une étude détaillée en est effectuée. On vérifie ainsi que l'analyse des sommes de composantes sinusoidales peut illustrer toutes les

(*) Communication faite le 16 mars 1983 devant les Sociétés de Statistique de Paris et de France.

(**) Cet Appendice II est la suite du Mémoire : « *Fréquence, Probabilité et Hasard* », et de son Appendice I, *Fréquences empiriques et Fréquences mathématiques* », publiés dans le *Journal de la Société de Statistique de Paris*. Série 124, n° 2, 2^e trimestre 1983, p. 70-102.

circonstances que l'on observe généralement dans la nature, qu'il s'agisse de phénomènes physiques, biologiques ou économiques, depuis le cas de séries très fortement autocorrélées jusqu'à celui, relativement exceptionnel, de séries dénuées de toute autocorrélation, avec toujours ce caractère général qu'elles se distribuent suivant la loi normale.

Il est ainsi établi que la structure déterministe vibratoire de l'univers peut engendrer des effets d'apparence aléatoire, et que le déterminisme peut imiter le hasard. Par là le Théorème (T) ouvre des perspectives nouvelles quant à l'interprétation des théories physiques, des séries temporelles de la géophysique, et des séries temporelles de l'économie.

THE THEOREM (T) AND THE SIMULATION OF CHANCE BY ALMOST PERIODIC FUNCTIONS

Time series generally considered as resulting from purely random processes as obeying the normal law can be interpreted as representing entirely deterministic processes.

A new theorem, Theorem (T), is presented according to which under very general conditions the distribution of the values $X(n)$ ($1 \leq n \leq N$) of the sum of l sinusoidal components $x_j(n)$ of incommensurable periods T_j considered at regularly spaced times n is asymptotically normal when the number N of the considered values and the number l of sinusoidal components increase indefinitely.

Numerical calculations illustrating Theorem (T) show that the convergence towards the normal law is generally very rapid.

It can also be shown that under some very general additional conditions and in a given interval ($1 \leq n \leq N$) the series of $X(n)$ may present all the appearances of a sequence of independent terms according to the different tests of the literature.

Theorem (T) is only an illustration of a very general property: the applicability of all theorems on "random variables" to deterministic functions of real variables, when these functions have the same distributions that these random variables and satisfy the same conditions of independence.

This theorem can be generalized to the continuous case, to the case where the components $x_j(n)$ of $X(n)$ are periodic functions, and to the cases of almost periodic functions in the sense of H. Bohr and of independent functions.

It remains approximately valid when, some periods T_j being commensurable, their least common multiple is very great relatively to the number N of observations, the time unit being equal to the interval between two successive values.

Three series S , T and U are presented as illustrations of Theorem (T) with a detailed study. Thus it can be shown that the analysis of the sums of sinusoidal components enables all possible circumstances to be illustrated, which are generally observed in nature, relating to physical, biological and economic phenomena, from the case of highly autocorrelated series to the case, quite exceptional, of series devoid of any auto-correlation, but all having the general characteristic of being normally distributed.

It is thus established that the deterministic vibrational structure of the universe can produce apparently random effects and that determinism can imitate chance. Thus Theorem (T) opens up new prospects for the interpretation of the theories of physics, of time series in geophysics, and of time series in economics.

SOMMAIRE

APPENDICE II — LE THÉORÈME (T) — LA SIMULATION DU HASARD PAR DES FONCTIONS PRESQUE PÉRIODIQUES

P — La simulation du hasard par le déterminisme

P.1. L'origine de cette étude; P.2. Les fluctuations conjoncturelles de l'économie; P.3. *Le Théorème (T)*; P.4. La possibilité d'une explication déterministe des apparences aléatoires des séries temporelles observées dans la nature; P.5. Trois illustrations; P.6. Méthodes d'analyse des séries — Critères de normalité, d'indépendance, d'arégularité locale, et des valeurs cumulées; P.7. Caractères généraux des fonctions presque périodiques considérées; P.8. Simulation de variables aléatoires indépendantes par des fonctions presque périodiques et convergence vers la moyenne; P.9. Régularités locales des séries temporelles observées et fonctions presque périodiques; P.10. Structure des séries temporelles stationnaires; P.11. Propriétés générales des fonctions presque périodiques; P.12. Illustration des propriétés des séries temporelles observées par les propriétés de fonctions presque périodiques; P.13. L'analyse des séries temporelles et l'algorithme des valeurs cumulées.

Q — Le Théorème (T)

Q.1. La formulation classique du Théorème central limite; Q.2. Reformulation fréquentielle du Théorème central limite; Q.3. *Théorème (T)*; Q.4. Lemmes; Q.5. Démonstration du Théorème (T); Q.6. Conditions d'application du Théorème central limite; Q.7. Processus aléatoire (P^*) correspondant au processus (P) des $X(n)$; Q.8. *Le Théorème (T)* et les fonctions arbitraires de Poincaré; Q.9. Le coefficient de Liapounoff; Q.10. Distribution des coefficients d'autocorrélation asymptotiques des $X(n)$; Q.11. Coefficients d'autocorrélation asymptotiques des coefficients d'autocorrélation asymptotiques; Q.12. Cas de l'égalité de tous les a_i ; Q.13. La convergence empirique vers la loi normale de la distribution des $X(n)$; Q.14. Généralisations du Théorème (T); Q.15. Cas où certaines fréquences sont commensurables; Q.16. Théorème d'équivalence.

R — Tests de normalité, d'indépendance, d'arégularité locale et des valeurs cumulées

R.1. Test de normalité : Ajustement de Henri suivant la loi normale; R.2. Test d'absence de toute autocorrélation — Corrélogramme; R.3. Test d'arégularité locale pour une série de valeurs indépendantes : Distribution des nombres de suites de valeurs positives et négatives; R.4. Test des valeurs cumulées; R.5. Représentation de Henri de la fonction de fréquence correspondant à une sinusoïde.

S — Série S. Analyse d'une série temporelle issue des données de l'observation et constituée par les 721 valeurs successives de la somme de 13 sinusoïdes dont la distribution est celle de la loi normale

S.1. La série S; S.2. Ajustement de la distribution des valeurs de $X(n)$ suivant la loi normale; S.3. Coefficients d'autocorrélation asymptotiques; S.4. Coefficients d'autocorrélation; S.5. Commentaire général; S.6. Annexes — Tableau S.1. et Graphiques S.1 à S.4.

T — Série T. Analyse d'une série temporelle artificielle constituée par les 1.000 valeurs successives d'une somme de 20 sinusoïdes et simulant une série aléatoire autocorrélée

T.1. La série T; T.2. Ajustement de la distribution des valeurs de $X(n)$ suivant la loi normale; T.3. Coefficients d'autocorrélation asymptotiques; T.4. Coefficients d'autocorrélation; T.5. Valeurs cumulées; T.6. Commentaire général; T.7. Annexes — Tableau T.1 et Graphiques T.1 à T.7.

U — Série U. Analyse d'une série temporelle artificielle constituée par les 1.000 valeurs successives d'une somme de 200 sinusoïdes et simulant une suite aléatoire de termes indépendants.

U.1. La série U; U.2. Ajustement de la distribution des valeurs de $X(n)$ suivant la loi normale; U.3. Coefficients d'autocorrélation asymptotiques; U.4. Coefficients d'autocorrélation; U.5. Valeurs cumulées; U.6. Distribution des suites de valeurs positives et négatives; U.7. Commentaire général; U.8. Annexes — Tableau U.1 et Graphiques U.1 à U.7.

(Les théories indéterministes) ne doivent pas être considérées comme concernant les réalités sous-jacentes, mais les phénomènes tels qu'ils sont vus et compris par nous; en d'autres mots l'indétermination ne réside pas dans la nature objective mais seulement dans notre interprétation subjective de la nature.

James JEANS ⁽¹⁾

Le jour viendra où, par une étude suivie de plusieurs siècles, les choses actuellement cachées paraîtront avec évidence, et la postérité s'étonnera que des vérités si claires nous aient échappé.

SÈNEQUE ⁽²⁾

Quelque belle que puisse être une théorie mathématique, quelque complète que puisse être son élaboration, ses faiblesses ou ses avantages ne peuvent être déterminés que lorsqu'elle a été soumise au test du calcul numérique.

Karl PEARSON ⁽³⁾

Cet *Appendice II* a pour objet de présenter les principaux résultats relatifs au *Théorème (T)* et à la simulation du hasard par des fonctions presque périodiques, et quelques commentaires essentiels. Il vient justifier, au moins partiellement, les propositions du § 10 de mon mémoire, *Fréquence, Probabilité et Hasard* ⁽⁴⁾.

Les idées directrices et les propositions essentielles de cet *Appendice* sont présentées ci-dessous dans la Section (P).

P. LA SIMULATION DU HASARD PAR LE DÉTERMINISME

P.1. *L'origine de cette étude*

De 1953 à 1960 j'ai été amené à analyser les séries d'observations correspondant au déplacement de l'azimut d'oscillation d'un pendule paraconique ⁽¹⁾. Ce mouvement présente des anomalies *tout à fait inexplicables* dans la théorie actuelle de la gravitation, qu'on la considère sous sa forme newtonienne ou sous sa forme einsteinienne.

Comme on constate que les séries temporelles données par l'observation du mouvement du pendule paraconique se distribuent en général suivant la loi normale, l'hypothèse toute naturelle qui s'est présentée tout d'abord à moi a été de considérer les effets observés comme purement fortuits. Les séries étant fortement autocorrélées, cette interprétation revenait à considérer les séries temporelles données par l'observation comme résultant d'un processus aléatoire autorégressif, et elles en avaient effectivement toutes les apparences.

(1) James JEANS, 1954, *Physique et Philosophie*, Chap. VII, p. 226.

(2) SÈNEQUE, *Questions Naturelles, Livre VII, Des Comètes*, § XXV. Traduction de Laplace, *Essai philosophique sur les probabilités*, I, p. 5.

(3) Karl PEARSON, 1921-1933, *The History of Statistics in the 17th and 18th centuries*, p. 456.

(4) *Journal de la Société de Statistique de Paris*, tome 124, n° 2, 2^e trimestre 1983, p. 70-102. Les références à ce Mémoire sont indiquées par les numéros des sections correspondantes.

(P.1) Ces séries correspondent à des expériences que j'ai réalisées dans cette période dans un laboratoire de l'*Institut de Recherche de la Sidérurgie (I.R.S.I.D)* à Saint-Germain-en-Laye.

Voir ALLAIS, 1957 a, b, c, et f; 1958 a et b; 1959 a et b. Sur l'analyse de ces séries, voir ALLAIS, 1957 d et e, 1960, et 1961. Pour une vue d'ensemble, voir ALLAIS, 1958 c et 1959 c.

Ces expériences m'ont valu d'être lauréat en 1959 de la *Gravity Research Foundation* (États-Unis), et de recevoir le *Prix Galabert* 1959 de la *Société Française d'Astronautique*.

Mais une hypothèse alternative était que la distribution suivant la loi normale des valeurs successives d'une série temporelle ne doit pas être nécessairement considérée comme de nature aléatoire.

J'ai été ainsi amené à examiner si la distribution des valeurs successives de la somme de différentes sinusoïdes peut être effectivement normale. En fait l'analyse harmonique par l'*Institut Hydrographique National* de la première série d'observations obtenue en juin-juillet 1954 suivant 13 ondes de la théorie des marées avait déterminé les amplitudes et les phases de l'ajustement correspondant, et la somme de ces 13 ondes avait été calculée. Non sans quelque étonnement j'ai constaté que les valeurs successives de cette somme se distribuaient suivant la loi normale ⁽²⁾. Il était donc établi que *la distribution normale d'une série temporelle ne prouve pas qu'elle dérive nécessairement du hasard*.

P.2. Les fluctuations conjoncturelles de l'économie

Dans ces vingt cinq dernières années j'ai été amené à reprendre cette analyse sous des formes diverses, notamment dans l'étude des séries temporelles relatives aux fluctuations conjoncturelles de l'économie. Ici encore ces séries se présentent sous les apparences des fluctuations aléatoires autocorrélées engendrées par des processus autorégressifs. Tel est le cas notamment des fluctuations des cours de bourse. De cette constatation de nombreux auteurs ont déduit qu'il s'agit là de mouvements purement fortuits ⁽³⁾.

Je suis quant à moi arrivé progressivement à la conviction qu'une telle conclusion ne s'impose pas nécessairement, *les structures observées pouvant parfaitement s'expliquer par l'action conjuguée de composantes périodiques*, ces composantes périodiques pouvant d'ailleurs comporter des amplitudes et des périodes variables avec le temps ⁽⁴⁾⁽⁵⁾.

Au cours de toutes ces recherches, j'ai été particulièrement frappé de *l'identité* des problèmes relatifs à la signification des données de l'expérience en économie et en physique. Rien n'a été plus suggestif pour moi que ce rapprochement de deux sciences apparemment si dissemblables ⁽⁶⁾.

(P.2) Les résultats de cette analyse sont présentés dans la Section S ci dessous.

(P.3) BACHELIER (1938, p. VI) écrit : « *Les résultats que fournit l'examen des cotes sont en parfait accord avec ceux que fournit le calcul... Cette concordance entre la théorie et l'observation... prouve que le marché obéit à la loi du hasard... Un tel marché, soumis constamment à une infinité d'influences variables et qui agissent dans divers sens, doit finalement se comporter comme si aucune cause n'était en jeu et comme si le hasard agissait seul... En général la diversité des causes permet leur élimination, l'incohérence même du marché est sa méthode, et c'est parce qu'il n'obéit à aucune loi qu'il suit fatalement la loi du hasard* ».

Un grand nombre d'auteurs ont repris la thèse de BACHELIER. Voir notamment COOTNER, 1964; et GRANGER et MORGENSTERN, 1970.

(P.4) *Dans l'état actuel de la science, ce n'est là naturellement qu'une hypothèse. Mais c'est là une hypothèse qu'il ne serait pas scientifique de refuser d'examiner sérieusement* (voir ALLAIS, 1980, et ci dessous Section Q).

(P.5) La théorie des cycles économiques, très à la mode dans les années trente, a connu depuis la guerre une longue période de récession en raison des difficultés auxquelles elle s'est heurtée et des échecs qu'elle a subis. Ces difficultés et ces échecs ont résulté de conceptions trop simplistes et à vrai dire erronées. On a supposé l'existence d'un petit nombre de composantes sinusoïdales d'amplitude et de périodes constantes, alors que la structure quasi-périodique des structures que l'on rencontre est beaucoup plus complexe. Sur ces difficultés et ces échecs voir notamment KENDALL, 1946, § 1.22, et 7.1 7.14, p. 7 et 63-65.

En voulant tout expliquer par le hasard on est aujourd'hui passé d'un extrême dans l'autre, et la théorie de l'analyse spectrale se limite presque entièrement aujourd'hui à l'étude des spectres continus générés par des modèles aléatoires. Cette position est tout aussi injustifiable que celle qui l'a précédée, car on ne saurait postuler *a priori* que seul le *hasard* peut expliquer la génération des séries temporelles observées.

(P.6) Le problème majeur que j'ai eu à résoudre dans mes recherches expérimentales sur le pendule paraconique a été de tester la réalité de certaines périodicités dans une série temporelle (sur ce point, voir ALLAIS, 1957 *d* et *e*, 1960 et 1961). Ce problème est en fait identiquement le même que celui que les économistes ont été amenés à étudier dans leurs recherches sur les cycles économiques. Cet exemple a une valeur absolument générale : tous les ouvrages d'économétrie exposent des méthodes d'analyse des séries temporelles qui s'appliquent identiquement en géophysique, et réciproquement les géophysiciens ont été amenés à résoudre des problèmes sur la structure des séries temporelles dont les économistes ne peuvent que tirer le plus grand profit.

Comme je l'ai indiqué, cette similitude de structure morphologique conduit à l'hypothèse d'une similitude des causes (voir *Fréquence, Probabilité et Hasard*, § 10, J.S.S.P., tome 124, n° 2, p. 86-87; et § P.9 ci dessous).

L'objet de cet *Appendice II* est de présenter très sommairement quelques résultats particulièrement significatifs ⁽⁷⁾.

P.3. Le Théorème (T)

Depuis de Moivre et Laplace l'idée a progressé, puis s'est imposée dans la littérature que sous des conditions très générales la somme de variables « aléatoires » indépendantes se distribue suivant la loi normale.

En réalité le *Théorème central limite* sur lequel repose cette théorie est tout à fait indépendant de toute considération de hasard et les variables qu'elle considère ne sont pas réellement des variables aléatoires ⁽⁸⁾. En fait on peut appliquer ce théorème à l'analyse des propriétés de fonctions presque périodiques ⁽⁹⁾.

Sous des conditions très générales on peut effectivement démontrer (*Théorème T*) que des sommes de l sinusoïdes

$$X(n) = \sum_{j=1}^l x_j(n) = \sum_{j=1}^l a_j \cos 2\pi f_j(n - n_j) \quad (1)$$

$$f_j = 1/T_j; \quad a_j > 0; \quad 1 \leq j \leq l; \quad 1 \leq n \leq N$$

dont les fréquences f_j sont irrationnelles et ne sont liées par aucune relation linéaire à coefficients entiers, se distribuent d'autant mieux suivant la loi normale que leur nombre l est plus élevé et que le nombre N des valeurs considérées est plus grand. *Les conditions considérées sont celles de Liapounoff*. Elles expriment notamment que le carré de chacune des amplitudes est relativement petit par rapport à la somme des carrés de toutes les amplitudes ⁽¹⁰⁾.

Les fonctions $X(n)$ définies par la relation (1) où les fréquences f_j sont irrationnelles sont des fonctions presque périodiques.

On constate de plus que, pour une valeur donnée de N et sous des conditions additionnelles également très générales, la suite des $X(n)$ peut présenter toutes les apparences d'une suite de termes indépendants. On obtient ainsi pour des séries d'une longueur donnée une simulation presque parfaite de ce qu'il est convenu d'appeler le hasard ⁽¹¹⁾.

Enfin, comme il sera indiqué plus loin dans le § Q 14, le domaine de validité du *Théorème (T)* peut être considérablement étendu, cette extension lui donnant la plus grande généralité.

P.4. La possibilité d'une explication déterministe des apparences aléatoires des séries temporelles observées dans la nature

Ce que l'on observe généralement dans la nature, qu'il s'agisse de phénomènes physiques, biologiques ou économiques, ce sont des séries temporelles autocorrélées, et on n'observe que très exceptionnellement des séries de termes indépendants. En tout cas le nombre de leurs termes est toujours fini; et il est rarement très grand. Très souvent ces séries, qu'elles soient autocorrélées ou non, présentent toutes

(P.7) En fait cet *Appendice* n'est qu'un résumé d'une étude plus étendue « *Simulation déterministe du hasard par des fonctions presque périodiques* » (ALLAIS, 1982) qui sera intégrée dans mon prochain ouvrage, « *Frequency, Probability and Chance* », 1984.

(P.8) Voir *Fréquence, Probabilité et Hasard*, Section I.C., J.S.S.P., p. 101 102, tome 124, n° 2, et ci dessous § Q.2.

(P.9) Voir ci-dessous § Q.2 et Q.6.

(P.10) Voir Section Q ci dessous.

Il convient de souligner que l'irrationalité des fréquences ne saurait suffire par elle-même pour assurer la normalité de la distribution des $X(n)$. On ne saurait donc dire que la normalité de la distribution des $X(n)$ résulte de l'irrationalité des fréquences et de ce que les distributions des décimales des nombres irrationnels présentent généralement, au moins à première vue, l'apparence de distributions aléatoires.

(P.11) Voir ci dessous Section U.

les apparences de ce qu'il est communément convenu d'appeler le « *hasard* » et elles sont considérées comme résultant d'un processus aléatoire.

En fait *l'analyse des sommes d'un nombre fini de composantes sinusoïdales permet d'illustrer toutes les circonstances que l'on observe*, depuis le cas de séries très fortement autocorrélées jusqu'à celui, très exceptionnel, de séries dénuées de toute autocorrélation, avec toujours ce caractère général qu'elles se distribuent approximativement suivant la loi normale ⁽¹²⁾.

Il est ainsi établi que la structure déterministe vibratoire de l'univers peut engendrer des effets d'apparence aléatoire ⁽¹³⁾, et que le déterminisme peut imiter le hasard.

Ainsi que je l'ai déjà indiqué, la simulation de ce qu'il est convenu d'appeler le « *hasard* » par des fonctions presque périodiques vient appuyer l'hypothèse du « *Facteur X* », suivant laquelle les fluctuations des séries temporelles que nous observons dans les phénomènes qui relèvent des sciences de la nature, des sciences de la vie et des sciences de l'homme, pourraient résulter par des effets de résonance de l'influence des innombrables vibrations qui sillonnent l'espace dans lequel nous vivons et dont l'existence est aujourd'hui une certitude ⁽¹³⁾.

P.5. Trois illustrations

Je présente ci-dessous trois illustrations de la distribution suivant la loi normale de certaines fonctions presque périodiques.

La première (*Série S*) est constituée par 721 valeurs successives de la somme des 13 sinusoïdes que j'avais analysée en 1956. Cette série est marquée par une forte autocorrélation, et bien que son graphique représentatif apparaisse comme manifestement caractérisé par une structure voisine d'une structure périodique, elle se distribue approximativement suivant la loi normale ⁽¹⁴⁾.

La seconde (*Série T*) est constituée par une série artificielle de $N = 1.000$ termes, sommes de 20 sinusoïdes. Cette série est présentée pour montrer que la distribution suivant la loi normale de la série S n'a rien d'exceptionnel bien que le nombre de ses composantes périodiques soit relativement faible ⁽¹⁵⁾.

La troisième (*Série U*) est une série artificielle de 1.000 valeurs, sommes de 200 composantes sinusoïdales. Cette série présente tous les caractères d'une série de termes aléatoires *indépendants* ⁽¹⁶⁾. Elle permet ainsi d'illustrer les conclusions erronées auxquelles conduirait l'application des tests habituels de la littérature quant à la structure de la série considérée *pour quiconque ignorerait sa structure effective* ⁽¹⁷⁾.

(P.12) Voir § P.5.7, et Sections S, T et U ci dessous.

(P.13) J'ai recherché attentivement dans toute la littérature si cette proposition a déjà été exprimée, et je n'ai trouvé en fait qu'un seul auteur, Baudez, 1952 et 1953, qui l'ait déjà présentée (je n'ai découvert le livre de 1952 de Baudez qu'en 1960). *Je dois souligner ici son antériorité.*

Cependant son analyse, qui ne repose sur aucune démonstration réelle, présente malheureusement de très nombreuses imperfections, et des affirmations très contestables et très souvent inexactes. Je me rallie pour une large part aux observations présentées par Maurice Fréchet lors de la communication de 1953 de Baudez devant la Société de Statistique de Paris : « *Il n'est pas impossible que nous soyons en présence d'une idée juste et intéressante. Mais nous nous trouvons devant une ébauche de théorie, exprimée dans un langage si imprécis, dans un ordre si peu cohérent qu'il est actuellement impossible de se prononcer dans un sens ou dans l'autre... Des objections analogues et d'autres encore s'appliquent au texte plus développé de l'ouvrage de M. Baudez* » (Discussion de la Communication du 18 mars 1982 de Baudez, J.S.S.P., juillet août 1953, p. 185).

(P.13') Voir ALLAIS, 1983 *d*, § 10, p. 86-87.

(P.14) Voir ci-dessus § P.1 et ci-dessous Section S. Voir tout particulièrement les graphiques G.4 et G.5.

(P.15) Voir ci-dessus Section T.

(P.16) Voir ci dessous Section U. De telles séries de termes indépendants ne se rencontrent *pratiquement jamais* dans la nature (voir ci dessus § P.4).

(P.17) Voir ci-dessus § U.7.

P.6. Méthodes d'analyse des séries

Critères de normalité, d'indépendance, d'arégularité locale et des valeurs cumulées

L'analyse qui a été effectuée des trois séries S , T , U et de nombreuses autres porte à la fois sur les séries elles-mêmes, sur leurs corrélogrammes et leurs valeurs cumulées. Elle permet d'apprécier à la fois la normalité de la distribution des valeurs des séries étudiées, leur autocorrélation, et éventuellement, pour quiconque ignorerait leur structure effective, leur absence d'autocorrélation et leur arégularité locale ⁽¹⁸⁾.

Les ajustements à la loi normale des distributions statistiques considérées ont été effectués en utilisant la méthode graphique de Henri, et la signification des ajustements a été analysée par référence au test χ^2 ⁽¹⁹⁾. Pour toutes les séries les corrélogrammes ont été calculés ⁽²⁰⁾.

En outre pour la Série U qui a toutes les apparences d'une suite de termes aléatoires indépendants, la distribution des suites de valeurs positives et négatives a été analysée. Cette distribution a été comparée à la distribution du modèle fréquentiel correspondant, avec une application du test χ^2 ⁽²¹⁾.

Enfin le calcul des valeurs cumulées a été effectué pour les séries T et U de manière à déceler, dans l'hypothèse où l'on ignorerait leur structure réelle, soit la présence d'une structure presque périodique sous-jacente, soit la compatibilité avec l'hypothèse d'indépendance des $x_j(n)$ ⁽²²⁾ ⁽²³⁾.

P.7. Caractères généraux des fonctions presque périodiques considérées

Pour illustrer le Théorème (T) un assez grand nombre de calculs ⁽²⁴⁾ ont été effectués et leurs résultats analysés. Les trois séries (S), (T) et (U) ci-dessus indiquées n'en constituent que trois illustrations ⁽²⁵⁾. Dans l'ensemble les conclusions suggérées par cette analyse sont les suivantes.

Tout d'abord, et naturellement, pour des valeurs entières de n , la distribution des N valeurs $X(n)$ de l sinusoïdes de fréquences f_j et de phases n_j se rapproche d'autant mieux d'une loi normale que les

(P.18) Ces différentes méthodes d'analyse font l'objet de la Section R ci dessous.

(P.19) § R.1 ci dessous.

(P.20) § R.2 ci dessous.

Comme le corrélogramme asymptotique des $X(n)$ est également une fonction presque périodique (§ Q.6 ci dessous), on a par là même de nouvelles illustrations de l'application du Théorème T aux fonctions presque périodiques (§ P.3 ci dessus).

(P.21) § R.3 ci dessous.

(P.22) § P.13 et R.4 ci dessous.

(P.23) Aucune analyse harmonique des séries S , T et U n'a été effectuée jusqu'ici avec les méthodes habituelles, puisqu'en fait leur structure presque périodique est parfaitement connue. Mais en fait une telle analyse pourrait cependant être très utile pour illustrer l'application des théories habituellement admises et les erreurs d'interprétation auxquelles elles sont susceptibles de mener pour quiconque ignorerait leur structure effective. Ce travail est en cours et fera l'objet d'une prochaine publication.

(P.24) Une série d'environ 30 calculs m'a permis de vérifier la théorie exposée dans la Section Q ci dessous, quant à l'influence notamment du nombre N de valeurs considérées, du nombre l des sinusoïdes, de leurs amplitudes a_j , et de leurs périodes T_j . Les phases n_j n'ont aucune influence sur la distribution des $X(n)$.

Une trentaine d'autres calculs ont été effectués en considérant des amplitudes et des périodes variables avec le temps. Pour simplifier je n'en fais pas état dans ce qui suit. J'indique simplement que dans ces calculs on prend dans la relation (1) ci dessus

$$T_j = T_j' \left/ \left[1 - \mu_j \cos \frac{2\pi}{\Theta_j} n \right] \right. \quad a_j = a_j' \left(1 + \lambda_j \cos \frac{2\pi}{\theta_j} n \right) \quad (1)$$

où les λ_j , μ_j , Θ_j et θ_j sont des constantes. Avec un tel modèle l'apparence aléatoire des séries calculées est beaucoup plus accusée.

Autant qu'on puisse en juger, l'existence de phénomènes caractérisés par des périodes et des amplitudes variables est très générale dans la nature.

En fait il résulte de la théorie des fonctions de Bessel que dans ce cas chacune des composantes $x_j(n)$ de $X(n)$ est en fait égale à une somme de sinusoïdes (voir note Q.18 ci dessous).

(P.25) Voir ci dessous Sections S, T et U.

Un grand nombre de lecteurs du manuscrit de ce mémoire n'ont attaché qu'une importance toute relative, voire aucune importance, aux Sections R, S, T et U ci dessous, pourtant tout à fait essentielles. C'est là une profonde erreur (voir la citation de K. Pearson présentée ci-dessus en épigraphe).

conditions générales d'application du *Théorème (T)* au cas de sommes de sinusoïdes ⁽²⁶⁾ sont mieux satisfaites :

- un nombre N relativement important de valeurs régulièrement espacées,
- un nombre l relativement important de sinusoïdes,
- l'irrationalité des fréquences f_j et l'absence de toute relation linéaire à coefficients entiers entre ces fréquences,
- des rapports a_j/a relativement peu différents de l'unité, a représentant la moyenne des amplitudes a_j .

Toutefois de bons ajustements suivant la loi normale ont été obtenus pour des valeurs de N de l'ordre de 200, des valeurs de l de l'ordre de 20, et une dispersion relativement grande des a_j . En fait la rapidité avec laquelle la convergence vers la loi normale peut avoir lieu dans certains cas est réellement étonnante.

En fait ces différentes conditions n'impliquent en aucune façon que deux valeurs successives de $X(n)$ soient indépendantes ⁽²⁷⁾.

Une circonstance *tout à fait essentielle* doit être soulignée. Une somme de sinusoïdes ne pouvant jamais excéder en valeur absolue la somme de ses amplitudes, on a nécessairement

$$-\sum a_j \leq X(n) \leq \sum a_j \quad (2)$$

Il résulte de là que si excellente que puisse être la représentation des distributions des valeurs successives $X(n)$ de ces séries par la loi normale, elle ne peut l'être *pour une valeur finie du nombre l de sinusoïdes* que dans un intervalle d'amplitude limitée. *Les valeurs infinies de la distribution sont ainsi totalement exclues* ⁽²⁸⁾ ⁽²⁹⁾. *Mais en fait c'est là le cas de la totalité des séries temporelles effectivement observées dans la nature.*

P.8. *Simulation de variables aléatoires indépendantes par des fonctions presque périodiques et convergence vers la moyenne*

Dans le cas de processus « aléatoires » on sait que la fréquence observée ne tend vers la « probabilité » (qu'en fait il convient d'appeler « fréquence intrinsèque ») qu'en « probabilité », c'est-à-dire en fréquence, ce qui implique que la convergence n'a pas lieu au sens de l'analyse ⁽³⁰⁾.

De même si l'on considère la moyenne

$$v_n = (u_1 + \dots + u_p + \dots + u_n) / n \quad (3)$$

des valeurs prises par des variables « aléatoires » indépendantes u_p ayant la même distribution normale, de moyenne nulle et de même écart type σ , on sait que v_n ne tend vers zéro comme $1/\sqrt{n}$ qu'en « probabilité », c'est-à-dire en fréquence ⁽³¹⁾.

Il en va tout autrement si on considère les valeurs successives $X(n)$ définies par la relation (1) ci-

(P.26) Voir ci-dessous Section Q, tout particulièrement les § Q.3 et Q.13.

(P.27) La série S en constitue une excellente illustration (voir ci-dessous Section S).

(P.28) Sur ce point voir § Q.13 ci-dessous.

(P.29) Mais, comme dans le cas du théorème central limite (§ Q.I. ci-dessous), rien n'empêche d'envisager pour des valeurs infinies de N et l des valeurs infinies pour $A = \sum a_j$ et pour l'écart type σ_X de $X(n)$.

En tout état de cause comme on a $\sigma_X^2 = \sum a_j^2 / 2$, σ_X est de l'ordre de $\sqrt{l}a$ où a est la moyenne des a_j . Comme $A = \sum a_j = la$ le rapport A/σ est de l'ordre de grandeur de \sqrt{l} . Pour $l = 400$, A/σ_X est ainsi de l'ordre de 20. C'est là un écart qui pour la loi normale ne se réalise pratiquement jamais.

(P.30) Voir *Fréquence, Probabilité et Hasard*, Appendice I, J.S.S.P., tome 124, n° 2, p. 92 102.

(P.31) Sur la représentation graphique de tels processus, voir Wold, 1965, p. 12 18.

dessus d'une fonction presque périodique, et cela même si la série $X(n)$ peut présenter toutes les apparences d'une série résultant d'un processus aléatoire u_n dans un très large domaine de variation de n ⁽³²⁾.

Posons en effet

$$Y(n) = \sum_{p=1}^n X(p) \quad \lambda_j = -2\pi f_j n_j \quad \mu_j = 2\pi f_j \quad (4)$$

On a d'après (1) (§ P.3) ⁽³³⁾

$$\frac{Y(n)}{n} = \frac{1}{n} \sum_{j=1}^l \sum_{p=1}^n x_j(p) = \frac{1}{n} \sum_{j=1}^l a_j \sum_{p=1}^n \cos 2\pi f_j (p - n_j) \quad (5)$$

$$\begin{aligned} \sum_{p=1}^n \cos 2\pi f_j (p - n_j) &= \cos(\lambda_j + \mu_j) + \cos(\lambda_j + 2\mu_j) + \dots + \cos(\lambda_j + n\mu_j) \\ &= \frac{\sin[(n+1)\mu_j/2]}{\sin(\mu_j/2)} \cos\left[\lambda_j + \frac{n\mu_j}{2}\right] - \cos \lambda_j \end{aligned} \quad (6)$$

Comme on a

$$2 \sin[(n+1)(\mu_j/2)] \cos\left[\lambda_j + n(\mu_j/2)\right] = \sin[n\mu_j + \lambda_j + (\mu_j/2)] + \sin[(\mu_j/2) - \lambda_j] \quad (7)$$

que f_j est irrationnel et que par suite $\sin(\mu_j/2)$ ne peut être nul, on voit que pour une valeur donnée de l , la valeur cumulée $Y(n)$ est elle-même une fonction presque périodique, qu'elle reste limitée supérieurement quel que soit n ⁽³⁴⁾ et que la moyenne des $X(1), X(2), \dots, X(n)$ tend vers zéro comme $1/n$ lorsque n augmente indéfiniment ⁽³⁵⁾.

On voit ainsi que la simulation d'une variable aléatoire par une fonction presque périodique déterministe, si excellente qu'elle puisse être, ne pourra se constater que dans un domaine de variation limité de n , si grand que ce domaine puisse être lui-même. Cependant on pourra toujours attribuer au hasard la convergence vers zéro de la fonction $Y(n)/n$ si on l'interprète comme une variable aléatoire et si la série considérée est d'une longueur N limitée.

Il résulte de là que les séries temporelles données par l'observation étant d'une longueur nécessairement limitée, une structure périodique et une structure aléatoire, le déterminisme et le hasard, pourront se révéler comme tout à fait indiscernables.

Au total on voit que si on considère les valeurs successives de fonctions presque périodiques, la fréquence tend vers la fréquence intrinsèque au sens de l'analyse, et cela même si les séries considérées peuvent présenter toutes les apparences d'une série aléatoire dans un domaine de variation de n très large mais fini. La raison en est essentiellement que les valeurs absolues d'une fonction presque périodique sont limitées supérieurement et que dès lors les valeurs très grandes de cette fonction sont impossibles.

(P.32) Comme c'est le cas de la série presque périodique $U(n)$ étudiée dans la Section U ci-dessous.

(P.33) Naturellement, dans le cas continu, le calcul est bien plus simple puisque

$$\int_{n_0}^n \cos 2\pi f_j (n - n_j) dn = \frac{1}{2\pi f_j} \left[\sin 2\pi f_j (n - n_j) - \sin 2\pi f_j (n_0 - n_j) \right]$$

On en déduit notamment que l'intégration de $X(n)$ amplifie l'influence des périodes T_j les plus grandes (voir ci-dessous § P.13).

(P.34) En fait la Section U ci-dessous présente l'exemple d'une somme cumulée $Y(n)$ se comportant dans un très large domaine de variation de n comme la somme cumulée $\sum u_p$ de la relation (3) ci-dessus (voir § U.5 ci-dessous).

(P.35) Dans la mesure où une variable considérée comme aléatoire ne serait en réalité qu'une fonction presque périodique, le postulat de Von Mises (1920-1956) de convergence de la fréquence vers la probabilité au sens de l'analyse, lorsque le nombre d'épreuves augmente indéfiniment, se trouverait entièrement justifié.

P.9. Régularités locales des séries temporelles observées et fonctions presque périodiques

Au cours de ces trente-cinq dernières années, j'ai constaté dans de très nombreux cas que les séries temporelles, notamment dans le domaine de l'économie et dans celui de la géophysique, sont caractérisées par l'existence de régularités, périodicités ou symétries locales, très remarquables⁽³⁶⁾, tout à fait analogues à celles que présentent certaines fonctions presque périodiques.

Malheureusement ces séries empiriques sont toujours trop courtes pour que l'on puisse découvrir leur structure presque périodique sous-jacente, à supposer naturellement qu'une telle structure existe.

En fait les valeurs cumulées des deux séries T et U ci-dessous analysées présentent de nombreuses régularités locales⁽³⁷⁾ qui illustrent cette propriété générale que les fonctions presque périodiques peuvent présenter des propriétés tout à fait analogues à celles qui se constatent très souvent dans les données de l'observation⁽³⁸⁾.

Cette analogie morphologique ne justifie naturellement pas à elle seule l'hypothèse du « Facteur X », suivant laquelle les fluctuations des séries temporelles observées pourraient résulter par des effets de résonance de l'influence des innombrables vibrations qui sillonnent l'espace, mais elle montre que si cette hypothèse était justifiée, elle permettrait de suggérer une explication d'un grand nombre de régularités observées⁽³⁹⁾.

P.10. Le Théorème (T) et la structure des séries temporelles stationnaires

D'après un théorème de Wold⁽⁴⁰⁾, tout processus stationnaire discret $\{\xi_t\}$ peut être décomposé en un processus complètement déterministe $\{\eta_t\}$ et un processus complètement non déterministe $\{\zeta_t\}$ ⁽⁴¹⁾ suivant la relation

$$\xi_t = \eta_t + \zeta_t \quad (8)$$

avec

$$\zeta_t = \varepsilon_t + b_1 \varepsilon_{t-1} + \dots + b_l \varepsilon_{t-l} + \dots \quad (9)$$

La composante η_t est une fonction périodique, ou presque périodique lorsque certaines composantes ont des périodes incommensurables.

Les coefficients b_i de la composante non déterministe sont des constantes, $\sum b_i^2$ est fini, et ε_t est une variable aléatoire non autocorrélée et dont la distribution est quelconque. Le processus $\{\zeta_t\}$ est dit « processus de sommation mobile ».

Le processus $\{\eta_t\}$ caractérise le déterminisme, et le processus $\{\zeta_t\}$ l'aléatoire.

La simulation de l'aléatoire par le déterminisme.

Au regard du théorème de décomposition de Wold la proposition suivant laquelle la composante déterministe η_t peut simuler l'aléatoire peut apparaître à première vue quelque peu paradoxale, et à vrai dire insoutenable. Elle n'en est pas moins très correcte.

(P.36) Voir ALLAIS, 1980, notamment p. 33,8-34.

(P.37) Pour les valeurs cumulées de la série U, ces régularités n'apparaissent que si l'on considère un domaine de variation de n , suffisamment grand, en l'espèce ($1 \leq n \leq 3\,000$). Voir la note (U.25) ci-dessous.

(P.38) Faute de place je ne puis présenter cette analyse dans le présent mémoire. Elle fera l'objet d'une prochaine publication.

(P.39) Voir *Fréquence, Probabilité et Hasard*, § 10, p. 27-28, J.S.S.P., tome 124, n° 2. Voir également le § P. 4 ci-dessus.

(P.40) WOLD, 1938, p. 89; 1953, p. 200; et 1965, p. 61. Voir également les théorèmes de décomposition de CRAMER (1940) et de KOLMOGOROV (1941). Pour un bref exposé de ces théorèmes, voir WOLD, 1965, p. 60-62. J'ai utilisé le théorème de Wold dans ma généralisation du test de périodicité de Schuster au cas des séries autocorrélées (ALLAIS, 1957 d; 1960; et 1961).

(P.41) C'est à-dire en considérant qu'il n'y a pas de partie singulière dans le spectre (c'est à-dire pas de partie continue, mais non différentiable), (WOLD, 1938, p. 202-234; GRENDER et ROSENBLATT, 1957, p. 35).

Si les coefficients b_i de la relation (9) ne sont pas trop différents les uns des autres, il résulte du *Théorème central limite* que la distribution asymptotique de la variable aléatoire ζ_i , lorsque le nombre de ses composantes ε_{i-l} , augmente indéfiniment est normale.

De même il résulte du *Théorème (T)* que sous certaines conditions la distribution de la fonction déterministe η_i tend vers la loi normale lorsque N et l augmentent indéfiniment.

On voit ainsi que dans les deux cas les distributions sont normales bien que les structures sous-jacentes soient totalement différentes.

Cas de l'indépendance des valeurs successives de la composante aléatoire ζ_i

Les valeurs successives de la variable aléatoire ζ_i , sont indépendantes si tous les b_i sont nuls. On a alors $\zeta_i = \varepsilon_i$, et si la distribution est normale, il en est de même de ζ_i .

Par contre, et même si le *Théorème (T)* s'applique, le coefficient d'autocorrélation asymptotique ρ_q^* (η_i, η_{i-q}) de la composante déterministe η_i , n'est jamais identiquement nul et il est lui-même une fonction presque périodique ⁽⁴²⁾.

Cas des séries observées

Dans la réalité les séries observées ne comprennent toujours qu'un nombre fini de termes. Il résulte de là que pour les deux processus déterministe et aléatoire les propriétés asymptotiques ne seront vérifiées qu'*approximativement*.

Il peut dès lors devenir impossible de distinguer les deux processus, car leurs apparences peuvent être pratiquement identiques.

Du point de vue de l'analyse de la réalité concrète, c'est naturellement une erreur que de s'intéresser *uniquement* aux propriétés *asymptotiques* des processus ⁽⁴³⁾. Quant à leur applicabilité à la réalité concrète, les théorèmes asymptotiques *ne sont effectivement intéressants que s'ils restent approximativement valides pour des séries de longueur limitée*, comme c'est le cas du *Théorème (T)* ⁽⁴⁴⁾. Les séries temporelles observées sont *toujours* en effet d'une longueur limitée.

De toutes ces indications, il résulte que concrètement il sera souvent très difficile, sinon impossible, de distinguer l'aléatoire de la presque périodicité, et la presque périodicité de la périodicité ⁽⁴⁵⁾.

(P.42) Voir ci dessous § Q.10.

(P.43) Comme le font trop de mathématiciens.

Il est caractéristique que trois probabilistes aussi éminents que J. Lindeberg, P. Levy et H. Cramer, dont les travaux théoriques sur le *Théorème central limite* sont si remarquables, n'aient effectué *aucun calcul numérique* pour déterminer les conditions d'une rapidité plus ou moins grande de la convergence vers la loi normale. Ce travers des mathématiciens est éminemment préjudiciable au progrès de la science.

(P.44) Démontré dans la Section Q ci-dessous et dont des illustrations sont présentées dans les Sections S, T et U ci-dessous (Graphiques S 2, T 2, T 5, U 2 et U 5).

(P.45) Le processus aléatoire $\{\zeta_i\}$ est caractérisé par un spectre continu et le processus déterministe $\{\eta_i\}$ est généralement caractérisé par un spectre discontinu.

Cependant si on considère la fonction de la variable entière n

$$X(n) = \int_{f_1}^{f_2} \left[\cos 2\pi f [n - n(f)] \right] h(f) df \quad (f_1 \leq f \leq f_2; 1 \leq n \leq N) \quad (1)$$

ou, dans le cas continu, la fonction de la variable continue t

$$X(t) = \int_{f_1}^{f_2} \left[\cos 2\pi f [t - \theta(f)] \right] h(f) df \quad (f_1 \leq f \leq f_2) \quad (2)$$

la fonction $h(f)$ représentant la densité de la distribution des amplitudes, les fonctions *déterministes* $X(n)$ et $X(t)$ ont des *spectres continus*. En fait, les spectres continus sont trop souvent interprétés comme résultant toujours de processus « *aléatoires* ».

P.11. *Propriétés générales des fonctions presque périodiques*

Au regard des illustrations numériques présentées ci-dessous dans les Sections S, T et U, il me paraît souhaitable de rappeler brièvement quelques propriétés fondamentales des *fonctions presque périodiques* en me limitant au cas discret ⁽⁴⁶⁾.

a) *Définition*

Pour simplifier je me borne dans le présent mémoire à la considération des fonctions presque périodiques définies comme des sommes de fonctions sinusoïdales dont certaines périodes sont incommensurables. Ces fonctions sont bornées supérieurement.

C'est là la définition considérée par P. Bohl, 1893, et reprise sous une forme analogue par E. Esclangon, 1904 et 1917 ⁽⁴⁷⁾.

b) *Théorème fondamental*

Pour toute valeur de $\varepsilon > 0$ on peut trouver un nombre l fonction de ε tel qu'il existe *au moins un* nombre Θ

$$0 < \Theta \leq l \quad (10)$$

tel que l'on ait pour toute valeur de n ⁽⁴⁸⁾

$$|X(n + \Theta) - X(n)| < \varepsilon \quad -\infty < n < +\infty \quad (11)$$

Toute valeur Θ peut être interprétée comme une « *presque période* ».

Ce théorème tout à fait fondamental sera illustré dans ce qui suit pour des valeurs finies de N ($1 \leq n \leq N$) ⁽⁴⁹⁾ ⁽⁵⁰⁾.

c) *Presque périodes*

Dans le cas de la relation (1) du § P.3 ci-dessus où les périodes sont irrationnelles il existe toujours des entiers p_j et des nombres Θ tels que

$$\Theta - p_j T_j = \gamma_j T_j \quad |\gamma_j| < \varepsilon' \quad (12)$$

où ε' est aussi petit que l'on veut. On a alors

$$\begin{aligned} x_j(n + \Theta) &= \cos 2\pi f_j(n + \Theta - n_j) = \cos \left[2\pi f_j(n - n_j) + 2\pi(p_j + \gamma_j) \right] \\ &= \cos \left[2\pi f_j(n - n_j) + 2\pi \gamma_j \right] \end{aligned} \quad (13)$$

(P.46) Pour un bon exposé d'ensemble des propriétés des fonctions presque périodiques, voir CORDUNEANU, 1968. Voir également FAVARD, 1933.

(P.47) Esclangon appelait ces fonctions « fonctions quasi périodiques » (voir ESCLANGON, 1904 et 1917).

(P.48) CORDUNEANU, *id.*, p. 45.

(P.49) Voir § T.5 a.

(P.50) Cette propriété des fonctions presque périodiques au sens de Bohl et Esclangon a été prise par H. Bohr comme *définition des fonctions presque périodiques*. En fait, la définition des fonctions presque périodiques de H. Bohr est plus générale que celle de Bohl et Esclangon, et c'est elle qui est généralement adoptée aujourd'hui.

Qu'il me suffise d'indiquer ici qu'une fonction presque périodique au sens de H. Bohr peut être approximée avec une précision arbitrairement grande par une somme de sinusoïdes (H. BOHR, 1923). Voir ci dessous § Q.14 c. Voir également CORDUNEANU, 1968, p. 9.

Il convient toutefois d'indiquer qu'Esclangon a considéré lui même la définition de H. Bohr (1904, p. 116) en lui donnant même la préférence en raison de sa plus grande généralité (*id.*, p. 124).

On voit que si ε' est suffisamment petit la condition (11) sera satisfaite et Θ sera une presque période. Les presque périodes Θ apparaissent ainsi comme des communs multiples approchés des périodes T_j ⁽⁵⁰⁾.

Si l'amplitude a_k correspondant à une période T_k est beaucoup plus grande que toutes les autres, cette période T_k pourra être considérée comme une presque période.

d) *Moyen mouvement*

Sous certaines conditions une fonction presque périodique $X(t)$ peut s'écrire sous la forme

$$X(t) = \varrho(t) \cos \left[\frac{2\pi t}{T} + \varphi(t) \right] \quad (14)$$

où $\varrho(t)$ et $\varphi(t)$ sont des fonctions presque périodiques ⁽⁵¹⁾. Le facteur d'amplitude $\varrho(t)$ ne s'annule jamais ⁽⁵²⁾ et $\varphi(t)$ est borné supérieurement.

Il en sera ainsi notamment si une des amplitudes a_k est supérieure à la somme de toutes les autres. Dans ce cas la période T sera égale à la période T_k correspondante ⁽⁵³⁾. Mais ici naturellement le *Théorème (T)* ne sera plus applicable.

e) *Sur les fonctions presque périodiques et l'imitation du hasard*

La propriété fondamentale (11) des fonctions presque périodiques peut apparaître à première vue comme contradictoire avec la proposition suivant laquelle les valeurs d'une fonction presque périodique peuvent simuler une suite de caractère aléatoire. Si en effet, il existe une presque période Θ satisfaisant à la condition (11) où ε est suffisamment petit, il peut paraître évident *a priori* que la série des $X(n)$ ne saurait imiter ce qu'il est convenu d'appeler le « *hasard* ».

Cependant c'est là précisément tout l'intérêt du *Théorème (T)* que sous des conditions très générales la distribution asymptotique des valeurs $X(n)$ d'une fonction presque périodique est normale. *En ce sens* une telle fonction presque périodique peut être considérée comme imitant ce qu'il est communément convenu d'appeler le « *hasard* » ⁽⁵⁴⁾.

P.12. *Illustration des propriétés des séries temporelles observées par les propriétés des fonctions presque périodiques*

En fait l'objet de cet *Appendice II* est double :

a) Il est tout d'abord de montrer que sous des conditions très générales la distribution des valeurs successives d'une fonction presque périodique tend vers une distribution normale lorsque le nombre N des valeurs $X(n)$ considérées et le nombre l des composantes sinusoidales $x_j(n)$ augmentent indéfiniment ⁽⁵⁵⁾; et que cette propriété reste approximativement valable pour des valeurs finies de N et l ⁽⁵⁶⁾.

(P.50') Ces propriétés correspondent aux théorèmes de Dirichlet et Kronecker. Voir CORDUNEANU, 1968, § 1, 2, p. 14-23 et § VI, 3, p. 146-149; et HARDY et WRIGHT, 1959, § 11, 12, p. 169 170; et Chapitre XXIII, p. 375 393.

(P.51) Sur les conditions d'existence d'un moyen mouvement pour une fonction presque périodique, voir : HARTMAN, von KAMPEN et A. WINTNER, 1934; B. JESSEN et H. TORNHAVE, 1945; H. BOHR et B. JESSEN, 1949.

(P.52) Ainsi $\varrho(t)$ peut être égal à la somme d'une constante et de composantes sinusoidales.

(P.53) Tel est le cas des longitudes des planètes qui sont des fonctions presque périodiques ayant un moyen mouvement.

Tel est le cas encore, à mon avis, de la série des taches du soleil, les périodes qui interviennent étant des périodes dérivées des périodes des mouvements planétaires. Cette question fera l'objet d'une prochaine publication.

Tel est le cas encore au moins dans un large domaine des valeurs cumulées $Y(n)$ et $\pi^*(q)$ du § T.5 ci dessus.

(P.54) Voir ci dessus § P.10.

En dernière analyse il semble bien que l'opinion commune considère qu'une grandeur est déterminée par le « *hasard* » si elle a une distribution statistique. C'est là en tout cas la définition mathématique d'une variable dite *aléatoire*. Sur ces questions, voir ALLAIS, 1983 e, § 1, 1; 1, 6; 1, 7 et 1, 8.

(P.55) C'est là l'objet du *Théorème (T)* : § P.3 ci-dessus et Section Q ci-dessus.

(P.56) Trois illustrations sont données dans les Sections S, T et U ci-dessus.

b) Il est en second lieu de montrer que sous des conditions très générales et pour une valeur finie de N la série des valeurs successives de $X(n)$ peut présenter toutes les apparences d'une série de termes indépendants⁽⁵⁷⁾.

Dans le cas (a) les séries des valeurs successives d'une fonction presque périodique sont toujours autocorrélées. C'est là également le cas en général des séries temporelles observées en géophysique et en économie. Dans le cas (b) les apparences que présente la suite des valeurs successives d'une fonction presque périodique sont celles de séries dénuées de toute autocorrélation, cas effectivement très rare dans la réalité observée.

Dans les deux cas (a) et (b) il y a simulation du hasard en ce sens que la distribution des valeurs $X(n)$ est très voisine d'une distribution normale. Mais dans le cas (b) il y a cette circonstance supplémentaire que les valeurs successives $X(n)$ paraissent dénuées de toute autocorrélation.

Bien que le Théorème (T) soit un théorème asymptotique rigoureusement valable seulement pour des valeurs infinies de N et l , il peut rester approximativement valable pour des valeurs finies de N et l , comme le montre l'analyse des séries S, T et U présentées ci-dessous.

En ce sens, l'objet général de cet Appendice II n'est naturellement pas l'étude des fonctions presque périodiques $f(t)$ en général pour $-\infty < t < +\infty$, mais celle des propriétés des suites discrètes $X(n)$ des valeurs prises par des fonctions presque périodiques $X(t)$ pour des valeurs entières régulièrement espacées $1, 2, \dots, n, \dots, N$ de t lorsque certaines conditions sont réalisées et sous la condition que N a une valeur finie.

P.13. L'analyse des séries temporelles et l'algorithme des valeurs cumulées

Au regard de l'interprétation des résultats de certains calculs de valeurs cumulées⁽⁵⁸⁾ il paraît nécessaire de souligner les propriétés très différentes de l'intégrale d'une fonction sinusoïdale et de la valeur cumulée de cette fonction pour des valeurs entières de l'argument.

L'intégrale de la fonction

$$x_j(t) = a_j \cos 2\pi f_j(t - t_j)$$

s'écrit

$$y_j(t) = a_j^* \sin 2\pi f_j(t - t_j) \quad a_j^* = a_j/2\pi f_j \quad (15)$$

Il y a ainsi amplification des amplitudes des composantes sinusoïdales de faible fréquence, c'est-à-dire de grande période. Quant à la période elle est la même pour l'intégrale et pour la fonction.

Il en est tout différemment de la valeur cumulée

$$y_j(n) = \sum_{p=j}^n x_j(p) \quad (16)$$

D'après (4), (5), (6) et (16) on a en effet

$$y_j(n) = a_j^* [\sin [2\pi f_j n + \varphi_j] + \sin \theta_j] \quad (17)$$

en posant

$$a_j^* = a_j/2 \sin \pi f_j \quad \varphi_j = \lambda_j + (\mu_j/2) \quad \theta_j = (\mu_j/2) - \lambda_j \quad (18)$$

On voit tout d'abord que si f_j est très petit l'amplitude a_j^* est très voisine de $a_j/2\pi f_j$, comme dans le cas

(P.57) Une illustration est donnée dans la Section U ci-dessous. Voir notamment le § U.7.

(P.58) Voir notamment § T.5 et note (U.25) ci-dessous.

de la relation (15) ⁽⁵⁹⁾. Mais ici il y aura également amplification de l'amplitude si f_j est voisin d'un nombre entier. Ainsi si

$$f_j = q + \varepsilon_j \quad q = \text{entier} \quad \varepsilon \ll 1 \quad (19)$$

on a

$$a_j^* \sim a_j / 2 \pi \varepsilon_j \quad (20)$$

Dans ce cas on a également

$$\sin(2 \pi f_j n + \nu_j) = \sin(2 \pi \varepsilon_j n + \nu_j) \quad (21)$$

La période est donc égale à

$$T_j' = 1/\varepsilon_j \quad (22)$$

Elle est beaucoup plus grande que la période T_j . On voit ainsi qu'une fréquence voisine d'un nombre entier entraîne à la fois une amplification de l'amplitude et une augmentation considérable de la période.

Il résulte de là que la série des valeurs cumulées $Y(n)$ de $X(n)$ pourra comporter des composantes $y_j(n)$ de grandes périodes et de grandes amplitudes, et l'une d'elles pourra l'emporter sur toutes les autres dès lors que la fréquence f_j qui lui correspond sera la plus voisine de l'unité ⁽⁶⁰⁾ ⁽⁶¹⁾.

Cette circonstance qui résulte seulement du fait que l'on considère uniquement ici des valeurs entières de n n'a naturellement pas de signification réelle quant à la structure de la fonction presque périodique $X(t)$ puisqu'elle n'est qu'une conséquence de la seule considération des valeurs entières de t .

Mais dans tous les cas, que l'on considère ou non des séries discrètes, la considération des valeurs cumulées est susceptible de donner des indications précieuses sur la structure quasi-périodique, ou non, d'une série temporelle donnée par l'observation.

Q. LE THÉORÈME (T)

Q.1. La formulation classique du Théorème central limite

La présentation du théorème central limite dans la littérature est la suivante :

« Soit

$$X = x_1 + \dots + x_j + \dots + x_l \quad (1)$$

la somme de l variables aléatoires indépendantes dont les fonctions de distribution peuvent être très différentes et dont les moyennes et les écarts types sont $m_1, \dots, m_j, \dots, m_l$ et $\sigma_1, \dots, \sigma_j, \dots, \sigma_l$.

(P.59) On aura une circonstance analogue si

$$f_j = (q/p) + \varepsilon \quad (p \text{ et } q \text{ entiers}) \quad (1)$$

On a ici

$$\sin(2 \pi f_j n + \nu_j) = \sin(2 \pi q (n/p) + 2 \pi \varepsilon_j n + \nu_j) \quad (2)$$

Pour $n = 1, 1 + p, 1 + 2p + \dots$ tous les points représentatifs de cette fonction se trouvent sur une sinusoïde de fréquence ε_j et de période $T_j' = 1/\varepsilon_j$.

Au total tous les points de la sinusoïde (2) se répartissent sur p sinusoïdes de période T_j . Mais ici l'amplitude

$$a_j^* = a_j / \sin \pi [(q/p) + \varepsilon] \quad (3)$$

n'est généralement pas amplifiée sauf si le rapport des deux entiers p et q est voisin de l'unité.

(P.60) Voir une illustration de cette circonstance : § T.5 ci-dessous.

(P.61) A cette période pourra correspondre un moyen mouvement de la fonction $y_j(n)$ au sens de la relation (14) ci-dessus. Pour une illustration voir le § T.5 ci-dessous.

Soit

$$m = m_1 + \dots + m_j + \dots + m_l \quad (2)$$

$$\sigma^2 = \sigma_1^2 + \dots + \sigma_j^2 + \dots + \sigma_l^2 \quad (3)$$

Sous des conditions très générales relatives aux distributions des x_i , on a

$$\lim_{l \rightarrow \infty} P = \text{Probabilité} \left[\frac{X - m}{\sigma} \leq u \right] = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt \quad (4)$$

Autrement dit la distribution de la variable aléatoire X est de plus en plus voisine de la loi normale lorsque l augmente indéfiniment » (1) (2).

Ce théorème n'implique nullement que l'écart type σ de X reste fini quand l augmente indéfiniment.

Liapounoff a donné une condition suffisante, mais non nécessaire, de validité du théorème central limite (3). Cette condition est la suivante :

$$\lim_{l \rightarrow \infty} \frac{\rho}{\sigma} = 0 \quad (5)$$

σ étant défini par la relation (3) et ρ par les conditions

$$\rho^3 = \sum_j \rho_j^3 \quad \rho_j^3 = M(|x_j - m_j|^3) \quad (6)$$

où M représente la valeur moyenne.

La condition (5) signifie qu'aucune des variables x_j , n'est relativement importante par rapport à l'ensemble des autres (4).

Q.2. Reformulation fréquentielle du Théorème central limite

Si l'on examine attentivement ce que sont *effectivement* les calculs qui sont effectués dans la démonstration du Théorème central limite, on constate que le sens de ces calculs est tout à fait différent de ce que suggère la formulation habituelle du Théorème central limite et sa terminologie

(Q.1) CRAMER, 1946, § 17.4 17 5, p. 214-216, et 1959, § 7 4, p. 114 116

(Q 2) Sur ce théorème et ses conditions d'application, voir également Paul Lévy, 1930 et 1937, CRAMER, 1962 (1937).

Le problème de la recherche des conditions les plus générales de validité du théorème central limite a été résolu par Paul Lévy, Khintchine et Feller.

Sur l'historique de ce théorème, voir notamment CRAMER, 1946, id., p. 214 et 219; et surtout FRÉCHET, 1937, p. 108 110. Fréchet indique justement que « c'est à Laplace que l'on doit la considération de la loi normale dans la théorie des erreurs et plus généralement dans l'étude asymptotique de la loi de probabilité d'une somme de variables aléatoires indépendantes » Dans ces conditions le moins que l'on puisse faire c'est d'appeler la loi normale . « Loi de Laplace Gauss ».

(Q.3) CRAMER, 1946, id., p. 215 216

(Q.4) Si par exemple on a

$$\sigma_1 = \dots = \sigma_j = \dots = \sigma_l \quad \rho_1 = \dots = \rho_j = \dots = \rho_l \quad (1)$$

on a

$$\sigma = \sqrt{l} \sigma_1 \quad \rho = \sqrt[3]{l} \rho_1 \quad (2)$$

et par suite

$$\rho/\sigma = \rho_1/\sqrt[6]{l} \sigma_1 \quad (3)$$

et la condition (5) est effectivement satisfaite.

Dans ce cas particulier σ et ρ augmentent indéfiniment avec l . En fait d'une manière générale la validité du théorème central limite n'implique en aucune façon que σ et ρ restent finis. Cette circonstance est illustrée très clairement dans le cas considéré par Liapounoff

a) L'emploi du terme « *variables aléatoires* » suggère que les valeurs des variables x , résultent d'un tirage au hasard. En fait il n'en est rien, car toutes les réalisations également possibles des x , sont considérées *simultanément*.

b) *Nulle part* dans les démonstrations du théorème central limite n'interviennent les concepts de *hasard* ou de *choix au hasard*.

c) Les démonstrations résultent d'un calcul de fréquences de configurations qui se réalisent *simultanément*.

Dans les démonstrations présentées dans la littérature ces trois propositions n'apparaissent pas explicitement, et elles sont plus ou moins masquées par la complexité des calculs relatifs aux probabilités continues ⁽⁵⁾ ⁽⁶⁾.

En fait au lieu de parler de *variables aléatoires*, il convient de parler de *variables fréquentielles*, définies comme des variables dont on prend en compte simultanément toutes les valeurs ⁽⁷⁾, et il convient d'énoncer le Théorème central limite ainsi qu'il suit :

X étant la somme de l variables fréquentielles indépendantes, on a sous des conditions très générales.

$$\lim_{l \rightarrow \infty} F = \text{Fréquence} \left\{ \frac{X - m}{\sigma} \leq u \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt \quad (7)$$

Le calcul correspondant à la démonstration de ce théorème suppose simultanément réalisées toutes les trajectoires possibles ⁽⁸⁾, et la fréquence F est la fréquence de ces trajectoires pour lesquelles on a

$$(X - m)/\sigma \leq u \quad (8)$$

Les distributions considérées sont ainsi des distributions *déterministes* de variables fréquentielles dont on considère simultanément toutes les valeurs possibles.

Il résulte des indications qui précèdent que si la *condition de Liapounoff* ⁽⁵⁾ ci-dessus suffit pour assurer la validité du théorème central limite, elle n'est nullement nécessaire.

Q.3. Théorème (T)

Par application du Théorème central limite pour les fréquences ⁽⁹⁾ nous allons démontrer le Théorème (T) suivant :

Hypothèses (H)

On considère une fonction X (n) des valeurs entières n satisfaisant aux hypothèses suivantes :

$$(A) \quad X(n) = \sum_{j=1}^l x_j(n) = \sum_{j=1}^l a_j \cos 2\pi f_j (n - n_j) \quad n = \text{nombre entier} \quad (9)$$

$$a_j > 0 \quad f_j = 1/T_j > 0 \quad 1 \leq n \leq N \quad 1 \leq j \leq l \quad (10)$$

n_j : valeurs réelles quelconques

$$m_N = \frac{1}{N} \sum_{n=1}^N X(n) \quad \sigma_N^2 = \text{variance de } X(n) \quad (11)$$

(Q.5) Il y a en effet dans ce cas pour chaque variable x , une infinité de trajectoires au sens de l'Appendice IA de mon précédent article (J.S.S.P., tome 124, n° 2, 1983, P. 93-98) et l'égalité possible se traduit par l'égalité des mesures de sous-ensembles

(Q.6) Sur tous ces points voir les indications déjà données dans mon précédent article (id. Appendice IC, p. 101-102).

(Q.7) Voir l'Appendice IA, § 3 à mon précédent article (id. p. 98).

(Q.8) Au sens de l'Appendice IA (id. p. 93-98).

(Q.9) § Q.2 ci-dessus

$$(B) \quad \text{Il n'existe aucun système d'entiers } \alpha_j, \beta \text{ tels que } \sum_j \alpha_j f_j = \beta \quad (12)$$

$$(C) \quad \lim_{i \rightarrow \infty} \left[\frac{\sqrt[3]{\sum a_j^3}}{\sqrt[2]{\sum a_j^2}} \right] = 0 \quad (13)$$

Théorème (T)

Sous les hypothèses (H) on a

$$\lim_{i \rightarrow \infty} \left\{ \lim_{N \rightarrow \infty} F_N = \text{Fréquence} \left[\frac{X(n) - m_N}{\sigma_N} \right] \right\} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^u e^{-\frac{t^2}{2}} dt \quad (14)$$

La démonstration de ce théorème s'appuie sur les lemmes suivants :

Q.4. Lemmes

Lemme I

La distribution des valeurs de t/T étant supposée uniforme, la densité de fréquence $f(x)$ des valeurs de

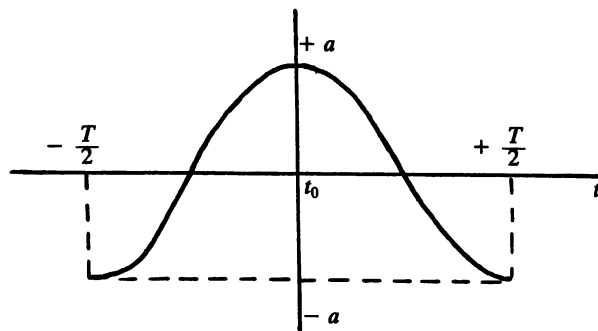
$$x = a \cos \frac{2\pi}{T} (t - t_0) \quad -T/2 \leq t - t_0 \leq T/2 \quad (15)$$

est

$$f(x) = 1/\pi \sqrt{a^2 - x^2} \quad -a \leq x \leq a \quad (16)$$

Considérons en effet les valeurs de x dans l'intervalle $-T/2 \leq t - t_0 \leq T/2$

FIG 1



Dans l'intervalle $-T/2 \leq t - t_0 \leq 0$ la densité $f(x)$ est définie par la condition

$$f(x) dx = 2 dt/T \quad -a \leq x \leq a \quad -T/2 \leq t - t_0 \leq 0 \quad (17)$$

Cette relation définit la *correspondance* entre x et t assurant l'égalité des mesures des sous-ensembles de cas *également possibles*. A deux intervalles égaux sur l'axe des t correspondent deux sous-ensembles de valeurs de x également possibles.

On a ainsi

$$f(x) = \frac{2}{T} \frac{1}{dx/dt} = \frac{2}{-a 2\pi \sin[2\pi(t-t_0)/T]} = \frac{1}{\pi\sqrt{a^2-x^2}} \quad (18)$$

Comme la courbe représentative de $x(t)$ est symétrique par rapport à $t = t_0$ l'expression (16) représente bien la densité de fréquence dans tout l'intervalle $-T/2 \leq t - t_0 \leq T/2$.

On a

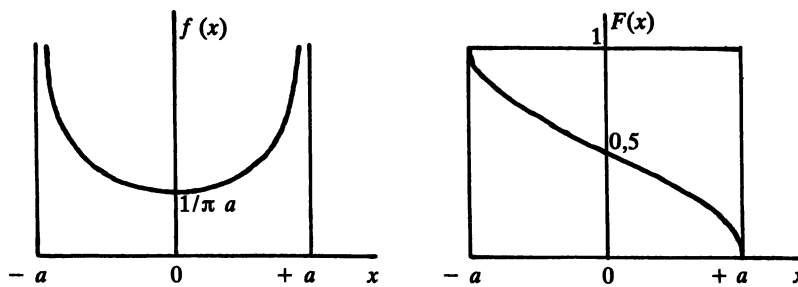
$$F(x) = \int_x^a f(x) dx = \frac{1}{2} - \frac{1}{\pi} \arcsin \frac{x}{a} \quad (-a \leq x \leq a) \quad (19)$$

avec

$$F(-a) = 1 \quad F(a) = 0 \quad (20)$$

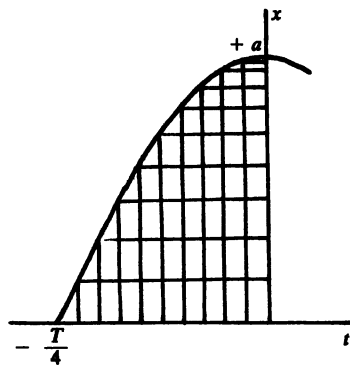
Les courbes représentatives de $f(x)$ et $F(x)$ sont les suivantes ⁽¹⁰⁾ :

FIG. 2



(Q.10) Il est facile de se rendre compte que la densité $f(x)$ est effectivement infinie pour $x = \pm a$ en considérant la figure suivante où sont représentées les valeurs positives de x correspondant à des valeurs régulièrement espacées de t . Ces valeurs sont d'autant plus rapprochées qu'elles se trouvent plus voisines de la valeur $+a$.

FIG 3



On a ⁽¹¹⁾ ⁽¹²⁾

$$m_x = \int_{-a}^{+a} x f(x) dx = 0 \quad \sigma_x^2 = \int_{-a}^{+a} x^2 f(x) dx = a^2/2 \quad (21)$$

$$\rho_x^3 = \int_{-a}^{+a} |x^3| f(x) dx = 4a^3/3\pi \quad (22)$$

Il est très remarquable que la distribution de la fonction $x(t)$ soit *indépendante* de la période T .

Lemme II

Lorsque N augmente indéfiniment la distribution discrète des fréquences des

$$x_j(n) = a \cos 2\pi f_j(n - n_j) \quad 1 \leq n \leq N \quad (23)$$

où n prend les valeurs entières successives $1, 2, \dots, N$, tend vers

$$f(x_j) = 1/\pi \sqrt{a^2 - x_j^2} \quad (24)$$

En effet, la fréquence f_j étant supposée irrationnelle d'après l'*Hypothèse (B)*, toutes les valeurs de x_j , lorsque N augmente indéfiniment, apparaissent avec la même densité de fréquence que dans le cas de la fonction continue $x(t)$ définie par la relation (15) ci-dessus ⁽¹³⁾, d'où, d'après le *Lemme I*, l'expression (24) pour la densité de fréquence $f(x_j)$.

Lemme III

Les conditions d'indépendance des variables fréquentielles x_{jn} sont asymptotiquement remplies.

(Q.11) GROBNER et HOFREITER, *Integral Tafel*, 1973, Springer, Wien und Innsbruck, 2^e partie, p. 38, relations 4a et 4b.

(Q.12) On peut naturellement obtenir m , σ^2 et ρ^3 par un calcul direct.

On a en effet

$$m = \frac{1}{T} \int_{-T/2}^{+T/2} a \cos \frac{2\pi}{T} (t - t_0) dt = 0 \quad (1)$$

$$\sigma^2 = \frac{1}{T} \int_{-T/2}^{+T/2} a^2 \cos^2 \frac{2\pi}{T} (t - t_0) dt = \frac{1}{T} \int_{-T/2}^{+T/2} \frac{a^2}{2} \left[1 + \cos \frac{4\pi}{T} (t - t_0) \right] dt = \frac{a^2}{2} \quad (2)$$

$$\begin{aligned} \rho^3 &= \frac{1}{T} \int_{-T/2}^{+T/2} a^3 \left| \cos^3 \frac{2\pi}{T} (t - t_0) \right| dt = \frac{2a^3}{T} \int_{-T/4}^{+T/4} \cos^3 \frac{2\pi}{T} (t - t_0) dt \\ &= \frac{a^3}{2T} \int_{-T/4}^{+T/4} \left[3 \cos \frac{2\pi}{T} (t - t_0) + \cos 3 \frac{2\pi}{T} (t - t_0) \right] dt = \frac{4a^3}{3\pi} \end{aligned} \quad (3)$$

(Q.13) Soit

$$\tau_{jn} = n - n_j - qT_j \quad 0 \leq n - n_j - qT_j \leq T_j \quad (1)$$

T_j étant supposé irrationnel et q étant un nombre entier satisfaisant à l'inégalité (1).

Lorsque N augmente indéfiniment la répartition des points d'abscisse τ_{jn} tend vers une distribution continue uniforme dans l'intervalle $(0, T_j)$ (voir notamment BOREL, 1926, chapitre I, RAUZY, 1976, chapitre I; et tout spécialement BASS, 1971, tome III, Théorème de Weyl, § V.3.6, p. 378-382) Sur le théorème de Weyl, voir également HARDY et WEIGHT, 1959, § 23.10, p. 390-393.

Comme on a

$$x_j(n) = a_j \cos \frac{2\pi}{T_j} (n - n_j) = a_j \cos \frac{2\pi}{T_j} (\tau_{jn} + qT_j) = a_j \cos \frac{2\pi}{T_j} \tau_{jn}$$

on vérifie que lorsque N augmente indéfiniment la densité f de $x_j(n)$ tend effectivement vers la densité $f(x_j)$ définie par la relation (24) et correspondant à la relation (23) (*Lemme I*)

a) *Condition d'indépendance*

Rappelons tout d'abord que la condition nécessaire et suffisante de l'indépendance asymptotique des fonctions $x_1(n), x_2(n), \dots, x_l(n)$ est

$$M [e^{i \sum \tau_j x_j(n)}] \underset{N \rightarrow \infty}{=} M [e^{i \tau_1 x_1(n)}] M [e^{i \tau_2 x_2(n)}] \dots M [e^{i \tau_l x_l(n)}] \quad (25)$$

quelles que soient les valeurs des variables réelles $\tau_1, \tau_2, \dots, \tau_l$, l'opérateur M signifiant que l'on considère les moyennes pour $1 \leq n \leq N$ (14).

Si on développe les fonctions exponentielles en séries de Taylor par rapport aux τ_j , on voit que la condition (25) équivaut aux conditions

$$M [x_1^p(n) x_2^q(n) \dots x_l^r(n)] \underset{N \rightarrow \infty}{=} M [x_1^p(n)] M [x_2^q(n)] \dots M [x_l^r(n)] \quad (26)$$

où p, q, \dots, r prennent respectivement toutes les valeurs entières possibles ou des valeurs nulles (15).

b) *Conditions d'absence de corrélation asymptotique de $x_j(n)$ et $x_k(n)$*

Pour simplifier la discussion des implications de la condition de l'indépendance asymptotique des $x_j(n)$ considérons tout d'abord le cas $p = 1, q = 1$, les autres exposants r étant égaux à zéro. L'équation (26) représente alors la condition d'absence de corrélation asymptotique entre $x_j(n)$ et $x_k(n)$ (16).

$$\underset{N \rightarrow \infty}{M} [x_j(n) x_k(n)] = 0 \quad (27)$$

On a

$$M [x_j(n) x_k(n)] = \frac{a_j a_k}{N} \sum_{n=1}^N \cos 2\pi f_j (n - n_j) \cos 2\pi f_k (n - n_k) \quad (1 \leq n \leq N) \quad (28)$$

$$2 \cos 2\pi f_j (n - n_j) \cos 2\pi f_k (n - n_k) = \cos [2\pi (f_j + f_k)n - 2\pi (n_j f_j + n_k f_k)] \\ + \cos [2\pi (f_j - f_k)n - 2\pi (n_j f_j - n_k f_k)] \quad (29)$$

On sait que

$$\cos u + \cos (u + h) + \dots + \cos (u + Nh) = \cos \left(u + \frac{Nh}{2} \right) \sin (N + 1) \frac{h}{2} / \sin \frac{h}{2} \quad (30)$$

Des conditions (28), (29), (30) et de l'*Hypothèse (B)* il résulte que l'on a ici

$$\sin (h/2) = \sin \pi (f_j \pm f_k) \neq 0 \quad (31)$$

et que la condition (27) est satisfaite.

(Q.14) CRAMER, 1946, § 21.3, p. 266, et 22.4, p. 296.

De la relation (25) on déduit que l'on a quel que soit t

$$M [e^{it \sum x_j(n)}] \underset{N \rightarrow \infty}{=} M [e^{it x_1(n)}] M [e^{it x_2(n)}] \dots M [e^{it x_l(n)}] \quad (1)$$

C'est cette condition qui précisément intervient dans la démonstration du théorème central limite de Liapounoff (CRAMER, id., p. 216, relation 17.4.4.).

(Q.15) Les conditions (26) sont à la fois des conditions nécessaires et suffisantes de l'indépendance suivant un théorème de Robert Fortet.

(Q.16) CRAMER, id., p. 265.

c) *Condition d'indépendance asymptotique de $x_j(n)$ et $x_k(n)$*

Considérons maintenant le cas où p et q ont des valeurs quelconques, les autres exposants r ayant des valeurs nulles. On a

$$M [x_{jn}^p x_{kn}^q] = a_j^p a_k^q M [(\cos 2\pi f_j(n - n_j))^p (\cos 2\pi f_k(n - n_k))^q] \quad (32)$$

$$M [x_{jn}^p] = a_j^p M [(\cos 2\pi f_j(n - n_j))^p] \quad (33)$$

$$M [x_{kn}^q] = a_k^q M [(\cos 2\pi f_k(n - n_k))^q] \quad (34)$$

On sait que suivant que p est pair ou impair on a (17).

$$2^{p-1} \cos^p u = \cos pu + C_p^1 \cos(p-2)u + C_p^2 \cos(p-4)u + \dots + \frac{1}{2} C_p^{p/2} \quad (35)$$

$$2^{p-1} \cos^p u = \cos pu + C_p^1 \cos(p-2)u + C_p^2 \cos(p-4)u + \dots + C_p^{p-1} \cos u \quad (36)$$

D'après (35) et (36) le calcul des trois quantités (32), (33) et (34) est analogue à celui de l'expression (27).

Pour toutes les sommes de cosinus, nous avons ainsi dans le cas de (33) ou (34)

$$\sin(h/2) = \sin \pi \alpha_j f_j \quad (37)$$

où α_j est entier, et dans le cas de (32)

$$\sin(h/2) = \sin \pi [\alpha_j f_j + \alpha_k f_k] \quad (38)$$

où α_j et α_k sont entiers. Ainsi d'après l'Hypothèse (B) on a dans le cas de (37) et (38)

$$\sin(h/2) \neq 0 \quad (39)$$

Si un des entiers p et q est impair, il résulte alors de relations semblables à (28), (29) et (30), et des relations (35), (36) et (39) que

$$\lim_{N \rightarrow \infty} M [x_j^p(n) x_k^q(n)] = 0 \quad \lim_{N \rightarrow \infty} M [x_j^p(n)] M [x_k^q(n)] = 0 \quad (40)$$

Si p et q sont tous deux pairs on a

$$\lim_{N \rightarrow \infty} M [x_j^p(n) x_k^q(n)] = [C_p^{p/2} a_j^p / 2^p] [C_q^{q/2} a_k^q / 2^q] \quad (41)$$

$$\lim_{N \rightarrow \infty} M [x_j^p(n)] = C_p^{p/2} a_j^p / 2^p \quad M [x_k^q(n)] = C_q^{q/2} a_k^q / 2^q \quad (42)$$

Il résulte des relations (40), (41) et (42) que les conditions (26) sont effectivement satisfaites dans le cas considéré.

d) *Cas général*

Dans le cas général le même raisonnement montre qu'il résulte de l'Hypothèse (B) que les conditions d'indépendance (26) sont satisfaites (18) (19).

(Q.17) Edwin P. ADAMS, 1947, *Smithsonian Mathematical Formulae*, p. 68-69, Smithsonian Institution.

(Q.18) On pourrait tout aussi bien fonder la démonstration de l'indépendance asymptotique des $x_j(n)$ et $x_k(n)$ en partant directement de la relation (25) et en considérant les développements en séries de Bessel

$$e^{i a_j \cos 2\pi f_j(n - n_j)} = 1 + 2 \sum_{k=1}^{\infty} i^k J_k(a_j) \cos 2\pi k f_j(n - n_j) \quad (1)$$

ou les J_k représentent les fonctions de Bessel. (RYSNIK et GRADSTEIN, 1957, *Tables*, Deutscher Verlag der Wissenschaften, p. 331.)

Le calcul des moyennes de la relation (26) se ferait alors à partir de la considération de relations analogues aux relations (32), (33) et (34).

Lemme IV

Lorsque N augmente indéfiniment, la distribution des $X(n)$ tend vers la distribution des

$$X(t) = \sum_{j=1}^l a_j \cos 2\pi f_j (t - n_j) \quad -T_j/2 \leq t - n_j \leq T_j/2 \quad (43)$$

Cette proposition résulte du *Lemme II*.

D'après les *Lemmes I, II et III* on a

$$m_x = 0 \quad \sigma_x^2 = \sum_j \sigma_j^2 = \sum_j a_j^2/2 \quad \rho_x^3 = \sum \rho_j^3 = 4 \sum_j a_j^3/3\pi \quad (44)$$

Lemme V

L'Hypothèse (C) entraîne la validité de la condition de Liapounoff ⁽²⁰⁾

$$\lim_{l \rightarrow \infty} \mathcal{L} = \frac{\rho_x}{\sigma_x} = 0 \quad (45)$$

On a en effet d'après les *Lemmes I, II, III*, les relations (44), et l'Hypothèse (C)

$$\lim_{l \rightarrow \infty} \mathcal{L} = \frac{\rho_x}{\sigma_x} = \lim_{N \rightarrow \infty} \delta \frac{3\sqrt{\sum a_j^3}}{2\sqrt{\sum a_j^2}} = 0 \quad (\delta = 2^{7/6}/\sqrt[3]{3\pi} = 1,0628..) \quad (46)$$

La condition (C) implique donc bien la condition de Liapounoff (45).

Q.5. Démonstration du Théorème (T)

Il résulte :

- des *Lemmes I et II* que les densités de fréquence des fonctions $x_j(n)$ sont définies par les relations (24);
- du *Lemme III* que ces fonctions sont asymptotiquement indépendantes;
- du *Lemme IV* que la distribution de $X(n)$ s'identifie avec celle de $X(t)$ lorsque N augmente indéfiniment;
- et du *Lemme V* que l'Hypothèse (C) entraîne la validité de la Condition de Liapounoff (45).

La démonstration du *Théorème (T)* résulte alors de la simple transposition ⁽²¹⁾ de la démonstra-

(Q.19) On pourrait également, presque sans aucun calcul, fonder la démonstration de l'indépendance asymptotique des $x_j(n)$ et $x_k(n)$ sur le fait qu'en posant

$$u_j(n) = [nf_j]d \quad (1)$$

où la notation $[v]_d$ désigne la partie décimale de v , on a d'après l'Hypothèse B

$$\text{Fréquence } \{\lambda_j < u_j(n) < \mu_j\} \underset{N \rightarrow \infty}{=} \mu_j - \lambda_j \quad (2)$$

$$\text{Fréquence } \{\lambda_1 < u_1(n) < \mu_1, \dots, \lambda_l < u_l(n) < \mu_l\} \quad (3)$$

$$\underset{N \rightarrow \infty}{=} \text{Fréquence } \{\lambda_1 < u_1(n) < \mu_1\} \dots \dots \dots \text{Fréquence } \{\lambda_l < u_l(n) < \mu_l\}$$

D'où il résulte que les fonctions $u_1(n), u_2(n), \dots, u_l(n)$ sont asymptotiquement indépendantes, et par suite, d'après un théorème bien connu (Harald CRAMER, 1946, *Mathematical Methods of Statistics*, § 14.5, p. 164) que les fonctions $x_j(n)$, fonctions des $u_j(n)$, sont indépendantes.

Naturellement et réciproquement si les fonctions $x_j(n)$ sont indépendantes il en est de même des fonctions $u_j(n)$.

(Q.20) § Q.1 et Q.2 ci dessus

(Q.21) En remplaçant toute variable aléatoire u_n par une fonction $u(n)$ de la variable réelle n , et le mot probabilité par fréquence, la démonstration reste identiquement la même

Sur la démonstration de Liapounoff du *Théorème central limite* voir H. CRAMER, 1946, id., § 17.4, p. 215-217

tion du *Théorème central limite* dans le cas considéré par Liapounoff lorsqu'on substitue aux variables aléatoires x_n les fonctions $x_j(n)$ de la variable réelle n .

La démonstration du *Théorème (T)* n'implique nullement que l'écart type σ_N de $X(N)$ reste fini quand l augmente indéfiniment. Sa validité est indépendante des phases n_j , et la distribution limite de $X(n)$ ne dépend que des amplitudes a_j .

La démonstration du *Théorème (T)* ⁽²²⁾ et les conditions qu'elle implique donnent lieu aux commentaires suivants.

Q.6. Conditions d'application du *Théorème central limite*

Tout d'abord, il convient de souligner que la validité du *Théorème central limite* pour une série temporelle $X_n = \sum_j x_{jn}$ n'est nullement subordonnée à la condition que deux valeurs successives des variables « dites aléatoires » x_{jn} soient indépendantes. Sa démonstration suppose en effet que toutes les valeurs de ces variables sont considérées *simultanément*, leur densité étant définie par leur loi de distribution ⁽²³⁾. Peu importe qu'on les écrive dans un certain ordre. La seule condition est que la fréquence de leur apparition soit bien celle de la fonction de distribution qui les caractérise. Ce que l'on considère en effet ce sont les distributions des valeurs des x_{jn} et X_n *indépendamment de leur ordre de succession*.

De ce point de vue la seule condition à laquelle la démonstration du *Théorème central limite* est subordonnée est que les variables considérées soient *mathématiquement indépendantes*, c'est-à-dire que la condition (25) du *Lemme III* soit effectivement satisfaite ⁽²⁴⁾.

Il résulte de là que le *Théorème central limite* s'applique tout aussi bien lorsqu'on substitue aux variables « aléatoires » x_{jn} des fonctions $x_j(n)$ de la variable réelle n , dès lors que les conditions correspondant aux *Lemmes II, III, IV, et V* sont satisfaites. *Le Théorème central limite s'applique ainsi à des phénomènes totalement déterministes.*

C'est là une propriété générale. Les variables considérées dans les traités sur le « *Calcul des probabilités* » n'ont *aucun caractère aléatoire* et les théories axiomatiques des probabilités *ne font intervenir en aucune façon le hasard*. Ces théories se réduisent à des théories des ensembles mesurables assorties de conditions d'indépendance. Une variable qualifiée d'« aléatoire » est simplement une variable à laquelle correspond une fonction de distribution. Rien n'est changé dans les théories axiomatiques des probabilités si l'on remplace les « variables aléatoires » u_i par des fonctions $u(t)$ d'une variable réelle t , ces fonctions ayant les mêmes distributions que les « variables aléatoires » u_i et satisfaisant aux mêmes conditions d'indépendance. En fait à toute proposition relative aux variables *arbitrairement qualifiées d'« aléatoires »* de ce qu'il est convenu d'appeler le « *Calcul des probabilités* » on peut faire correspondre une proposition analogue pour des fonctions d'une variable réelle *de caractère entièrement déterministe*, comme le montre l'applicabilité du *Théorème central limite* à des sommes de sinusoides satisfaisant aux *Hypothèses (H)*. Les propositions et les démonstrations sont mathématiquement identiques. *Seules diffèrent la sémantique utilisée et les interprétations qu'elle suggère.*

La sémantique d'aujourd'hui, profondément et malheureusement enracinée par l'usage, *ne peut que susciter de faux problèmes et de lourdes erreurs d'interprétation, en physique notamment.* La ten-

(Q.22) La démonstration du *Théorème (T)* et les conditions qu'elle implique ont fait l'objet de ma part de deux communications à l'Académie des Sciences : ALLAIS, 1983 a et 1983 b.

(Q.23) C'est une circonstance *tout à fait essentielle* qui à ma connaissance n'est explicitée *nulle part* dans la littérature. Voir mon article précédent, *Fréquence, Probabilité et Hasard*, Appendice I, § A.2 A.4, p. 96-98, et § C.1 C.3, p. 101-102.

(Q.24) Il suffit par exemple de se reporter à la démonstration du théorème de Liapounoff donnée par Cramer (1946, p. 215-218) pour le vérifier. Toute cette démonstration repose sur la relation 17.4 4 (p. 216), elle-même déduite de 15.2.1 (p. 188).

Il en est de même des démonstrations beaucoup plus élaborées données par Paul Lévy (1930 et 1937) et par Cramer (1937) pour des cas beaucoup plus généraux.

dance moderne à l'abstraction et à la généralité tend à masquer la nature réelle des calculs qui sont effectués et l'interprétation qu'elle implique.

Q.7. *Processus « aléatoire » (P*)*
correspondant au processus déterministe (P) des $X(n) = \sum_j x_j(n)$

Au processus déterministe P défini par la relation

$$X(n) = \sum_j x_j(n) \quad (1 \leq j \leq l ; 1 \leq n \leq N) \quad (47)$$

on peut faire correspondre un processus « aléatoire » P^* (qu'il vaut mieux appeler « fréquentiel ») défini par la relation

$$X_n^* = \sum_j x_{j,n}^* \quad (48)$$

où les variables « aléatoires » $x_{j,n}^*$ (qu'il vaut mieux appeler « fréquentielles ») ont pour densités de fréquence les fonctions

$$f(x_{j,n}^*) = 1/\pi \sqrt{a_j^2 - x_{j,n}^{*2}} \quad (49)$$

Le processus (P) peut être considéré comme une réalisation particulière du processus (P^*)⁽²⁵⁾, et il y a ergodicité en ce sens que la distribution des valeurs de la fonction $X(n)$ de la variable réelle n ($1 \leq n \leq N$) est asymptotiquement identique lorsque N augmente indéfiniment à la distribution de la variable fréquentielle (dite aléatoire) X_n^* dont les composantes $x_{j,n}^*$ ont pour densités de fréquence les fonctions (49)⁽²⁶⁾.

La distinction entre le processus déterministe (P), auquel correspond une fonction presque périodique $X(n)$, et le processus fréquentiel (P^*) qui peut lui être associé est absolument fondamentale, et il résulte des indications présentées ci-dessus que tous les théorèmes du « Calcul des Probabilités » applicables au processus fréquentiel (P^*) peuvent s'appliquer au cas du processus déterministe (P)⁽²⁷⁾. Comme je l'ai déjà souligné, la raison en est fondamentalement que la démonstration de ces théorèmes implique en fait que toutes les valeurs des variables « aléatoires » $x_{j,n}^*$ sont supposées, au moins implicitement, réalisées simultanément, leur densité étant définie par la relation (49).

Q.8. *Le Théorème (T) et les régularités constatées dans certains processus empiriques*

Pour expliquer les régularités constatées dans certains processus empiriques Poincaré a montré que quelles que soient, aux différents instants considérés, les fonctions arbitraires de répartition des fréquences correspondantes, la fréquence moyenne résultante est indépendante de ces fonctions arbitraires⁽²⁷⁾. En fait, les calculs effectués reposent simplement sur le calcul de certaines distributions de fréquence à partir de la considération d'autres distributions.

Le Théorème (T) dans sa formulation la plus générale permet de donner une « interprétation » bien plus suggestive encore puisqu'il n'est plus nécessaire de considérer à chaque instant une distribution

(Q.25) Naturellement si on considère toutes les réalisations possibles du processus (P^*) la fréquence du processus particulier (P) est nulle.

(Q.26) Sur l'ergodicité, voir notamment WOLD, 1953, § 3.3, p. 73; § 9.4, p. 156-158; et § 10 4, p. 166-168; WOLD, 1965, § 2.3.1., p. 52 54; KOOPMANS, 1974, § 2.9., p. 53 61, et NERLOVE, GREYER et CARVALHO, 1979, § 11 2, p. 23 30.

(Q.27) Voir POINCARÉ, 1912, § 92-93, p. 148-152; et, 1906, Chap. XI, p. 227 235.

de fréquences et que l'on considère simplement la distribution au cours du temps de la fonction $X(t)$ résultant de la somme des effets $x_j(t)$ caractérisant le processus considéré, la distribution des $X(t)$ convergeant vers la loi normale sous des conditions très générales. On est ramené ainsi à la conception de Laplace suivant laquelle le hasard résulte simplement de la superposition de causes très nombreuses, indépendantes, et dont aucune ne peut être considérée comme prépondérante.

Q.9. Le coefficient \mathcal{L} de Liapounoff

La condition (46) de Liapounoff, qui d'ailleurs *n'est qu'une condition suffisante et non nécessaire*, implique que les amplitudes a_j des $x_j(n)$ ne diffèrent pas trop les unes des autres. Mais c'est là dans tous les cas une condition générale du Théorème central limite.

En fait d'après (44) pour une valeur donnée de σ_x la valeur minimale du rapport ρ_x/σ_x est atteinte lorsque tous les a_j ont une même valeur; sa valeur maximale est atteinte lorsque tous les a_j sont nuls sauf un. On a ainsi d'après (46)

$$\delta/\sqrt{l} \leq \mathcal{L} = \rho_x/\sigma_x \leq \delta \quad (\delta = 2^{7/6}/\sqrt[3]{3\pi} = 1,0628\dots) \quad (50)$$

quels que soient les a_j . Posons

$$a_j = (1 + \alpha_j) a \quad a = \sum_j a_j/l \quad (51)$$

On a

$$\sum_j \alpha_j = 0 \quad (52)$$

et par suite

$$\mathcal{L}/\delta = \sqrt[3]{\sum_j a_j^3}/\sqrt{\sum_j a_j^2} = \sqrt[3]{l + 3\sum_j \alpha_j^2 + \sum_j \alpha_j^3}/\sqrt{l + \sum_j \alpha_j^2} \quad (53)$$

d'où

$$\mathcal{L}/\delta = \sqrt[3]{1 + \frac{3}{l} \sum_j \alpha_j^2 + \frac{1}{l} \sum_j \alpha_j^3} / \sqrt{l} \sqrt{1 + \frac{1}{l} \sum_j \alpha_j^2} \quad (54)$$

On a ainsi

$$\mathcal{L}/\delta \leq K/\sqrt{l} \quad K = \sqrt[3]{1 + \frac{3}{l} \sum_j \alpha_j^2 + \frac{1}{l} \sum_j \alpha_j^3} \quad (55)$$

Il suffit donc que la quantité K reste limitée supérieurement pour que \mathcal{L} décroisse indéfiniment avec l . Il en sera ainsi si $\sum_j \alpha_j^2$ et $\sum_j \alpha_j^3$ croissent moins rapidement que l . On peut encore dire qu'il suffit que les valeurs moyennes des carrés et des cubes des écarts restent limitées supérieurement.

Compte tenu de l'homogénéité de l'expression du coefficient de Liapounoff on peut toujours supposer sans perte de généralité que la moyenne a des a_j est égale à l'unité. La condition de Liapounoff implique alors qu'aucun a_j ne soit très différent de l'unité.

Q.10. Distribution asymptotique des coefficients d'autocorrélation asymptotiques ρ_q^* des $X(n)$

Lorsque N augmente indéfiniment le coefficient d'autocorrélation de la série temporelle $X(n)$ définie par la relation (9) a pour expression asymptotique ⁽²⁸⁾

$$\rho_q^* = \sum_j a_j^2 \cos 2\pi f_j q / \sum_j a_j^2 \quad (1 \leq j \leq l) \quad (56)$$

(Q 28) On a en effet

$$\rho_q^* = \lim_{N \rightarrow \infty} M[X_n X_{n+q}] / M[X_n^2] \quad (1 \leq n \leq N) \quad (1)$$

Un calcul analogue à celui correspondant au Lemme III ci dessus permet alors de déduire la relation (56).

Cette expression montre que le coefficient d'autocorrélation asymptotique ϱ_q^* est lui-même une fonction presque périodique. Elle est tout à fait analogue à l'expression (9) de $X(n)$, le paramètre n étant remplacé par le paramètre q , les a_j étant remplacés par les $a_j^2/\sum_j a_j^2$, et les n_j ayant ici des valeurs nulles.

Le coefficient de Liapounoff correspondant est d'après (46)

$$\mathcal{L}_1 = \rho/\sigma = \delta \sqrt[3]{\sum_j a_j^6} / \sqrt{\sum_j a_j^4} \quad \delta / \sqrt[6]{l} \leq \mathcal{L}_1 \leq \delta \quad (57)$$

On passe en effet de l'expression (9) de $X(n)$ à l'expression (56) de ϱ_q^* par remplacement des a_j par les $a_j^2/\sum_j a_j^2$.

Il résulte de là que si les a_j ne sont pas trop différents de leur moyenne le Théorème (T) du § Q.3. s'applique ici encore, de sorte qu'en général les ϱ_q^* se distribuent approximativement suivant la loi normale.

Mais la condition de Liapounoff est ici plus restrictive puisque les a_j sont remplacés par les $a_j^2/\sum_j a_j^2$ dont la dispersion est plus grande (29).

*Ordre de grandeur des ϱ_q^**

Si les coefficients a_j sont peu différents de leur moyenne, les coefficients d'autocorrélation asymptotiques ϱ_q^* des $X(n)$ seront relativement petits. Il suffit pour s'en rendre compte de considérer le cas où tous les a_j sont égaux à leur moyenne a (30). On a d'ailleurs

$$\text{variance } \varrho_q^* = M \left[\sum_j a_j^2 \cos 2\pi f_j q \right]^2 / \left(\sum_j a_j^2 \right)^2 \quad (58)$$

d'où

$$\text{variance } \varrho_q^* = \sum_j a_j^4 / 2 \left(\sum_j a_j^2 \right)^2 \quad (59)$$

On a (31)

$$1/2 l \leq \sum_j a_j^4 / 2 \left(\sum_j a_j^2 \right)^2 \leq 1/2 \quad (60)$$

Q.11. Coefficients d'autocorrélation asymptotiques ϱ_r^* des coefficients d'autocorrélation asymptotiques ϱ_q^*

Considérons le coefficient d'autocorrélation asymptotique ϱ_r^* des ϱ_q^*

$$\varrho_r^* = \text{coefficient de corrélation } (\varrho_q^*, \varrho_{q+r}^*) \quad (61)$$

Pour obtenir l'expression de ϱ_r^* , il suffit, compte tenu de (9) et (56), de remplacer dans (56) a_j par $a_j^2/\sum_j a_j^2$. On a ainsi

$$\varrho_r^* = \sum_j a_j^4 \cos 2\pi r f_j / \sum_j a_j^4 \quad (62)$$

$$\text{variance } \varrho_r^* = \sum_j a_j^8 / 2 \left(\sum_j a_j^4 \right)^2 \quad (63)$$

On voit ainsi que le coefficient de corrélation asymptotique ϱ_r^* est également une fonction presque périodique. Le coefficient de Liapounoff correspondant est d'après (46) et (62)

$$\mathcal{L}_2 = \rho/\sigma = \delta \sqrt[3]{\sum_j a_j^{12}} / \sqrt{\sum_j a_j^8} \quad \delta / \sqrt[6]{l} \leq \mathcal{L}_2 \leq \delta \quad (64)$$

(Q.29) Une bonne illustration de cette circonstance est donnée par l'analyse de la série s (Section S ci-dessous).

(Q.30) Voir ci-dessous § Q.12.

(Q.31) Cette proposition se démontre comme au § Q.4. Lemme III.

Il résulte de là encore que si les a_j ne sont pas trop différents de leur moyenne, la distribution des q_j^* sera approximativement normale et leurs valeurs seront relativement faibles.

Q.12. *Cas de l'égalité de tous les a_j*

Dans le cas où tous les a_j sont égaux à leur moyenne a , on peut prendre $a = 1$ sans aucune perte de généralité, et on a d'après (9), (56) et (62)

$$X(n) = \sum_j \cos 2\pi f_j (n - n_j) \quad (65)$$

$$q_q^* = \frac{1}{l} \sum_j \cos 2\pi q f_j \quad q_r^* = \frac{1}{l} \sum_j \cos 2\pi r f_j \quad (66)$$

Les coefficients de Liapounoff \mathcal{L} , \mathcal{L}_1 et \mathcal{L}_2 relatifs à ces trois séries ont pour expression d'après (46), (57) et (64)

$$\mathcal{L} = \mathcal{L}_1 = \mathcal{L}_2 = \delta/\sqrt[6]{l} \quad \delta = 2^{7/6}/\sqrt[3]{3\pi} = 1,0628 \quad (67)$$

Il résulte des relations (66) que les distributions de q_q^* ($q_1 \leq q \leq q_2$) et de q_r^* ($r_1 \leq r \leq r_2$) sont identiques pour $r_1 = q_1$, $r_2 = q_2$.

Il est visible que les quantités $X(n)$, q_q^* , q_r^* ont des valeurs d'autant plus faibles que les parties décimales des fréquences sont plus uniformément réparties dans l'intervalle (0, 1) ⁽³²⁾.

D'après (44), (59) et (63) on a ici pour les *valeurs asymptotiques des écarts type*

$$\sigma_x^* = \sqrt{l/2} \quad \sigma(q_q^*) = 1/\sqrt{2l} \quad \sigma(q_r^*) = 1/\sqrt{2l} \quad (68)$$

Lorsque N et l augmentent indéfiniment les valeurs moyennes de $X(n)$, q_q^* et q_r^* tendent vers zéro.

Cas de l'équipartition de l'énergie

Le cas de l'égalité des a_j peut être considéré comme susceptible de résulter du principe physique de l'équipartition de l'énergie.

Supposons en effet que l'on ait

$$x_j = \frac{d}{dt} \left[b_j \sin \frac{2\pi}{T_j} (t - t_j) \right] = a_j \cos \frac{2\pi}{T_j} (t - t_j) \quad (69)$$

en posant $a_j = 2\pi b_j/T_j$

S'il s'agit d'un système physique, il résulte de là que l'énergie moyenne E correspondant au système des x_j a pour expression

$$E = \sum_j a_j^2/2 \quad (70)$$

et dans ce cas l'équipartition de l'énergie implique que tous les a_j soient égaux.

Q.13. *La convergence empirique vers la loi normale de la distribution des $X(n)$*

Le *Théorème (T)* ci-dessus de convergence vers la loi normale de la distribution d'une somme $X(n)$ ($1 \leq n \leq N$) de l sinusoides pour des valeurs infinies de N et l sous les *Hypothèses (H)* n'est qu'un

(Q.32) Pour $q = r = 0$ on a $q_q^* = q_r^* = 1$. Il résulte de là que les premières valeurs de q_q^* et q_r^* ne pourront avoir des valeurs petites que si la plupart des périodes T_j n'ont pas de grandes valeurs relativement à l'intervalle séparant deux valeurs consécutives de $X(n)$, intervalle égal à l'unité.

théorème asymptotique qui ne précise pas les conditions d'une plus ou moins grande rapidité de la convergence vers la loi normale.

En fait, seul le calcul numérique peut indiquer pour quelles classes de valeurs des N , l , a_j , T_j les $X(n)$ ont effectivement une distribution *voisine* de la loi normale ⁽³³⁾.

En tout état de cause la condition de Liapounoff *n'est qu'une condition suffisante*, et ce n'est pas une condition nécessaire de cette convergence, de sorte qu'il est parfaitement possible qu'une distribution normale soit approximativement assurée sans que l prenne une valeur extrêmement grande ⁽³⁴⁾.

Par ailleurs il convient que le nombre N de valeurs considérées prenne des valeurs d'autant plus élevées que les périodes T_j sont elles-mêmes plus grandes pour que les conditions du *Lemme II* soient approximativement satisfaites ⁽³⁵⁾ ⁽³⁶⁾.

De toute façon, pour tous les cas susceptibles d'être analysés, les nombres N et l sont finis. La distribution des $X(n)$ ne peut donc qu'être *approximativement* normale ⁽³⁷⁾ ⁽³⁸⁾.

Pour qu'en outre la suite des $X(n)$ ait l'apparence d'une suite de valeurs *indépendantes* il est nécessaire que la plus grande partie des périodes T_j n'aient pas des valeurs d'un ordre de grandeur très supérieur à l'unité ⁽³⁹⁾.

Les trois illustrations présentées dans les Sections S, T et U ci-dessous ⁽⁴⁰⁾ ne constituent que trois illustrations parmi bien d'autres. Faute de place je ne peux présenter ici l'analyse d'autres cas, si suggestive qu'elle pourrait être.

En tout état de cause le calcul numérique confirme que la distribution approximative d'une somme de sinusoides suivant la loi normale est réalisée dans des circonstances *extrêmement générales*.

(Q.33) Voir les indications déjà données dans le § P 7 ci dessus. De toute façon, les calculs numériques effectués ne peuvent comporter une précision supérieure à celle des ordinateurs. Il résulte en particulier de là que dans les calculs numériques les conditions (B) ne peuvent être satisfaites qu'approximativement

En fait, l'hypothèse (B) du § Q.3 ne peut être satisfaite que pour des valeurs irrationnelles des fréquences f_j . Mais tout nombre irrationnel peut être approximé par un nombre rationnel avec une précision aussi grande que l'on désire. Ainsi le nombre irrationnel $e = 2,718.281.828 \dots$ peut être approximé par le nombre rationnel 2,718.281.828 avec une erreur inférieure à 10^{-9} .

(Q.34) Tel est notamment le cas de la série S ci dessous analysée dans la Section S $X(n)$ est la somme de *seulement* 13 sinusoides et on a

$$\sqrt[6]{l} = \sqrt[6]{13} = 1,53$$

(Q.35) Si en effet la grandeur des T_j n'intervient pas dans la démonstration du *Théorème (T)* (§ Q.3 à Q.5) ci dessus, par contre, il est supposé que N augmente indéfiniment, ce qui implique que pour N donné, les T_j doivent être en général relativement petits pour que le *Théorème (T)* soit approximativement vérifié

En outre la représentation par la loi normale paraît d'autant meilleure que la répartition des parties fractionnaires des fréquences est d'autant plus voisine d'une répartition uniforme. C'est là une condition tout à fait essentielle dès que le nombre de sinusoides est relativement faible

(Q.36) Comme on a nécessairement

$$|X_n| \leq A = \sum_j a_j$$

les valeurs très grandes de $X(n)$ pour des valeurs données des a_j sont impossibles.

En fait, comme les périodes T_j sont incommensurables, il est tout à fait exceptionnel que les sinusoides composantes $x_j(n)$ soient toutes en phase Il en résulte qu'en général $|X(n)|$ est très inférieure à A .

On se trouve ici dans une situation très comparable à celle étudiée par POINCARÉ pour les longitudes des planètes (POINCARÉ, 1906, p. 231-233; et 1912, § 93, p. 150-152). La fréquence d'alignements simultanés de toutes les planètes avec le soleil est presque certainement nulle.

(Q.37) Comme $|X(n)| \leq A = \sum a_j$, on peut s'attendre *a priori* à ce que la représentation de la distribution des $X(n)$ par la loi normale soit moins bonne pour les grandes valeurs de $X(n)$. C'est ce que l'on constate effectivement *pour les calculs effectués* pour $N \leq 1.000$ dès que $|X(n)|$ dépasse $2\sigma_x$.

(Q.38) En tout cas, il semble bien que N doit prendre des valeurs très supérieures à l , et que le rapport N/l doit prendre des valeurs d'autant plus grandes que les T_j ont des valeurs plus élevées relativement à l'intervalle pris ici égal à l'unité séparant deux valeurs consécutives de $X(n)$ (voir la note Q.32 ci-dessus)

(Q.39) Cette condition résulte de l'expression (56) ci dessus de ρ_q^* . Il est en effet visible que si la plupart des T_j ont des valeurs très supérieures à l'unité les premiers coefficients de corrélation ρ_1^* , ρ_2^* , ... ne peuvent avoir de petites valeurs.

Voir à titre d'illustration la distribution des valeurs des T_j (et des valeurs équivalentes θ_j) de la série U (§ U 1 ci-dessous).

(Q.40) Voir notamment les Graphiques S.2, T.2, T.5, U.2 et U.5 ci dessous. Voir également les Tableaux comparatifs R.1 et R.2 des § R.1 c et R.1 g ci dessous.

Q.14. *Généralisations du Théorème (T)*

Le *Théorème (T)* apparaît susceptible de nombreuses généralisations, dont notamment les suivantes :

a) *Extension au cas continu*

La démonstration effectuée dans le cas discontinu peut être étendue au cas continu où l'on a

$$X(t) = \sum_{j=1}^{j=N} a_j \cos 2\pi f_j (t - t_j) \quad (-\infty < t < +\infty) \quad (71)$$

Dans ce cas l'*Hypothèse (B)* doit être remplacée par l'hypothèse (B*) : « Il n'existe aucune relation linéaire $\sum_j \alpha_j f_j = 0$ à coefficients entiers α_j » (41).

b) *Extension au cas où les composantes $x_j(n)$ de $X(n)$ sont des fonctions périodiques*

Dans ce cas l'indépendance des $x_j(n)$ résulte de l'indépendance des

$$u_j(n) = [nf_j]_d \quad (72)$$

où la notation $[]_d$ représente la partie décimale de nf_j (42). Il en résulte que *la démonstration du Théorème (T) peut être étendue à des sommes de fonctions périodiques.*

c) *Extension à des fonctions presque périodiques au sens de H. Bohr*

Si certaines périodes sont incommensurables la série

$$X(t) = \sum_{j=1}^{\infty} a_j \cos 2\pi f_j (t - t_j) \quad \left(a_j > 0 ; \sum_{j=1}^{\infty} a_j \leq A \right) \quad (73)$$

uniformément et absolument convergente, est une fonction presque périodique. Mais elle n'est qu'un cas particulier (celui considéré par P. Bohl et E. Esclangon) des fonctions presque périodiques au sens de H. Bohr (43). Comme ces fonctions peuvent être approximées avec une précision arbitrairement grande par des séries du type (73), le *Théorème (T)* peut être étendu au cas des fonctions presque périodiques au sens de H. Bohr.

d) *Extension à des sommes de fonctions du temps indépendantes*

Le *Théorème (T)* s'étend également à des sommes

$$X(n) = \sum_{j=1}^l g_j(n) \quad (1 \leq n \leq N) \quad (74)$$

sous les conditions que les fonctions $g_j(n)$ soient asymptotiquement indépendantes et que la condition de Liapounoff soit satisfaite. Il en est naturellement encore de même sous les mêmes conditions pour la fonction continue du temps

$$X(t) = \sum_{j=1}^l g_j(t) \quad (-\infty < t < +\infty) \quad (75)$$

(Q.41) Si on prend comme unité de mesure un intervalle de temps quelconque, la relation (B) $\sum_j \alpha_j f_j = \beta$ où les α_j et β sont entiers s'écrit (B') $\sum_j \alpha_j f'_j = \beta \Delta$ avec $f'_j = f_j \Delta$, où Δ représente l'intervalle de temps entre les instants n et $n + 1$. Lorsque Δ tend vers zéro, la relation (B') devient (B*) $\sum_j \alpha_j f'_j = 0$. C'est là la condition d'« indépendance arithmétique » considérée dans la théorie des fonctions presque périodiques où la variable temps varie de manière continue (voir ESCLANGON, 1904, p. 70; FAVARD, 1933, p. 44-45; CORDUNEANU, 1968, p. 40-41; HARDY et WRIGHT, 1959, p. 381).

(Q.42) Voir la note (Q.19) ci-dessus.

(Q.43) Voir § P.11 ci-dessus.

Q.15. *Cas où certaines fréquences sont commensurables*

Dans le cas où il existe des entiers α_j et β positifs ou nuls tels que $\sum_j \alpha_j f_j = \beta$, mais où les rapports α_j/N et β/N ont des valeurs très supérieures à l'unité, les nombres N et l ayant des valeurs finies mais suffisamment grandes, et le coefficient \mathcal{L} de Liapounoff ayant une valeur suffisamment petite, il est visible que la distribution des $X(n)$ sera approximativement normale.

Q.16. *Théorème d'équivalence*

On a d'après la relation (9) ci-dessus

$$x_j(n) = a_j \cos 2\pi f_j(n - n_j) \quad (76)$$

Soit G un nombre fractionnaire ou irrationnel et posons

$$[G]_d = \text{partie décimale de } G \quad (77)$$

Théorème (44)

Si tous les n_j sont des nombres entiers, les fonctions $x_j(n)$, et toutes les grandeurs que l'on peut en déduire, restent invariantes par le remplacement des fréquences f_j par les fréquences φ_j telles que

$$\varphi_j = [f_j]_d + k(j) \quad (78)$$

où $k(j)$ est une fonction de j qui a une valeur entière positive ou négative.

Cette propriété résulte immédiatement de ce que m étant entier on a nécessairement (45)

$$m [f_j]_d = [mf_j]_d \quad (79)$$

Puisque par hypothèse $n - n_j$ est entier, on a alors d'après (79)

$$\begin{aligned} \cos 2\pi f_j(n - n_j) &= \cos 2\pi [f_j(n - n_j)]_d = \cos 2\pi [[f_j]_d(n - n_j)]_d \\ &= \cos 2\pi [[\varphi_j]_d(n - n_j)]_d = \cos 2\pi [\varphi_j(n - n_j)]_d \end{aligned} \quad (80)$$

d'où finalement

$$x_j(n) = a_j \cos 2\pi \varphi_j(n - n_j) \quad (81)$$

Il résulte de ce théorème que si les n_j ont des valeurs entières il y a une infinité de systèmes de fréquences, et donc de périodes, pour lesquelles les $x_j(n)$ gardent les mêmes valeurs, et donc également les $X(n) = \sum_j x_j(n)$

(Q 44) Une application de ce Théorème sera faite dans le § U.1 ci dessous.

(Q.45) Ainsi pour

$$f_j = \alpha \sqrt{j} \quad \alpha = \sqrt[4]{23} \quad j = 17 \quad m = 151$$

(§ U.1 ci dessous) on a

$$\begin{aligned} [f_j]_d &= [9,029348]_d = 0,029348 \\ [m[f_j]_d]_d &= [4,431636]_d = 0,431636 \\ [mf_j]_d &= [1363,431636]_d = 0,431636 \end{aligned}$$

R. TESTS DE NORMALITÉ, D'INDÉPENDANCE, D'ARÉGULARITÉ LOCALE,
ET DES VALEURS CUMULÉES

La présente Section présente les tests dont il est fait application dans les Sections S, T et U ci-dessous, soit pour illustrer la validité approximative du *Théorème (T)* pour des séries comportant des nombres *finis* N et l de termes et de sinusoides, soit pour tester l'indépendance apparente des valeurs successives $X(n)$ des séries considérées *pour quiconque ignorerait leur structure effective* ⁽¹⁾.

R.1. *Test de normalité : Ajustement de Henri suivant la loi normale*

L'ajustement des variables considérées suivant la loi normale a été effectué selon les principes suivants :

a) *Répartition des valeurs observées en $r + 1$ classes*

Pour ajuster les valeurs observées d'une variable u suivant la loi normale on répartit les N valeurs considérées dans $r + 1$ intervalles dont les r limites sont régulièrement espacées.

$$\begin{array}{ccccccc} & n_0 & n_1 & & n_{r-1} & n_r & \\ & | & | & & | & | & \\ -\infty & & u_1 & u_2 & & u_{r-1} & u_r & +\infty \end{array}$$

En principe les limites des intervalles ont été choisis de manière qu'il y ait *symétrie* par rapport à la valeur 0, ce qui correspond à un nombre impair de classes et à un nombre pair de points. En outre on a fait en sorte que dans chacun des deux intervalles extrêmes (symétriques ainsi qu'il vient d'être indiqué) il y ait un nombre de points de l'ordre d'au moins 10 à 15 ⁽²⁾.

Soient $n_0, n_1, \dots, n_s, \dots, n_r$ les nombres de valeurs u dans chacun des $r + 1$ intervalles considérés. On calcule les fréquences cumulées

$$F_s = F(u_s) = 1 - \sum_{i=0}^s (n_i/N) \quad (1)$$

n_0 correspond au nombre de valeurs u situées dans l'intervalle $(-\infty, u_1)$.

b) *Graphique de Henri*

On utilise le *Graphique de Henri* ⁽³⁾ dont l'échelle des abscisses est linéaire et dont l'échelle des ordonnées représente linéairement les valeurs de z de la loi normale

$$F^*(z) = \frac{1}{\sqrt{2\pi}} \int_z^{+\infty} e^{-\frac{z^2}{2}} dz \quad (2)$$

La graduation des ordonnées représente les valeurs de $F^*(z)$ en pourcents.

Pour chaque abscisse u_s on porte en ordonnée la fréquence correspondante F_s . Pour les $r+1$ classes on a r points représentatifs. Si la distribution était rigoureusement normale, les points $M_s(u_s, F_s)$ se répartiraient sur une même droite. Comme elle ne l'est pas on calcule la corrélation

$$z_s = a u_s + b + \varepsilon_s \quad (3)$$

(R.1) § P.6 ci dessus. Les tests ci dessous exposés sont utilisés pour l'analyse des séries S, T et U (Sections S, T et U ci-dessous).

(R.2) Pour l'application du test χ^2 Cramer (1946, p. 359) indique un nombre minimum de valeurs de 10. Le seul graphique de Henri pour lequel la condition de symétrie n'a pas été réalisée est le Graphique S.2 ci dessous. Il nous a paru inutile de refaire les calculs pour satisfaire à cette condition.

(R.3) Sur le Graphique de Henri et son utilisation, voir notamment HALD, 1951, § 6.6-6.7, p. 127 140; MORICE et CHARTIER, 1954, vol. I, p. 105 107; et ALLAIS, 1954, Annexe VII.

Sur les applications de cette méthode d'ajustement, voir Sections S, T et U ci-dessous, Graphiques S.2, T.2., T.5, U.2 et U.5.

La droite d'ajustement est

$$z^* = (u - \bar{u})/\sigma \quad G^* = G(z^*) \quad (4)$$

$$\sigma = 1/a \quad \bar{u} = -b/a \quad (5)$$

\bar{u} et σ représentent respectivement les estimations de la valeur moyenne \bar{u} et de l'écart type σ par cette méthode graphique. La médiane u_m des valeurs observées de \bar{u} est déterminée à partir des N valeurs u considérées ⁽⁴⁾.

c) Coefficient $1-R^2$ et écarts relatifs des fréquences cumulées

R représentant le coefficient de la corrélation (z, z^*), le coefficient $1-R^2$ représente la fraction de la variance de z « non expliquée » par la corrélation. Si l'ajustement est bon, la variance de z est peu différente de la variance de z^* qui d'après les propriétés de la distribution normale est égale à l'unité.

Les écarts relatifs des fréquences cumulées peuvent se caractériser par le coefficient

$$E^2 = \frac{1}{H} \sum_i H_i \left(\frac{F_i - F_i^*}{F_i^*} \right)^2 \quad (6)$$

avec ⁽⁵⁾

$$\begin{cases} H_s = 1 - F_s^* & \text{pour } F_s^* > 0,5 \\ H_s = F_s^* & \text{pour } F_s^* < 0,5 \\ H = \sum_{s=1}^r H_s \end{cases} \quad (7)$$

Le Tableau R.1 ci-contre donne les valeurs comparées de $1-R^2$, $\sqrt{1-R^2}$, E^2 et E , pour les ajustements S, T, et U présentés dans les Sections S, T, et U ci-dessous. A titre indicatif les caractéristiques de l'ajustement relatif à la série U prolongée jusqu'à $N = 3.000$ sont indiquées.

Les valeurs de $1-R^2$ sont très faibles, ce qui correspond au fait que les valeurs des z_s s'écartent relativement peu de leurs valeurs théoriques dans l'hypothèse de normalité.

Il convient cependant de souligner que l'hypothèse de normalité pour une réalisation particulière doit « normalement » comporter des écarts par rapport à la droite de Henri ⁽⁶⁾.

On constate enfin que pour $N = 3.000$ la distribution des $X(n)$ est plus voisine de la distribution normale que pour $N = 1.000$.

d) Chenal de confiance à 95 %

Comme on a

$$\frac{1}{\sqrt{2\pi}} \int_{-z_1}^{+z_1} e^{-\frac{z^2}{2}} dz = 0,95 \quad \text{pour } z_1 = 1,9600 \quad (8)$$

(R.4) Dans le cas où les points se répartissent approximativement sur une même droite on peut sans aucun calcul déterminer des valeurs approximatives de la moyenne \bar{u} et de l'écart type σ à partir du tracé à vue de la droite interpolant « au mieux » les points représentatifs des fréquences cumulées. La moyenne \bar{u} correspond en effet à l'intersection de la droite d'ajustement avec l'horizontale correspondant à $G = 0,5$ et l'écart type se déduit immédiatement des conditions

$$F^* (-1) \sim 0,841 \quad F^* (1) \sim 0,159$$

(R.5) Il convient de souligner que si l'on considère les $1 - F_i^*$ au lieu des F_i^* la droite de Henri est remplacée par sa symétrique relativement à la verticale correspondant à la valeur moyenne \bar{u}

Il résulte de là qu'il convient de pondérer les carrés des écarts relatifs $[(F_i - F_i^*)/F_i^*]^2$ par les pourcents H_i correspondants, égaux suivant le cas à F_i^* ou à $1 - F_i^*$, ainsi qu'il est indiqué dans les relations (7).

(R.6) Voir ci-dessous § R.1 e.

le chenal de confiance à 95 % pour une fréquence F^* est défini d'après la distribution binomiale par les valeurs

$$F^*(u) \pm \Delta F^*(u) \quad (9)$$

avec

$$\Delta F^* = 1,960 \sqrt{F^*(1-F^*)/N} \quad (10)$$

Pour chaque valeur de u , on porte les points correspondant aux valeurs $F^* - \Delta F^*$ et $F^* + \Delta F^*$. Ces points définissent les limites du chenal.

TABLEAU R.1

SERIES S, T ET U

Tableau comparatif des coefficients caractéristiques

$$\sqrt{1-R^2} \text{ et } E$$

relatifs aux fréquences cumulées

correspondant aux ajustements de Henri pour les séries S, T, et U ⁽¹⁾

Séries	S	T		U		U ₁ ⁽²⁾	
	X _n	X _n	e _n [*]	X _n	e _n [*]	X _n	e _n [*]
Graphiques	S.2	T.2	T.5	U.2	U.5	U.16	U.17
l	13	20	20	200	200	200	200
N	721	1 000	1 000	1 000	1 000	3 000	3 000
Limites de l'ajustement	98,5 %	98,4 %	98,4 %	97,7 %	98,6 %	99 %	99,6 %
	4 % ⁽²⁾	1,6 %	1,6 %	2,3 %	1,4 %	1 %	0,4 %
1 - R ²	8,9 · 10 ⁻⁴	3,4 · 10 ⁻⁴	10 · 10 ⁻⁴	9,4 · 10 ⁻⁴	8,7 · 10 ⁻⁴	2,4 · 10 ⁻⁴	6,8 · 10 ⁻⁴
E ²	9,8 · 10 ⁻⁴	11 · 10 ⁻⁴	6,8 · 10 ⁻⁴	17 · 10 ⁻⁴	11 · 10 ⁻⁴	2,0 · 10 ⁻⁴	17 · 10 ⁻⁴
(1 - R ²)/E ²	0,91	0,30	1,5	0,57	0,81	1,2	0,39
$\sqrt{1-R^2}$	3,0 · 10 ⁻²	1,8 · 10 ⁻²	3,2 · 10 ⁻²	3,1 · 10 ⁻²	2,9 · 10 ⁻²	1,5 · 10 ⁻²	2,6 · 10 ⁻²
E	3,1 · 10 ⁻²	3,4 · 10 ⁻²	2,6 · 10 ⁻²	4,1 · 10 ⁻²	3,3 · 10 ⁻²	1,4 · 10 ⁻²	4,2 · 10 ⁻²

1. Distribution des

$$X(n) = \sum_{j=1}^l x_j(n) = \sum_{j=1}^l a_j \cos 2\pi f_j(n - n_j)$$

$$1 \leq j \leq l \quad 1 \leq n \leq N$$

Voir § P.3, P.5 et P.6 ci-dessus.

2. Voir la note (R.2) ci-dessus.

3. La série U₁ a la même structure que la série U, mais elle est considérée pour 1 ≤ n ≤ 3.000 alors que pour la série U on a : 1 ≤ n ≤ 1.000. Faute de place les Graphiques U 16 et U 17 ne sont pas présentés dans la Section U ci-dessous.

e) Test χ^2

La valeur théorique n_s^* correspondant à l'intervalle (u_s, u_{s+1}) est

$$n_s^* = N f_s^* = N [F^*(u_s) - F^*(u_{s+1})] \quad (11)$$

On calcule alors

$$\chi^2 = \sum_{s=0}^{s=r} (n_s - N f_s^*)^2 / N f_s^* \quad (12)$$

La valeur u_0 est $-\infty$ et la valeur u_{r+1} est $+\infty$. On a

$$F^*(u_0) = 1 \quad F^*(u_{r+1}) = 0 \quad (13)$$

Comme les $n_s^* = N f_s^*$ ont été calculés à partir de la droite de Henri le nombre de degrés de liberté est

$$v = (r + 1) - 1 - 2 = r - 2 \quad (14)$$

le nombre de classes étant égal à $r + 1$.

La valeur de la probabilité P pour que χ^2 soit supérieur à la valeur calculée a été prise dans les *Tables de Hartley et Pearson* (1).

On a ici

$$\text{moyenne } \chi^2 = v = r - 2 \quad \text{variance } \chi^2 = 2v = 2(r - 2) \quad (15)$$

l'égalité ne valant naturellement que pour la distribution asymptotique de χ^2 (8). On en déduit

$$\text{variance } [(\chi^2/v) - 1] = 2/v \quad (16)$$

Si seul le « hasard » intervient on doit donc s'attendre non pas à ce que le rapport χ^2/v ait une valeur très faible, mais à ce que ce rapport ait une valeur se situant au voisinage de l'unité. Le coefficient

$$\lambda = [(\chi^2/v) - 1] / \sqrt{2/v} \quad (17)$$

caractérise l'écart relatif de χ^2/v par rapport à sa valeur moyenne.

A tort les manuels ne considèrent que la probabilité P pour que χ^2 ait une valeur supérieure à une valeur donnée. *Il faut également considérer* la probabilité P' pour que χ^2 ait une valeur inférieure à une valeur donnée. En fait les très faibles valeurs du rapport χ^2/v sont tout aussi « invraisemblables » que les très grandes valeurs de ce rapport.

f) Écarts relatifs $(n_s - n_s^*)/n_s^*$

On a pour la moyenne, pondérée suivant les fréquences théoriques, des carrés des écarts relatifs $(n_s - n_s^*)/n_s^*$.

$$e^2 = \sum_{s=0}^{s=r} f_s^* \left(\frac{n_s - n_s^*}{n_s^*} \right)^2 = \sum_{s=0}^{s=r} f_s^* \left(\frac{f_s - f_s^*}{f_s^*} \right)^2 \quad (18)$$

soit d'après (12)

$$e^2 = \chi^2 / N \quad (19)$$

(R.7) HARTLEY et PEARSON, *New Tables of Statistical Functions, Biometrika*, vol. 37, décembre 1950, p. 313-323.

(R.8) CRAMER, 1946, p. 417 (voir également p. 234 et 427), et HARTLEY et PEARSON, *id.*, p. 313.

Les valeurs médianes de χ^2 pour $v = 6, 7, 10, 14$ et 16 sont respectivement $5,45; 6,30; 9,65; 13,6; 15,4$ (N. B. OWEN, *Handbook of Statistical Tables*, Addison Wesley, 1962, p. 50-51).

Pour la distribution de χ^2/v , voir HALD, 1952, *Statistical Tables*, p. 44-46.

g) *Tableau comparatif des ajustements*

Le Tableau R.2 ci-contre présente les valeurs des coefficients caractéristiques des différents ajustements obtenus par la méthode de Henri correspondant aux séries S, T et U analysées ci-dessous dans les Sections S, T et U.

TABLEAU R.2

SERIES S, T ET U
Tableau comparatif des coefficients caractéristiques
 χ^2/ν et e
relatifs aux fréquences dans chaque classe
correspondant aux ajustements de Henri pour les séries S, T, et U (1)

Séries	S	T		U		U ₁	
	X_n	X_n	e_q^*	X_n	e_q^*	X_n	e_q^*
Graphiques	S. 2	T. 2	T. 5	U. 2	U. 5	U. 16	U. 17
l	13	20	20	200	200	200	200
N	721	1 000	1 000	1 000	1 000	3 000	3 000
ν	7	6	14	16	10	14	12
χ^2	6,90	5,37	13,2	9,70	4,94	11,2	13,6
P	0,44	0,49	0,51	0,88	0,89	0,67	0,33
χ^2/ν	0,986	0,895	0,943	0,606	0,494	0,804	1,13
$\lambda = \left(\frac{\chi^2}{\nu} - 1\right) / \sqrt{\frac{2}{\nu}}$	-0,026	-0,182	-0,151	-1,11	-1,13	-0,520	0,35
$e^2 = \chi^2/N$	$9,57 \cdot 10^{-3}$	$5,37 \cdot 10^{-3}$	$13,2 \cdot 10^{-3}$	$9,70 \cdot 10^{-3}$	$4,94 \cdot 10^{-3}$	$11,2 \cdot 10^{-3}$	$4,5 \cdot 10^{-3}$
e	$9,78 \cdot 10^{-2}$	$7,33 \cdot 10^{-2}$	$11,5 \cdot 10^{-2}$	$9,85 \cdot 10^{-2}$	$7,02 \cdot 10^{-2}$	$10,6 \cdot 10^{-2}$	$6,7 \cdot 10^{-2}$

1. Voir les notes 1 et 3 du tableau R.1.

Il résulte de ce Tableau R.2 que pour les sept ajustements considérés les valeurs de P et λ correspondent à ce que l'on peut « normalement » attendre dans le cas de distributions normales.

R.2. Test d'absence de toute autocorrélation — Corrélogramme

L'ajustement d'une variable u suivant la loi normale ne tient pas compte de l'ordre dans lequel se succèdent les différentes valeurs u considérées.

En général les valeurs successives $X(n)$ des sommes de l sinusoides, des coefficients d'autocorrélation asymptotiques e_q^* , et des coefficients d'autocorrélation r_q sont autocorrélées, mais, dans certains

cas, les séries des $X(n)$ et r_q peuvent apparaître, *pour tout analyste ignorant de leur structure réelle*, comme dénuées de toute autocorrélation ⁽⁹⁾.

Pour tester l'absence apparente de toute autocorrélation, il faut vérifier que les coefficients d'autocorrélation r_q ont bien l'ordre de grandeur correspondant à des suites dénuées de toute autocorrélation, et il faut vérifier qu'il en est bien ainsi *pour le nombre N considéré de valeurs consécutives de la série étudiée*.

On sait que la variance du coefficient d'autocorrélation r_q entre les valeurs u_n et u_{n+q} d'une série de N couples de termes u_n indépendants et normalement distribués est pour une valeur de N suffisamment élevée ⁽¹⁰⁾

$$\text{var. } r_q \sim 1/N \qquad \sigma(r_q) \sim 1/\sqrt{N} \qquad (N \gg 1) \qquad (20)$$

En fait, ce cas se réduit à celui de la corrélation entre deux séries indépendantes de longueur N , et les Tables de la distribution du coefficient de corrélation entre deux séries normalement distribuées et indépendantes s'appliquent.

Pour tester l'absence apparente d'autocorrélation, on peut donc utiliser les *Tables de David* ⁽¹¹⁾. En utilisant ces Tables on constate qu'à partir de $N = 50$ la distribution du coefficient de corrélation r , lorsque la véritable valeur ρ est nulle, est presque parfaitement normale ⁽¹²⁾.

a) Correction à apporter en raison des modalités de calcul du corrélogramme

Si on calcule les Q premiers coefficients de corrélation d'une série u_n comportant N valeurs, le premier coefficient de corrélation correspond à $N - 1$ couples de valeurs; mais le $q^{\text{ème}}$ coefficient de corrélation correspond à $N - Q$ couples de valeurs. La valeur moyenne du nombre de couples de valeurs considérées est ainsi

$$\bar{N} = [N - 1 + N - Q]/2 = N - (1 + Q)/2 \qquad (21)$$

et on a pour l'écart type théorique des Q premiers coefficients d'autocorrélation r_q

$$\sigma_r \sim 1/\sqrt{N - (Q/2)} \qquad (N \gg 1) \qquad (22)$$

dans l'hypothèse où la série considérée r_q est constituée de termes indépendants.

b) Intervalle de confiance à 95 % du $q^{\text{ème}}$ coefficient d'autocorrélation r_q d'une série de N valeurs supposées indépendantes

Comme pour un nombre suffisamment élevé de couples de valeurs la distribution du coefficient d'autocorrélation est approximativement normale, l'intervalle de confiance à 95 % du $q^{\text{ème}}$ coefficient d'autocorrélation d'une série de N valeurs supposées indépendantes est approximativement (d'après la Table des valeurs de la loi normale)

$$-1,96 \sigma_q \leq r_q \leq 1,96 \sigma_q \quad \text{pour} \quad N - q \geq 50 \qquad (23)$$

avec (relation 20 ci-dessus)

$$\sigma_q = 1/\sqrt{N - q} \qquad (24)$$

(R.9) Voir ci dessus § P 4, Q.10, Q 11, et ci dessous Section U

(R.10) KENDALL et STUART, 1976, tome III, § 48,2, p. 452 (relation 48.10).

(R.11) DAVID, *Tables of the Correlation Coefficient*, The University Press, Cambridge, England, 1938.

(R.12) Ainsi qu'on peut le vérifier en représentant par des droites de Henri (§ R.1 c ci dessus) les distributions données par David (p. 48 55). Les ajustements sont excellents. On trouve pour $N = 50 : 1 - R^2 = 1,63 \cdot 10^{-4}$; pour $N = 100 : 1 - R^2 = 5,3 \cdot 10^{-5}$; pour $N = 200 : 1 - R^2 = 5 \cdot 10^{-6}$; pour $N = 400 : 1 - R^2 = 3 \cdot 10^{-6}$. Les valeurs de σ , déduites du graphique de Henri sont respectivement $1/1,0036 \sqrt{50}$, $1/1,0042 \sqrt{100}$, $1/1,00022 \sqrt{200}$, $1/1,00031 \sqrt{400}$.

On retrouve donc bien les valeurs théoriques de σ , à un très faible écart près.

R.3. Test d'arégularité locale pour une série de valeurs indépendantes :
Distribution des nombres de suites de valeurs positives et négatives

La considération du corrélogramme d'une série de N valeurs u_n ne suffit pas par elle-même pour conclure que cette série peut être considérée comme résultant du hasard ou comme le simulatant. Ce corrélogramme ne constitue qu'un moyen de tester l'indépendance des valeurs successives de X_n , c'est-à-dire de tester l'absence de toute régularité significative dans la suite des X_n .

En fait il existe un grand nombre de tests d'arégularité locale ⁽¹³⁾ et on peut d'ailleurs en imaginer une foule d'autres. Le test que j'utiliserai est le suivant.

S'il n'existe aucune corrélation significative entre les valeurs successives des X_n (ou des q_q^* et des r_q) on doit constater que les longueurs des suites de valeurs positives et négatives des X_n (ou des q_q^* et des r_q) se distribuent suivant le modèle fréquentiel correspondant à la distribution binomiale correspondant à une fréquence $1/2$ ^(13').

Considérons une suite de n tirages dans une urne contenant un nombre égal de boules blanches et noires. Il y a 2^n trajectoires. On considère les suites de boules blanches et noires, et on considère la distribution de leurs longueurs. Soit q_{sn} le nombre de suites de s boules blanches ou de s boules noires, et soit

$$Q_n = \sum_{s=1}^n q_{sn} \quad (25)$$

Pour $n = 3$ par exemple, on a $2^3 = 8$ cas également possibles : $B B B$; $B B N$; $B N B$; $B N N$; $N B B$; $N B N$; $N N B$; $N N N$. On a ainsi $q_{13} = 10$; $q_{23} = 4$; $q_{33} = 2$; $Q_3 = 16$.

D'une manière générale j'ai démontré par récurrence que ⁽¹⁴⁾

$$q_{sn} = (n + 3 - s) 2^{n-s-1} \quad \text{pour} \quad 1 \leq s \leq n - 1 \quad (26)$$

$$q_{nn} = 2 \quad \text{pour} \quad s = n \quad (27)$$

$$Q_n = (n + 1) 2^{n-1} \quad Q_{sn} = \sum_{s=1}^n q_{sn} = (n - s + 2) 2^{n-s} \quad (28)$$

Il convient de souligner que la formule (26) ne s'applique que pour $s \leq n - 1$. Elle ne s'applique pas pour $s = n$.

On a naturellement

$$\sum_{s=1}^n s q_{sn} = n 2^n \quad (29)$$

(R.13) Sur les tests d'arégularité locale, voir KENDALL et SMITH, 1938; et KENDALL et STUART, 1976, vol. III, § 45.15, p. 364-375. A ma connaissance le test ci dessus exposé n'a jamais été présenté dans la littérature. Le test qui en est le plus proche est celui du « gap test » de KENDALL et SMITH, 1938, *id.*, p. 154 et 159. Cependant ce test n'est rigoureusement valable qu'asymptotiquement alors que celui qui est utilisé ici vaut rigoureusement pour toute valeur de n . En tout état de cause Kendall et Smith font une application défectueuse de leur test en calculant les fréquences théoriques d'après le nombre total de « gaps » observé au lieu de les calculer à partir de la valeur moyenne théorique de ce nombre (*id.* p. 159).

(R.13') ALLAIS, 1983 d, § A2, p. 95-97.

(R.14) ALLAIS, 1978.

Dans le cas général où la fréquence est $f = 1/j$ j'ai démontré par récurrence que :

— pour $1 \leq s \leq n - 1$:

$$q_{sn} = \{(j - 1) [(n - s)(j - 1) + j + 1] + (n - s + 2j - 1)(j - 1)^s\} j^{n-s-2} \quad (1)$$

— pour $s = n$:

$$q_{nn} = 1 + (j - 1)^n \quad (2)$$

Ce cas est celui d'une urne contenant k boules blanches et $k(j - 1)$ boules noires. Les suites de B et N restent définies comme ci-dessus. Les expressions (1) et (2) généralisent les expressions (26) et (27); pour $j = 2$, les relations (1) et (2) redonnent les expressions (26) et (27). On a naturellement

$$\sum_{s=1}^n s q_{sn} = n j^n \quad (3)$$

(ALLAIS, 1978, *id.*).

Soit r_{sn} le nombre moyen de suites de longueur s dans une suite de n tirages. On a

$$r_{sn} = q_{sn}/2^n = (n + 3 - s)/2^{s+1} \quad \text{pour} \quad 1 \leq s \leq n - 1 \quad (30)$$

$$r_{nn} = 2/2^n = 1/2^{n-1} \quad \text{pour} \quad s = n \quad (31)$$

$$R_n = Q_n/2^n = (n + 1)/2 \quad R_{sn} = Q_{sn}/2^n = (n - s + 2)/2^s \quad (32)$$

Pour $n = 1.000$ on prend : $s_1 = 6$. On a ainsi $Q_{s,n} = 15,56$; et le tableau suivant.

TABLEAU R.3

s	1	2	3	4	5	$s_1 = 6 \leq s \leq n$	$1 \leq s \leq n$
r_{sn}	250,5	125,12	62,5	31,22	15,59	$R_{6,1000} = 15,56$	$R_n = 500,5$

Test χ^2

La confrontation des nombres théoriques r_{sn} et R_{sn} ci-dessus définis avec les valeurs observées n_s et $n (s_1 \leq s) = \sum_{s_1}^n n_s$ permet de calculer

$$\chi^2 = \sum_{s=1}^{s_1} (n_s - r_{sn})^2 / r_{sn} \quad s_1 = n - 1 \quad (33)$$

le dernier terme étant naturellement égal à

$$[n (s_1 \leq s) - R_{s,n}]^2 / R_{s,n} \quad (34)$$

On choisit s_1 de manière que $R_{s,n}$ soit supérieur ou égal à 10 ⁽¹⁵⁾.

Ici le nombre de degrés de liberté ν est égal au nombre de classes, car la valeur observée de R_n n'est pas nécessairement égale à sa valeur théorique $(n + 1)/2$ ⁽¹⁶⁾ ⁽¹⁷⁾.

(R.15) CRAMER, 1946, p. 359.

(R.16) Ce test des suites s'est révélé très efficace dans différentes applications. Il est appliqué ci-dessous dans le § U.6.

(R.17) Sur la théorie des suites (*runs*) voir notamment : KERMACK et Mac KENDRIK, 1936-1937; DAVIS, 1941, p. 164-174; KENDALL et SMITH, 1938 et 1939; MOOD, 1940; KENDALL, 1948, vol. II, p. 124 127; HALD, 1952, p. 342 357 et 372 373; TINTNER, 1952, p. 234-238; FELLER, 1957, p. 40, 60, 180, 279 et 299 304; KENDALL, 1976, p. 21 28. Toutes ces références reposent généralement sur des formulations entièrement différentes de celle exposée ci-dessus et la plupart correspondent à des tests *non paramétriques*. Aucune d'elles ne repose sur les formules (30) et (31) ci-dessus.

Le principe des tests *non paramétriques* est que l'on considère comme également possibles les $n!$ permutations de n nombres donnés indépendamment de la loi de distribution (KENDALL, 1948, *id.*, § 21.45, p. 216) alors que dans le test ci-dessus exposé on considère comme également possibles les 2^n trajectoires correspondant à n tirages. Le premier cas correspond au battage de cartes dans un jeu de n cartes. Le second correspond au tirage avec remplacement dans une urne. Ce sont là deux problèmes entièrement différents et les formulations applicables à l'un ne sont pas applicables à l'autre.

Dans les deux cas les cas possibles sont les mêmes, mais ils interviennent avec des pondérations différentes. Dans le cas du battage de n cartes on a $n!$ cas également possibles. Dans le cas du tirage avec remplacement dans une urne contenant K boules dont k_1 blanches et k_2 noires (k_1 et k_2 premiers entre eux), on a K^n cas également possibles. On retrouve ici les questions tout à fait essentielles du dénombrement des cas également possibles (tout à fait analogue à celle qui se rencontre dans les probabilités géométriques; voir mon précédent article § 3 et § A.2.4, *J.S.S.P.*, tome 124, n° 2, 1983, p. 76 et 97), et de la définition du « choix au hasard » (*id.* § 9, p. 85 86) qui ne sauraient trouver de réponse *a priori* sans considérer le processus qui engendre tous les cas possibles. En fait la distribution des suites est inséparable du processus par lequel on les obtient.

Il résulte de là que si l'on considère les suites de valeurs positives et négatives dans une série temporelle de termes indépendants, les décomptes correspondant aux tests non paramétriques ne peuvent que conduire à des conclusions erronées. Le lecteur pourra utilement comparer avec les relations (30) et (31) ci-dessus la relation (avec mes notations)

$$r_{sn} = 2 (n - s - 2) (s^2 + 3s + 1) / (s + 3)! \quad (1)$$

indiquée par KENDALL (1948, *id.*, p. 125, relation 21.73). En fait seules les relations (30) et (31) sont applicables aux séries temporelles dans le cas où la fréquence intrinsèque est égale à 1/2.

R.4. Tests des valeurs cumulées

Le calcul des valeurs cumulées

$$U_n = \sum_{n=1}^n u_n \quad (35)$$

d'une série u_n dont on ne connaît pas la structure effective mais dont on peut supposer que la moyenne est nulle, peut être utilisé soit pour déceler des régularités non apparentes sur les graphiques représentatifs des quantités considérées par amplification des amplitudes de certaines composantes périodiques s'il en existe, soit pour tester l'hypothèse d'indépendance ⁽¹⁸⁾.

Lorsque par hypothèse une série u_n est constituée par une suite de termes indépendants, il est très instructif d'examiner le comportement de leurs valeurs cumulées de manière à vérifier si ce comportement est « vraisemblable » au regard de cette hypothèse.

Trois critères déduits de la considération de la série des u_n peuvent permettre d'apprécier dans quelle mesure les u_n peuvent être considérés comme indépendants et normalement distribués ⁽¹⁹⁾.

a) Chenaux correspondant à des valeurs données des écarts

Si les u_n sont indépendants et normalement distribués on a

$$\sigma_U(n) = \sigma_u \sqrt{n} \quad (36)$$

où $\sigma_U(n)$ est l'écart type de la valeur cumulée U_n .

Si on considère les courbes définies par

$$\alpha_n = \pm \sigma_u \sqrt{n} \quad \beta_n = \pm 2 \sigma_u \sqrt{n} \quad (37)$$

il résulte de la table de la loi normale que 68 % des points de la trajectoire U_n sont à l'intérieur du chenal des courbes α , et que 95,4 % des points de la trajectoire U_n sont à l'intérieur des courbes β .

Si donc l'hypothèse d'indépendance des u_n est justifiée on doit constater que la courbe représentative de U_n doit rester comprise dans sa plus grande partie entre les courbes β .

b) Maximum de U_n dans l'intervalle $1 \leq n \leq N$

Soit M_{UN} la valeur maximale de U dans l'intervalle $1 \leq n \leq N$. On sait que la fréquence d'une valeur de M_{UN} supérieure ou égale à $\sigma_U \sqrt{N}$ est égale au double de la fréquence d'un écart relatif supérieur en valeur absolue à $\sigma_U \sqrt{N}$ ⁽²⁰⁾. Autrement dit

$$F_1 = \text{Fréquence } \{\lambda \geq \lambda_1\} = 2 \int_{\lambda_1}^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{\lambda^2}{2}} d\lambda \quad (38)$$

en posant

$$\lambda = M_U / \sigma_U \sqrt{N} \quad (39)$$

Pour que l'hypothèse d'indépendance soit « vraisemblable » au regard du maximum observé M_U il faut que la fréquence F_1 correspondant à la valeur observée λ_1 de λ n'ait pas une valeur trop petite.

c) Distance entre deux passages successifs par zéro de U_n

C'est là une circonstance générale que les valeurs cumulées U_n d'une variable aléatoire indépendante u_n tendent à se trouver en général d'un même côté de l'axe des n ⁽²¹⁾.

(R.18) Voir le § P.12 ci dessus. Voir des illustrations dans les § T.5 et U.5 ci-dessous.

(R.19) Ces critères sont appliqués dans le § U.5 ci dessous.

(R.20) BACHELIER, 1938, p. 36; et 1939, p. 35. Voir également dans le cas du jeu de pile ou face : Paul LÉVY, 1931, p. 110.

(R.21) Voir par exemple les graphiques très suggestifs de WOLD, 1965, p. 8 32; et de FELLER, 1957, § 111,7, p. 84.

Si donc la courbe représentative de U_n reste d'un même côté de l'axe des n pendant un long intervalle (n_1, n_2) il convient de vérifier qu'un tel intervalle ne doit pas être considéré comme anormal au regard de l'hypothèse de normalité et d'indépendance des u_n .

Dans le cas d'un jeu de pile ou face pour lequel $X_n = \pm 1$ Feller donne la fréquence f_n ($n = 2q$) pour que le premier retour à zéro de

$$Y_n = \sum_{n=1}^n X_n \quad (40)$$

ait lieu à l'abscisse n .

On a ⁽²²⁾

$$f_{2q} = \frac{1}{2q-1} C_{2q-1}^a \frac{1}{2^{2q-1}} = \frac{1}{q} \frac{2q-2!}{q-1!} \frac{1}{2^{2q-1}} \quad (41)$$

On en déduit d'après la formule de Stirling ⁽²³⁾

$$f_N \sim \sqrt{2} / \sqrt{\pi} N^{3/2} \quad (42)$$

Il résulte de là que la fréquence F des trajectoires pour lesquelles le premier retour à zéro a lieu pour $n \geq N$ est

$$F(N) \sim \sqrt{2/\pi} \sum_{n=N}^{\infty} (1/n^{3/2}) \quad (43)$$

On a

$$\sum_{n=N}^{\infty} \frac{1}{n^{3/2}} \sim \int_N^{+\infty} \frac{dn}{n^{3/2}} = \left[-\frac{2}{\sqrt{n}} \right]_N^{+\infty} = \frac{2}{\sqrt{N}} \quad \text{pour } N \gg 1 \quad (44)$$

d'où finalement

$$F(N) \sim 2\sqrt{2/\pi} / \sqrt{N} \sim 1,5957 / \sqrt{N} \quad \text{pour } N \gg 1 \quad (45)$$

Cette formule valable dans le cas d'un jeu de pile ou face pour lequel $X_n = \pm 1$, dont l'écart type est $\sigma = 1$, vaut encore pour un jeu pour lequel $X_n = \pm a$, où a est entier, pour lequel l'écart type est égal à a , et pour lequel l'écart type des valeurs cumulées est égal à $a\sqrt{n}$.

Dans le cas où la distribution de X_n est normale et d'écart type σ la distribution asymptotique de Y_n ne dépend que de $\sigma\sqrt{n}$. On peut donc en déduire que la formule (45) reste applicable ⁽²⁴⁾.

Pour que l'hypothèse d'indépendance soit « vraisemblable » au regard des valeurs constatées n_1 et n_2 , il faut que la valeur de $F(n_2 - n_1)$ ne soit pas trop faible.

R.5. Représentation de Henri de la distribution des fréquences correspondant à une sinusoïde

Comme il a été indiqué ⁽²⁵⁾, la fonction de fréquence $F(x)$ correspondant à

$$x = a \cos [2\pi(t - t_0)/T] \quad (46)$$

est

$$F(x) = \frac{1}{2} - \frac{1}{\pi} \arcsin \frac{x}{a} \quad (-a \leq x \leq a) \quad (47)$$

(R.22) FELLER, 1957, tome 1, p. 76, relation 4.11; WOLD, 1965, p. 31, relation 75.

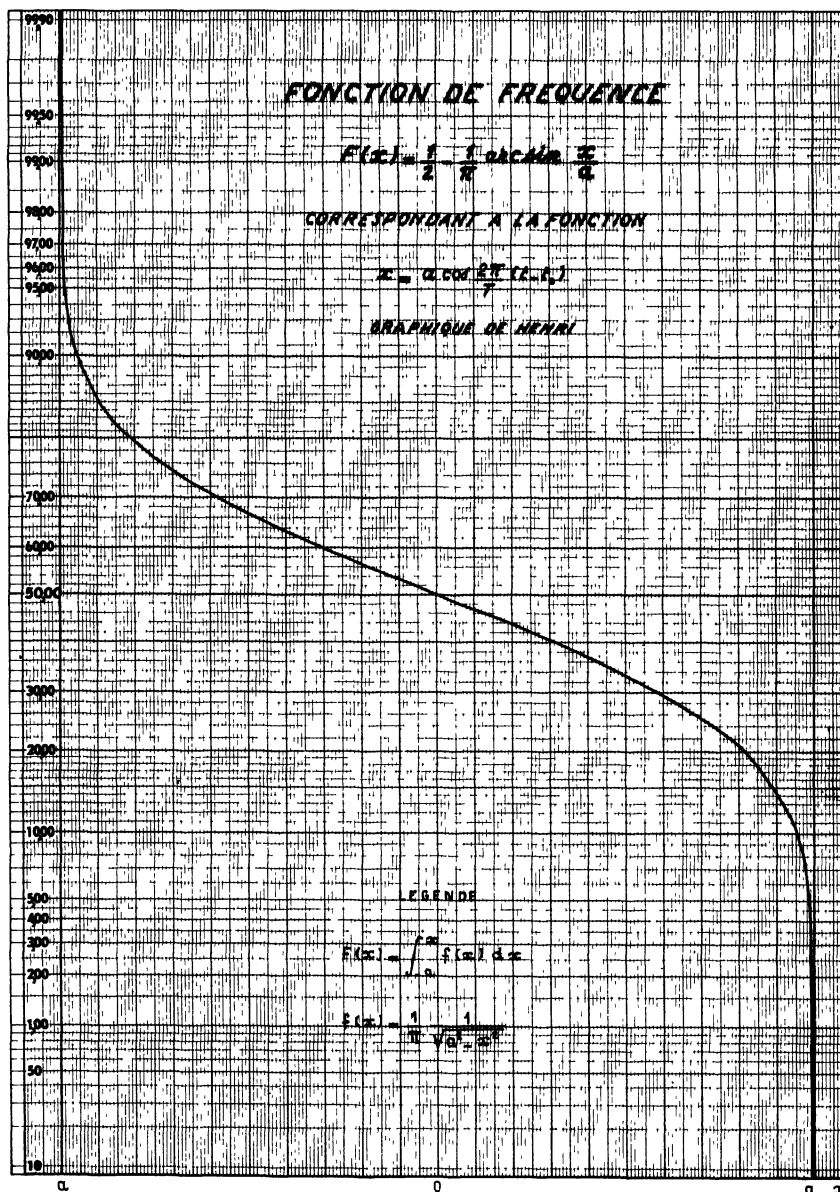
(R.23) FELLER, *id.*, p. 87, relation 8, 4; WOLD, 1965, *id.*, p. 31 donne bien la même formule (41), mais la valeur asymptotique indiquée $f_N \sim 2\sqrt{2/\pi} N^{3/2}$ double de celle de Feller est erronée.

(R.24) Avec une approche mathématique entièrement différente et une formulation plus générale, Bachelier (1915) a étudié un cas analogue.

(R.25) § Q.4. Lemme I.

Le Graphique R.1 représente la fonction $F(x)$ suivant la méthode de Henri. On voit que la courbe représentative est très différente ⁽²⁶⁾ de celle correspondant à une distribution normale à laquelle correspond une droite. Le Théorème (T) ci-dessus ⁽²⁷⁾ n'en prend que plus de signification.

GRAPHIQUE R.1 (§ R 5)



(R.26) Le Graphique R.1 montre une prédominance des valeurs extrêmes comme il résulte de la figure de la note (Q.10) ci-dessus. 60 % des valeurs de x/a sont extérieures à l'intervalle $(-0,6, +0,6)$, et 30 % correspondent à des valeurs telles que $0,9 < |x/a| < 1$. La médiane des valeurs de x comprises entre 0 et a est définie par la relation $\sin x/a = \pi/4$, d'où $x/a = 0,707$.

(R.27) § Q.3 ci-dessus.

S. SÉRIE S

ANALYSE D'UNE SÉRIE TEMPORELLE ISSUE DES DONNÉES DE L'OBSERVATION
ET CONSTITUÉE PAR LES 721 VALEURS SUCCESSIVES DE LA SOMME DE 13 SINUSOÏDES
DONT LA DISTRIBUTION EST CELLE DE LA LOI NORMALE

S.1. La série S

Comme il a été indiqué ⁽¹⁾, la série S est constituée par les sommes

$$X(n) = \sum_{j=1}^l x_j(n) = \sum_j a_j \cos 2\pi f_j (n - n_j) \quad (1 \leq j \leq l = 13; 1 \leq n \leq N = 721) \quad (1)$$

de 13 sinusoides dont les périodes sont les périodes principales de la théorie des marées.

La structure de la série S est indiquée sur le *Tableau S.1*. La série comprend 721 valeurs qui sont représentées sur le *Graphique S.1*.

S.2. Ajustement de la distribution des valeurs de X (n) suivant la loi normale

Le *Graphique S.2* représente l'ajustement de la distribution des X (n) suivant la loi normale, ajustement certainement *très remarquable*. Il l'est d'autant plus que la série des X (n) ne comporte que 13 sinusoides et que l'une d'entre elles a une amplitude plus de 3 fois plus élevée que la moyenne des amplitudes et que le rapport de son carré à la somme des carrés des amplitudes est voisin de 1/2. Encore faut-il souligner que les périodes 6,1 et 6,21, 8 et 8,28 sont respectivement peu différentes; que la période 12,19 est presque égale à la moyenne des périodes voisines $(12 + 12,42)/2 = 12,21$, que la période 24,84 est également presque égale à la moyenne des deux périodes voisines $(23,93 + 25,82)/2 = 24,875$.

Les valeurs de la moyenne relative $m_x/\sigma_x = -5,61 \cdot 10^{-3}$ et de l'écart type $\sigma_x = 41,0$ de X (n), déduites de l'ajustement de Henri, sont très voisines des valeurs asymptotiques.

$$\mu_x^* = 0 \quad \sigma_x^* = \sqrt{\sum_j a_j^2 / 2} = 39,6$$

Les mêmes valeurs de m_x/σ_x et de σ_x calculées à partir des 721 valeurs de X (n) sont respectivement $-5,64 \cdot 10^{-3}$ et 40,0 (*Graphique S.1*).

Alors que la valeur maximale susceptible d'être atteinte par X (n) est $A = \sum_j a_j = 155,6$, 95 % environ des valeurs de X (n) correspondent à l'intervalle $(-90, +70)$, leurs points représentatifs étant pratiquement alignés sur le *Graphique S.2* de Henri. On a $A/\sigma_x = 155,6/39,6 = 3,93$.

L'accord entre la valeur 40,0 de σ_x correspondant à l'ensemble des 721 valeurs de X (n) et la valeur 41,0 correspondant à l'ajustement de Henri montre d'ailleurs que l'ajustement normal reste approximativement valable en dehors de l'intervalle $(-90, +70)$ ⁽²⁾.

S.3. Coefficients d'autocorrélation asymptotiques de la série des X (n)

Le *Graphique S.3* représente les valeurs asymptotiques

$$\rho_q^* = \left(\sum_j a_j^2 \cos 2\pi f_j q \right) / \sum_j a_j^2 \quad (2)$$

(S.1) Voir ci dessus § P.1 et P.5

(S.2) Sur la comparaison de cet ajustement avec les autres ajustements obtenus voir les Tableaux R.1 et R.2 des § R.1.c et R.1.e ci dessus.

du coefficient d'autocorrélation de la série des $X(n)$ pour $1 \leq q \leq 600$ ⁽³⁾. Comme il est visible ce corrélogramme est dominé par l'onde Q_1 de période 26,87 et d'amplitude 40, dont l'effet est amplifié puisque le carré de son amplitude l'emporte de loin sur le carré de chacune des autres.

Il résulte de là que *bien qu'ils se présentent comme des sommes de sinusoides, les Q_q^* ne se distribuent pas approximativement suivant la loi normale; ce résultat confirme que le Théorème (T) n'est plus valide lorsque les amplitudes des composantes sinusoidales s'écartent par trop d'un ordre de grandeur comparable* ⁽⁴⁾.

Du fait de l'importance relative considérable de l'onde Q_1 , le Graphique de Henri pour les Q_q^* a en fait une allure comparable à celle représentée sur le *Graphique R.1* ci-dessus correspondant à la fonction de fréquence correspondant à une seule sinusoides ⁽⁵⁾.

S.4. Coefficients d'autocorrélation de la série des X (n)

Le *Graphique S.4* représente le corrélogramme calculé r_q ($1 \leq q \leq 600$) à partir des $N = 721$ valeurs de la série S. Ce corrélogramme est *pratiquement identique* au corrélogramme asymptotique Q_q^* du *Graphique S.3* sauf que le corrélogramme asymptotique est plus régulier dans ses amplitudes. Cela tient à ce que le corrélogramme calculé l'est sur 720 valeurs de X_n pour r_1 et seulement sur 121 valeurs de $X(n)$ pour r_{600} .

S.5. Commentaire général

Ces résultats montrent que le *Théorème (T)* ⁽⁶⁾ peut s'appliquer à des fonctions presque périodiques, *fortement autocorrélées, dont le nombre d'ondes composantes est relativement faible*.

Ils sont d'autant plus significatifs que la démonstration du *Théorème (T)* implique la sommation d'un nombre l infini de sinusoides et d'un nombre N également infini de valeurs de n .

Il convient enfin de souligner que *la définition de la série S ne comporte aucun élément d'arbitraire, sa structure étant totalement déterminée par l'expérience*.

(S.3) § Q.10 ci-dessus, relation (56).

(S.4) § Q.9 et Q.13 ci-dessus. Le coefficient \mathcal{L}_1/δ correspondant (§ Q.10 ci-dessus), égal à 0,941, est en fait très voisin de l'unité.

(S.5) § R.5 ci-dessus. Faute de place il n'est pas représenté ici.

(S.6) § Q.3 ci-dessus.

TABLEAU S. 1

SÉRIE S
Somme de 13 sinusoides dont les périodes correspondent
à la théorie des marées
 $l = 13 \quad N = 721$

Données structurelles

$$X(n) = \sum_{j=1}^l x_j(n) \quad x_j(n) = a_j \cos 2\pi f_j(n - n_j) \quad f_j = 1/T_j \quad 1 \leq n \leq N = 721$$

T_j	6,10	6,21	8,00	8,28	12,00	12,19	12,42	12,66	12,87	23,93	24,84	25,82	26,87
a_j	4,50	3,50	5,40	5,60	17,00	17,10	5,80	6,70	1,90	14,90	18,00	15,20	40,00
n_j	1,68	3,90	6	0,60	0,10	2,33	10,8	11,6	1,18	10,5	0,06	3,95	25,7
a_j/a	0,376	0,292	0,451	0,468	1,42	1,43	0,485	0,560	0,159	1,25	1,50	1,27	3,34
$a_j^2/\Sigma a_j^2$	0,006	0,004	0,009	0,010	0,092	0,093	0,011	0,014	0,001	0,071	0,103	0,074	0,511

Coefficients caractéristiques (1)

$$A = \sum_j a_j = 155,6 \quad a = A/l = 11,97 \quad A/\sigma_x^* = 3,93$$

$$\sigma_x^* = \sqrt{\sum_j a_j^2}/2 = 39,6 \quad \mathcal{L}/\delta = \sqrt[3]{\sum_j a_j^3/\sum_j a_j^2} = 0,793$$

$$\sigma_e^* = \sqrt{\sum_j a_j^4/2}/\sum_j a_j^2 = 0,387 \quad \mathcal{L}_1/\delta = \sqrt[3]{\sum_j a_j^6/\sum_j a_j^4} = 0,941$$

Obtention de la Série S (2)

La série S correspond à l'ajustement des 721 moyennes horaires observées de l'azimut du pendule paraconique du 9 juin 1954, 0 h. T.U., au 9 juillet 1954, 0 h. T.U. (3) suivant 13 ondes de la théorie des marées : $M_{s4}, M_4, S_3, M_3, S_2, L_2, M_2, N_2, \mu_2, K_1, M_1, O_1, Q_1$ (4).

Les a_j, n_j correspondent aux amplitudes et aux phases trouvées dans ce calcul pour les 13 ondes considérées. Les valeurs de a_j sont données en dixièmes de grade.

Analyse de la Série S

Graphique S.1. : Graphique des $X(n)$

Graphique S.2. : Distribution des $X(n)$

Graphique S.3. : Corrélogramme asymptotique ρ_q^* des $X(n)$

Graphique S.4. : Corrélogramme r_q des $X(n)$

1. § Q.4, Q.9 et Q.10 ci dessus.

2. Voir § P.1 ci-dessus.

3. ALLAIS, 1957 a et 1957 b.

L'ajustement a été effectué par l'Institut Hydrographique National.

4. SCHUREMAN, 1941, p. 164-167.

GRAPHIQUE S.1 (§ S.1)

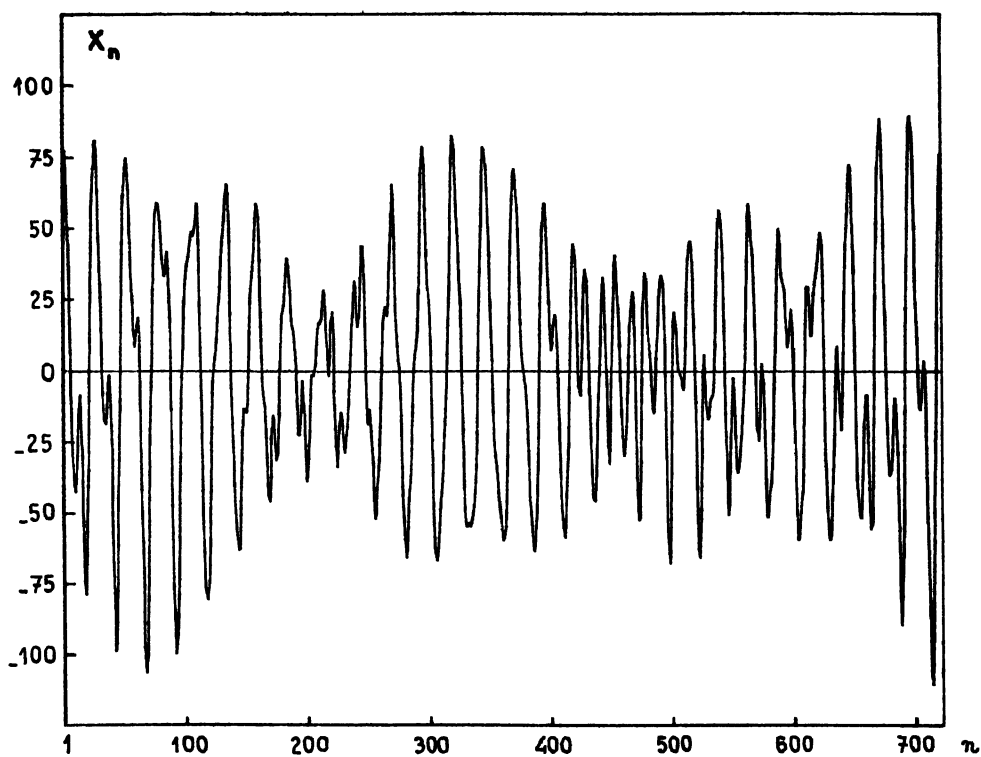
SÉRIE S

Sommes $X(n)$ de 13 sinusoïdes
dont les fréquences correspondent à la théorie des marées

$$X(n) = \sum_{j=1}^l x_j(n) = \sum_{j=1}^l a_j \cos 2\pi f_j (n - n_j)$$

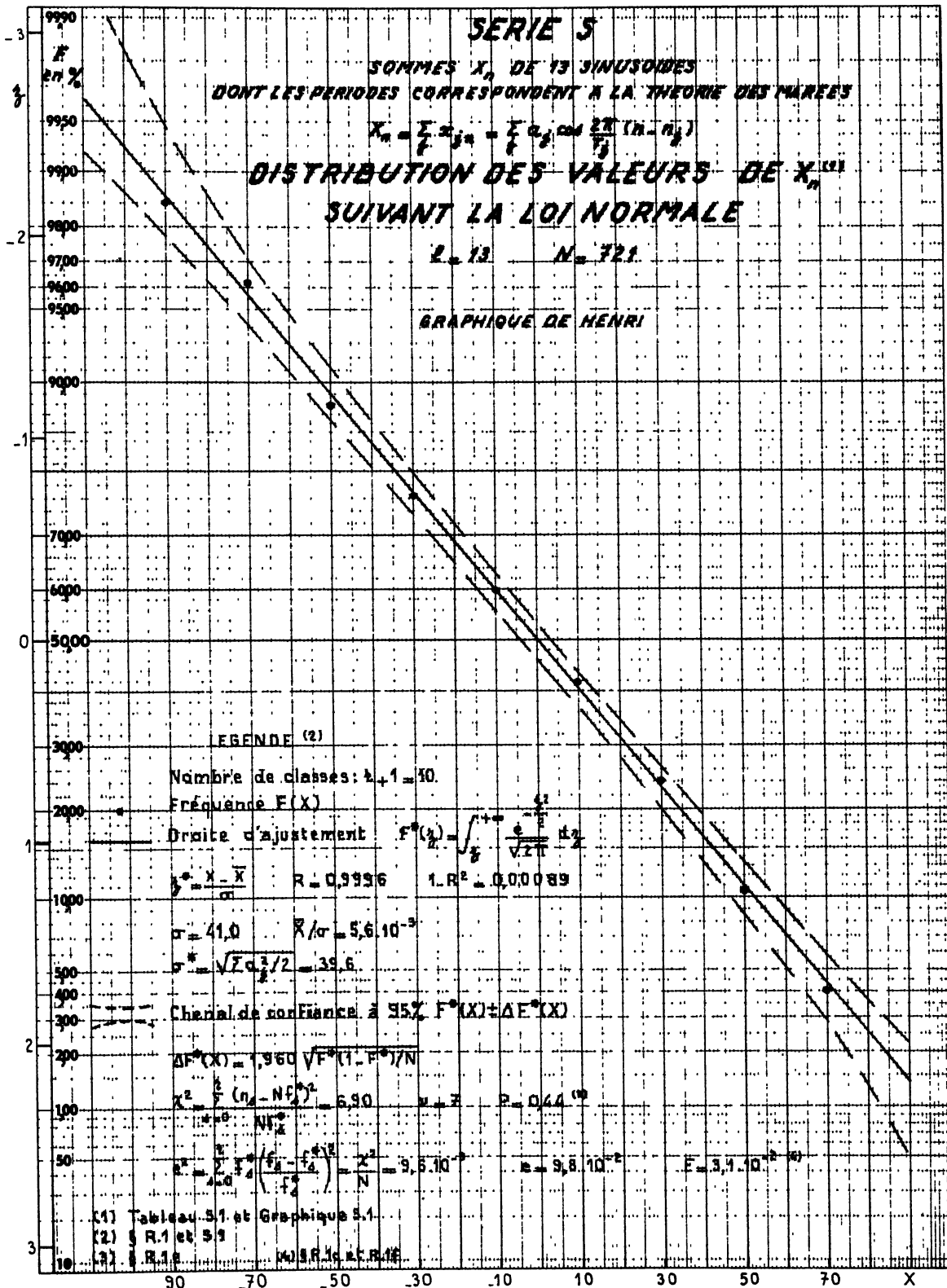
$$l = 13 \quad N = 721$$

$$1 \leq j \leq l \quad 1 \leq n \leq N$$



LÉGENDE voir Tableau S.1
(valeurs des $X(n)$: Tableau 7136).

GRAPHIQUE S.2 (§ S.2)

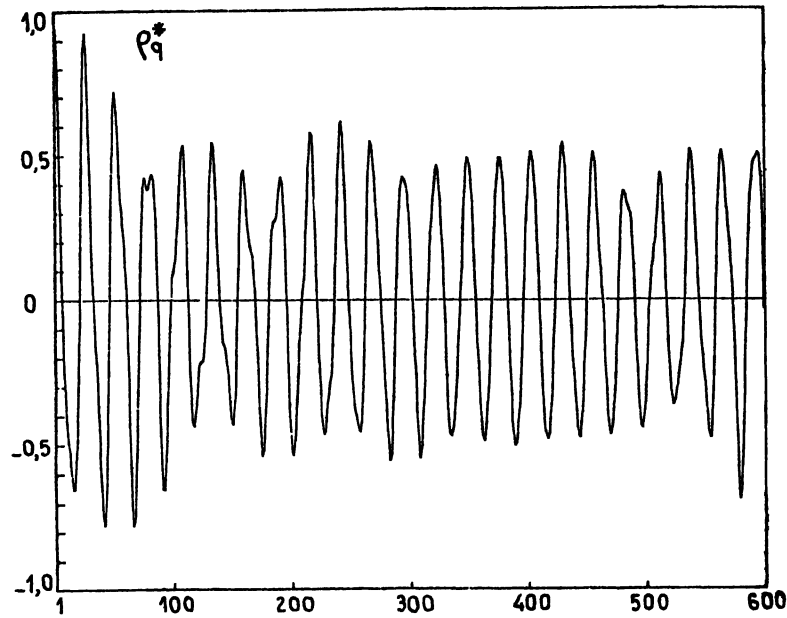
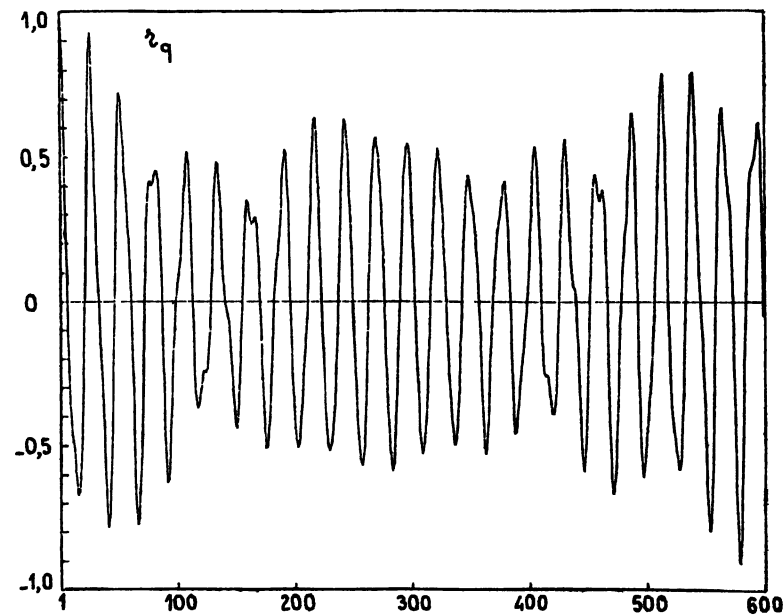


GRAPHIQUES S.3 ET S.4 (§ S.3 4)

SÉRIE S

13 ondes dont les périodes correspondent à la théorie des marées

$$X(n) = \sum_{j=1}^l x_j(n) = \sum_{j=1}^l a_j \cos \frac{2\pi}{T_j} (n - n_j) \quad l = 13 \quad N = 721$$

G.S.3. Corrélogramme asymptotique $\rho_q^* = \sum_j a_j^2 (\cos 2\pi f_j q) / \sum_j a_j^{2(1)}$ ($1 \leq q \leq 600$)G.S.4. Corrélogramme calculé $r_q = r[X(n), X(n+q)]$ ($1 \leq n \leq 721; 1 \leq q \leq 600$)

(1) Voir § Q.10.

T. SÉRIE T
ANALYSE D'UNE SÉRIE TEMPORELLE ARTIFICIELLE
CONSTITUÉE PAR 1 000 VALEURS SUCCESSIVES D'UNE SOMME DE 20 SINUSOÏDES
ET SIMULANT UNE SÉRIE ALÉATOIRE AUTOCORRÉLÉE

T.1. La série T

Comme il a été indiqué ⁽¹⁾ la série T est constituée par une série artificielle de $N = 1.000$ valeurs, sommes de $l = 20$ sinusoides. Sa structure est indiquée sur le *Tableau T.1* et les valeurs correspondantes

$$X(n) = \sum_j a_j \cos 2\pi f_j (n - n_j) \quad (1 \leq j \leq l = 20; 1 \leq n \leq N = 1.000) \quad (1)$$

sont représentées sur le *Graphique T.1*.

La définition de cette série est naturellement totalement arbitraire. Son seul objet est de montrer que si les a_j sont relativement peu différents, la distribution des $X(n)$ est approximativement normale, même pour un nombre réduit de sinusoides constituantes. Les valeurs des T_j , a_j , n_j ont été déterminées par des expressions faciles à calculer (*Tableau T.1*). Les périodes T_j satisfont aux conditions d'irrationalité (B) du § Q.3 dessus ^(1').

T.2. Ajustement de la distribution des valeurs de X (n) suivant la loi normale

Le *Graphique T.2* représente l'ajustement de la distribution des $X(n)$ suivant la loi normale, ajustement qui ici encore est *très remarquable* au regard du faible nombre de sinusoides considéré.

Les valeurs de la moyenne relative $m_x/\sigma_x = -0,0507$ et de l'écart type $\sigma_x = 3,32$ de $X(n)$ déduites de l'ajustement de Henri sont très voisines de leurs valeurs asymptotiques 0 et $\sigma_x^* = \sqrt{\sum_j a_j^2}/2 = 3,37$

Les valeurs de m_x/σ_x et de σ_x calculées à partir des 1.000 valeurs de $X(n)$ sont respectivement 0,00088 et 3,37 (*Graphique T.1*).

Alors que la valeur maximale susceptible d'être atteinte pour $X(n)$ est égale à $A = \sum a_j = 20,5$, 96 % environ des valeurs de $X(n)$ correspondent à l'intervalle $(-7, +7)$, leurs points représentatifs étant pratiquement alignés sur le *Graphique de Henri*. On a $A/\sigma_x = 20,5/3,37 = 6,17$.

L'accord entre la valeur 3,37 de σ_x calculé à partir des 1.000 valeurs de $X(n)$ et la valeur 3,32 déduite du *Graphique de Henri* montre d'ailleurs que l'ajustement normal reste approximativement valable en dehors de l'intervalle $(-7, +7)$ ⁽²⁾.

T.3. Coefficients d'autocorrélation asymptotiques de la série des X (n)

A l'expression (1) des $X(n)$ correspond l'expression asymptotique

$$\varrho_q^* = (\sum_j a_j^2 \cos 2\pi f_j q) / \sum a_j^2 \quad 1 \leq j \leq l = 20 \quad (2)$$

des coefficients d'autocorrélation ⁽³⁾

a) Graphique représentatif

Le *Graphique T.3* représente le corrélogramme asymptotique ϱ_q^* pour $1 \leq q \leq 600$. Les vingt premières valeurs de ϱ_q^* sont représentés à une échelle agrandie.

(T.1) Voir ci dessus § P.5.

(T.1') A l'approximation naturellement des calculs effectués sur ordinateur. Voir note Q.33 ci dessus.

(T.2) Sur la comparaison de cet ajustement avec les autres ajustements obtenus voir les Tableaux R.1 et R.2, des § R.1.c et R.1.e ci-dessus.

(T.3) § Q. 10 ci-dessus, relation (56).

L'écart type et la moyenne rapportée à l'écart type des 600 premières valeurs de ϱ_q^* sont respectivement 0,176 et $-5,7 \cdot 10^{-4}$. Leurs valeurs théoriques sont respectivement $\sigma(\varrho_q^*) = 0,178$ et $\mu(\varrho_q^*) = 0$

b) *Ajustement de la distribution des coefficients d'autocorrélation asymptotique ϱ_q^* suivant la loi normale*

Le *Graphique T.5* représente l'ajustement de la distribution des coefficients d'autocorrélation ϱ_q^* suivant la loi normale pour $1 \leq q \leq 1.000$. L'ajustement est tout aussi remarquable que les précédents. Compte tenu de l'expression (2) de ϱ_q^* on a ici une nouvelle illustration de la validité du *Théorème (T)* ⁽⁴⁾.

Les valeurs de la moyenne $\bar{\varrho}_q^* = 1,00 \cdot 10^{-4}$ et de l'écart type $\sigma(\varrho_q^*) = 0,176$ déduites de l'ajustement de Henri sont très voisines de leurs valeurs moyennes asymptotiques

$$\bar{\varrho}_q^* = 0 \quad \sigma(\varrho_q^*) = \sqrt{\frac{\sum_j a_j^4 / 2}{\sum_j a_j^2}} = 0,178 \text{ }^{(5)}$$

Les valeurs de $\bar{\varrho}_q^*$ et $\sigma(\varrho_q^*)$ calculées à partir des 1.000 premières valeurs de ϱ_q^* sont respectivement $0,205 \cdot 10^{-4}$ et 0,177.

On voit sur le *Graphique T.5* que 97 % des valeurs de ϱ_q^* sont comprises dans l'intervalle $(-0,375; +0,375)$ et que leurs 16 points représentatifs prennent place au voisinage immédiat de la droite d'ajustement.

La valeur maximale de ϱ_q^* susceptible d'être atteinte est égale à l'unité et on a

$$1/\sigma(\varrho_q^*) = 1/0,178 = 5,62.$$

L'accord entre la valeur 0,177 de l'écart type calculé de ϱ_q^* et la valeur 0,176 déduite du *Graphique de Henri* montre d'ailleurs que l'ajustement normal reste approximativement valable en dehors de l'intervalle $(-0,375; +0,375)$ ⁽⁶⁾.

T.4. Coefficients d'autocorrélation r_q de la série des $X(n)$

Les coefficients d'autocorrélation r_q sont calculés à partir de la série des $X(n)$.

a) *Graphique représentatif*

Le *Graphique T.4* représente les 600 premières valeurs du coefficient d'autocorrélation r_q calculé à partir des $N = 700$ premières valeurs des $X(n)$ ⁽⁷⁾. Pour les petites valeurs de r_q ce corrélogramme est très peu différent du corrélogramme asymptotique ϱ_q^* du *Graphique T.3*. Pour les grandes valeurs de q les différences sont un peu plus grandes. Il convient cependant de souligner que le corrélogramme calculé l'est sur 700 valeurs de $X(n)$ pour r_1 et seulement sur 100 valeurs pour r_{600} .

L'écart type et la moyenne rapportée à l'écart type des 600 premières valeurs de r_q sont respectivement 0,179 et $-2,12 \cdot 10^{-3}$. La valeur de l'écart type est ainsi en excellent accord avec l'écart type de 0,177 correspondant aux 600 premières valeurs du coefficient de corrélation asymptotique ϱ_q^* .

b) *Ajustement des coefficients d'autocorrélation r_q suivant la loi normale*

L'ajustement de la distribution des coefficients d'autocorrélation r_q suivant la loi normale est tout aussi remarquable que les précédents. En fait le graphique correspondant est très voisin du *Graphique*

(T.4) § Q.3 ci-dessus.

(T.5) § Q.10 ci-dessus, relation (59).

(T.6) Sur la comparaison de cet ajustement avec les autres ajustements obtenus, voir les Tableaux R.1 et R.2 des § R.1.c et R.1.e ci-dessus.

(T.7) Cette limitation a été imposée par la capacité de l'ordinateur utilisé.

T.5⁽⁸⁾ correspondant à l'ajustement des coefficients de corrélation asymptotiques ρ_q^* et c'est pourquoi il est inutile de le représenter.

Les 600 premiers coefficients d'autocorrélation r_q fluctuent autour d'une valeur peu différente de zéro et leur écart type est égal à 0,179, valeur très voisine de $\sigma(\rho_q^*) = 0,178$.

Pour apprécier la signification de cette valeur, il convient de rappeler que l'ordre de grandeur de l'écart type des 600 premiers coefficients d'autocorrélation calculés à partir des 700 premières valeurs d'une série supposée constituée de termes indépendants est⁽⁹⁾

$$\sigma_r \sim 1/\sqrt{700 - (600/2)} = 0,05$$

L'autocorrélation constatée de la série T est ainsi plus de trois fois plus forte que celle qui correspondrait à l'autocorrélation d'une série de termes indépendants⁽¹⁰⁾. On peut en déduire que pour un analyste qui ignorerait la structure effective de la série $X(n)$ ce résultat permet de conclure qu'il ne s'agit pas d'une suite de termes indépendants^(10').

T.5. Valeurs cumulées $Y(n)$ des $X(n)$ et $\pi^*(q)$ des ρ_q^*

Comme il a été indiqué et dans l'hypothèse où l'on ne connaît pas la structure effective d'une série temporelle⁽¹¹⁾, l'algorithme des valeurs cumulées permet d'amplifier les amplitudes de certaines périodes et de mettre en évidence des régularités sous jacentes non visibles sur les graphiques représentatifs des quantités considérées.

Les valeurs cumulées des $X(n)$ et des ρ_q^* ont pour expression

$$Y(n) = \sum_{p=1}^n X(p) \quad \pi_q^* = \sum_{p=1}^q \rho_p^* \quad (3)$$

a) Valeurs cumulées $Y(n)$ des $X(n)$

Le Graphique T.6 représente les valeurs cumulées $Y(n)$ de $X(n)$. En fait on voit apparaître sur ce graphique des valeurs cumulées une régularité tout à fait remarquable correspondant à un mouvement presque périodique d'ensemble d'une presque période moyenne de l'ordre de 80 avec une période secondaire de l'ordre de 50. On constate également que la courbe représentative de $Y(n)$ se reproduit presque identiquement par une translation Θ de l'ordre de 330.

A première vue ces résultats peuvent paraître quelque peu surprenants au regard du fait que toutes les périodes T_j et toutes les amplitudes a_j sont de l'ordre de l'unité (Tableau T.1). Mais en fait ces résultats s'expliquent sans difficulté au regard des propriétés des sommes cumulées de sinusoides pour des valeurs entières de l'argument⁽¹²⁾. En fait deux valeurs des fréquences f_j sont très voisines d'un entier. D'où pour les valeurs cumulées des $x_j(n)$ des périodes T_j^* et des amplitudes a_j^* relativement grandes comme il résulte du Tableau ci-dessous⁽¹²⁾.

(T.8) Les valeurs de la moyenne \bar{r}_q^* et de l'écart type $\sigma(r_q^*)$ déduites du Graphique de Henri correspondant aux 600 valeurs de r_q sont pratiquement identiques aux valeurs correspondantes du Graphique de Henri correspondant aux 1.000 valeurs de ρ_q^* .

(T.9) § R.2 a ci dessus.

(T.10) $0,179/0,05 = 3,58$.

(T.10') Cette conclusion est confirmée par l'analyse des valeurs cumulées de $X(n)$ et de ρ_q^* (§ T.5 ci dessus).

(T.11) § R.4 ci-dessus.

(T.12) Voir ci-dessus § P.13. Les composantes $y_{10}(n)$ et $y_{11}(n)$ correspondent respectivement aux deux nombres premiers $r_{10} = 29$ et $r_{11} = 31$ (Tableau du § T.5.a et Tableau T.1).

TABLEAU T.2

j	T_j	f_j	T_j^*	$a_j^* = a_j/2 \sin \pi f_j$
10	0,979	$1 + 0,0213$	$1/0,0212 = 46,893$	7,47
11	1,0132	$1 - 0,0122$	$1/0,0122 = 82,169$	13,08

On vérifie ainsi que les oscillations de $Y(n)$ sont dominés par sa composante $y_{11}(n)$. Comme par ailleurs on a

$$82,169 \times 4 = 328,676 = 328,464 + 0,212$$

$$46,893 \times 7 = 328,251 = 328,464 - 0,213$$

on voit qu'au regard de l'importance relativement faible de toutes les autres amplitudes, $Y(n)$ admet la valeur $\Theta = 328,464$ comme presque période.

On peut remarquer que la somme des 18 sinusoides composantes de la valeur cumulée $Y(n)$ autres que celles correspondant aux valeurs 11 et 10 de j fluctue dans un intervalle de l'ordre de 20 alors que la somme des deux composantes $y_{10}(n)$ et $y_{11}(n)$ fluctue dans un intervalle égal à $14,94 + 26,16 = 41,10$ ⁽¹³⁾ ⁽¹⁴⁾.

En raison de l'influence dominante des composantes $y_{11}(n)$ et $y_{10}(n)$ le *Graphique T.6* présente en outre de nombreuses régularités très remarquables ⁽¹⁵⁾ ⁽¹⁶⁾.

Il résulte de là que pour tout analyste qui ignorerait la structure effective de la série T , le *Graphique T.6* met en pleine évidence que cette série ne saurait être considérée même en première approximation comme une suite de termes indépendants, et que sa structure sous jacente est de nature presque périodique.

b) Valeurs cumulées π_q^* des q_q^*

Le *Graphique T.7* représente les valeurs cumulées π_q^* des q_q^* . Il met en évidence des régularités tout à fait semblables à celles du *Graphique T.6* des valeurs cumulées des $Y(n)$.

En fait aux échelles près le *Graphique T.7* est très peu différent du *graphique T.6* et il donne lieu aux mêmes observations.

T.6. Commentaire général

La série T ne constitue qu'une illustration parmi bien d'autres d'un même fait, savoir que le théorème central limite peut s'appliquer à des fonctions presque périodiques, fortement autocorrélées, dont le nombre d'ondes composantes est relativement faible.

(T.13) Au regard de l'importance respective des variations de $y_{11}(n)$ et des 19 autres composantes de $Y(n)$ on peut considérer comme vraisemblable que la fonction $Y(n)$ ne peut se mettre dans tout l'intervalle $(-\infty < n < +\infty)$ sous la forme de l'expression (14) du § P.11 où la période correspondant au moyen mouvement serait égale à T_{11}^* bien que dans l'intervalle $(1 \leq n \leq 1000)$, la fonction $Y(n)$ (*Graphique T.6*) présente toutes les apparences d'une fonction ayant un moyen mouvement de période égale à T_{11}^* .

Si on retranche de $Y(n)$ la somme des deux composantes correspondant aux indices $j = 10$ et $j = 11$, les valeurs de la série résiduelle se distribuent suivant la loi normale. On obtient ainsi une nouvelle confirmation de la validité approximative du *Théorème (T)*.

(T.14) La composante $x_{17}(n)$ correspondant au nombre premier $r = 59$ (Tableau T.1) a pour fréquence

$$f_{17} = 1/T_{17} = 1/1,39657 = 0,71604 = 5/7 + 0,00175$$

D'après l'expression (17) de $y_j(n)$ du § P.13 les points de la composante $y_{17}(n)$ de la série $Y(n)$ déduite de la série T se répartissent sur 7 sinusoides de périodes toutes égales à $1/0,00175 = 570,1$. Voir la note P.59 ci dessus.

(T.15) Tout à fait analogues à celles que l'on constate dans de nombreuses séries temporelles (voir ALLAIS, 1980, et § P.9 ci dessus).

(T.16) Le fait que les fluctuations de $Y(n)$ ont lieu autour d'une valeur différente de zéro provient simplement de la date initiale $n = 1$ considérée et du fait que la constante d'intégration est en fait prise égale à zéro. Avec une constante d'intégration appropriée les fluctuations de $Y(n)$ auraient lieu autour d'une valeur nulle.

On vérifie ainsi qu'il n'est nullement nécessaire pour sa validité de considérer des valeurs extrêmement grandes du nombre l de sinusoides et du nombre N des valeurs successives.

TABLEAU T.1 (§ T.1)

SÉRIE T

Série artificielle constituée par les sommes de 20 sinusoides

$$l = 20 \quad N = 1\,000$$

Données structurelles ⁽¹⁾

$$X(n) = \sum_{j=1}^l x_j(n) \quad x_j(n) = a_j \cos 2\pi f_j (n - n_j) \quad 1 \leq n \leq N = 1000$$

$$T_j = \sqrt{r_j/5,5} \quad r_j = \text{nombre premier de rang } j \quad (r_1 = 2, \dots, r_{20} = 71)$$

$$0,257 \leq T_j \leq 1,53$$

$$a_j = 1,55 - 0,05 j \quad 0,55 \leq a_j \leq 1,5$$

$$n_j = j/10 \quad 0,1 \leq n_j \leq 2$$

Coefficients caractéristiques ⁽²⁾

$$A = \sum_j a_j = 20,50 \quad a = A/l = 1,025 \quad A/\sigma_x^* = 6,09$$

$$\sigma_x^* = \sqrt{\sum_j a_j^2/2} = 3,37 \quad \mathcal{L}/\delta = \sqrt[3]{\sum_j a_j^3/\sqrt[2]{\sum_j a_j^2}} = 0,628$$

$$\sigma(\varrho_q^*) = \sqrt{\sum_j a_j^4/2}/\sum_j a_j^2 = 0,178 \quad \mathcal{L}_1/\delta = \sqrt[3]{\sum_j a_j^6/\sqrt[2]{\sum_j a_j^4}} = 0,663$$

Analyse de la Série T

Graphique T.1 : Représentation des $X(n)$ Graphique T.2 : Distribution des $X(n)$ Graphique T.3 : Corrélogramme asymptotique ϱ_q^* Graphique T.4 : Corrélogramme calculé r_q des $X(n)$ ⁽³⁾Graphique T.5 : Distribution des coefficients de corrélation ϱ_q^* Graphique T.6 : Représentation des valeurs cumulées $Y(n) = \sum_{p=1}^n X(p)$ Graphique T.7 : Représentation des valeurs cumulées $\pi^*(q) = \sum_{p=1}^q \varrho_q^*$

1. Voir § P.5 et T.1 ci-dessus.

2. § Q.4, Q.9 et Q.10 ci-dessus.

3. Faute d'une capacité suffisante de l'ordinateur utilisé, le calcul a été fait à partir des 700 premières valeurs de $X(n)$.

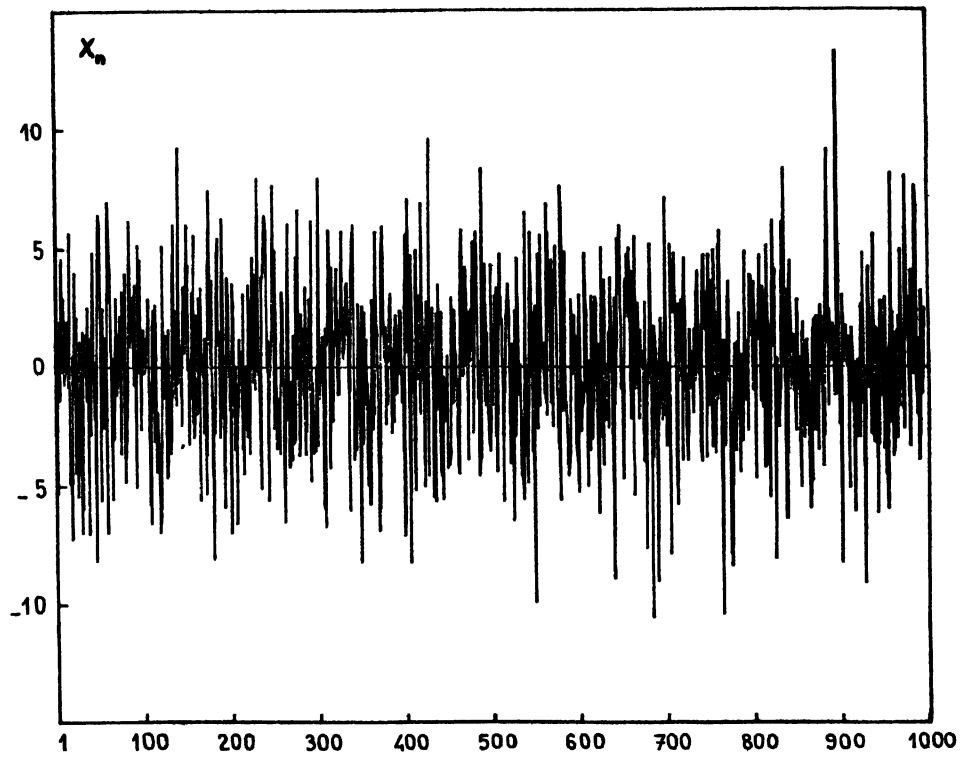
GRAPHIQUE T.1 (§ T.1)

SÉRIE T

Série artificielle correspondant aux sommes $X(n)$ de 20 sinusôides

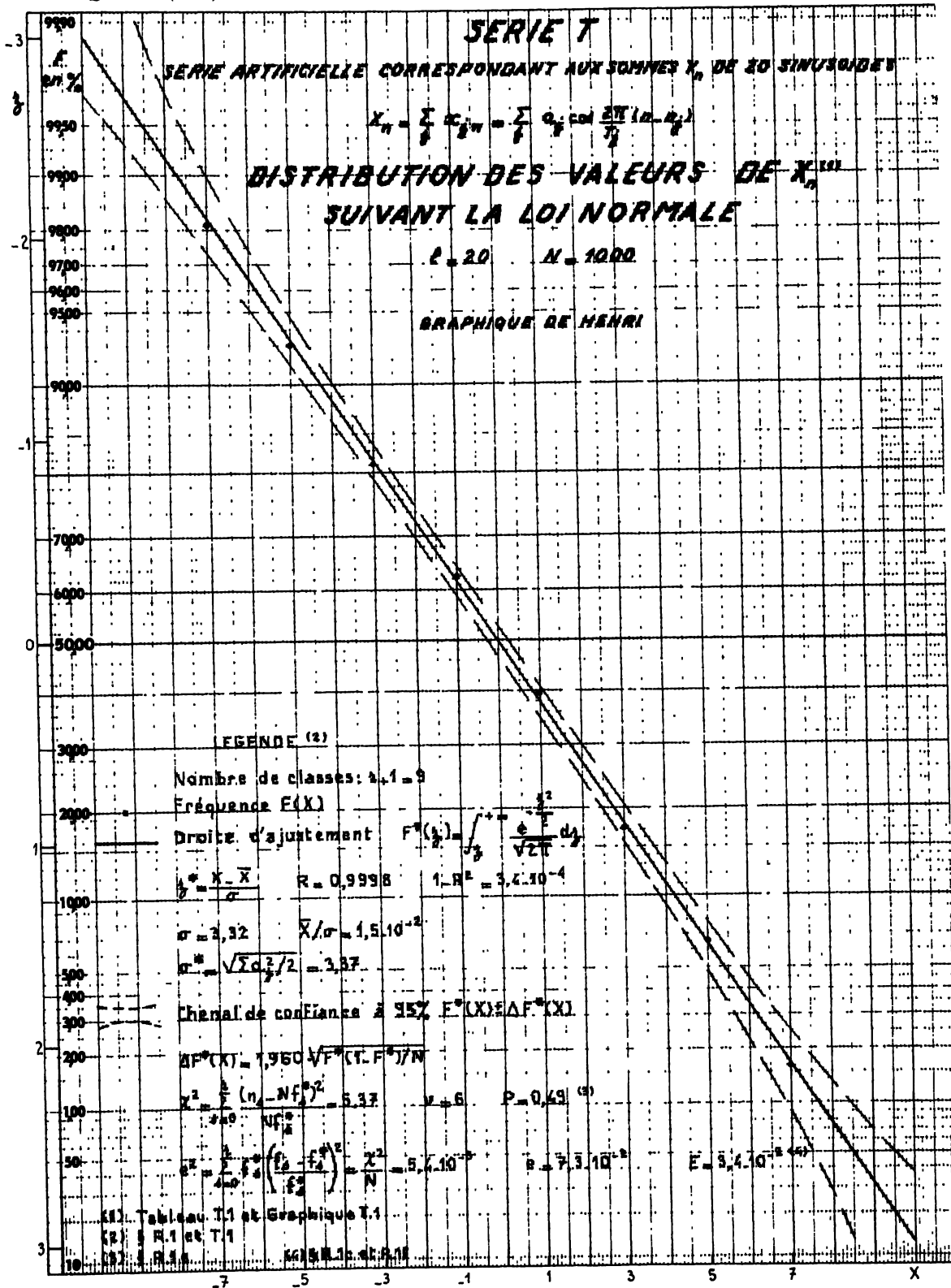
$$X(n) = \sum x_j(n) = \sum a_j \cos 2\pi f_j(n - n_j)$$

$$1 \leq j \leq l \quad 1 \leq n \leq N \quad l = 20 \quad N = 1000$$



LEGENDE voir Tableau T.1
(valeurs des $X(n)$: graphique 8244).

GRAPHIQUE T.2 (§ T.2)



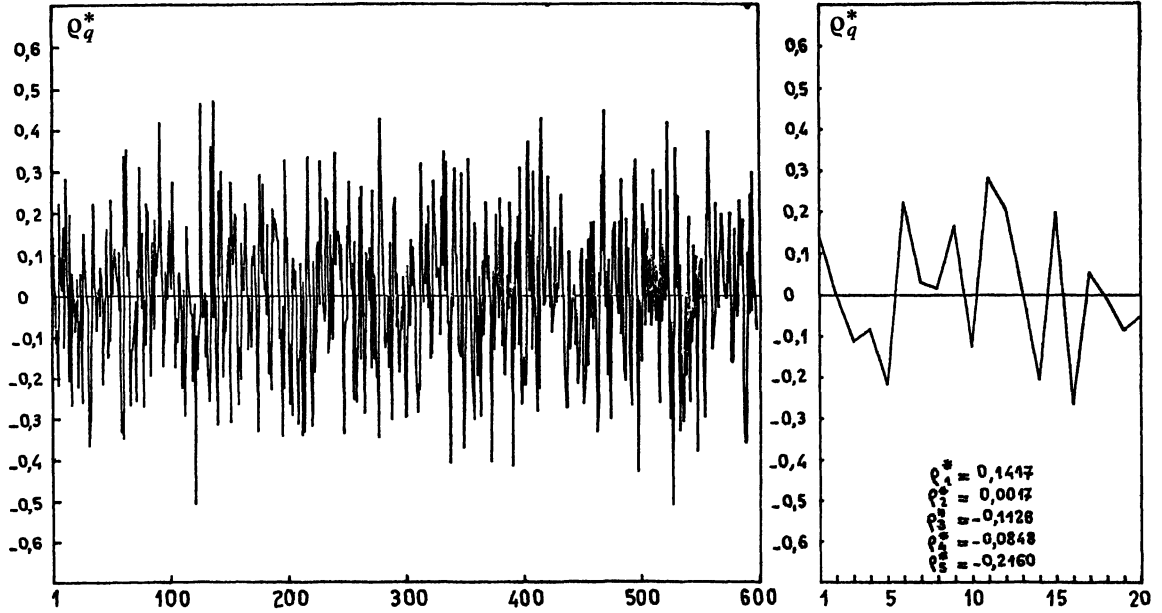
GRAPHIQUES T.3 ET T.4 (§ T.3 4)

SÉRIE T

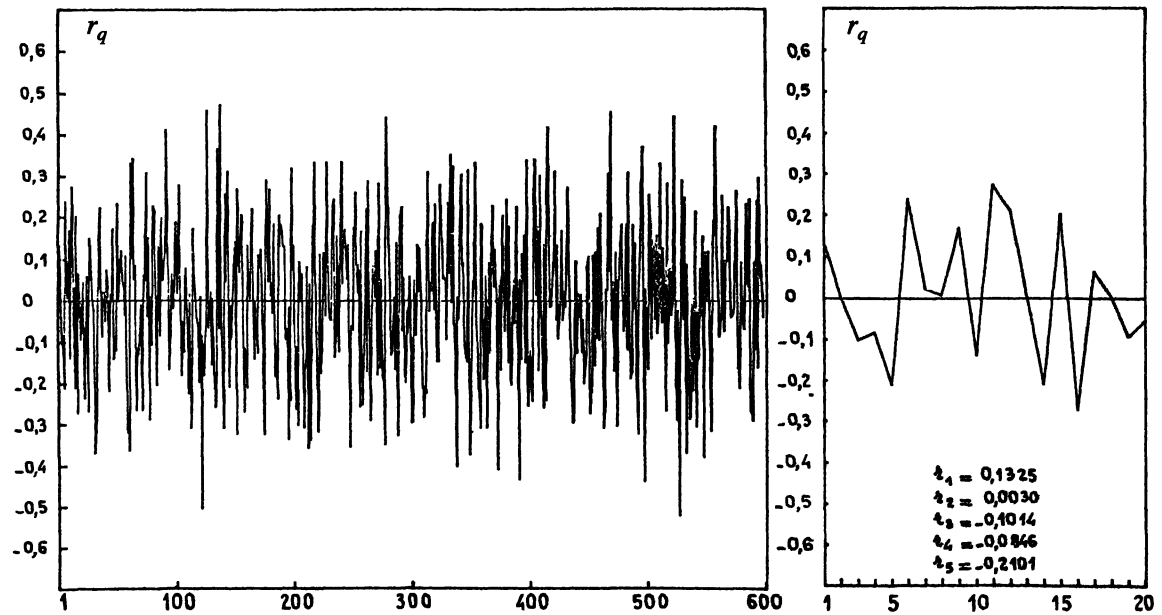
Série artificielle correspondant aux sommes $X(n)$ de 20 sinusôides

Corrélogrammes asymptotique et calculé

G.T.3 Corrélogramme asymptotique $\varrho_q^* = \frac{\sum_j a_j^2 (\cos 2\pi f_j q)}{\sum_j a_j^2}$ ($1 \leq q \leq 600$)

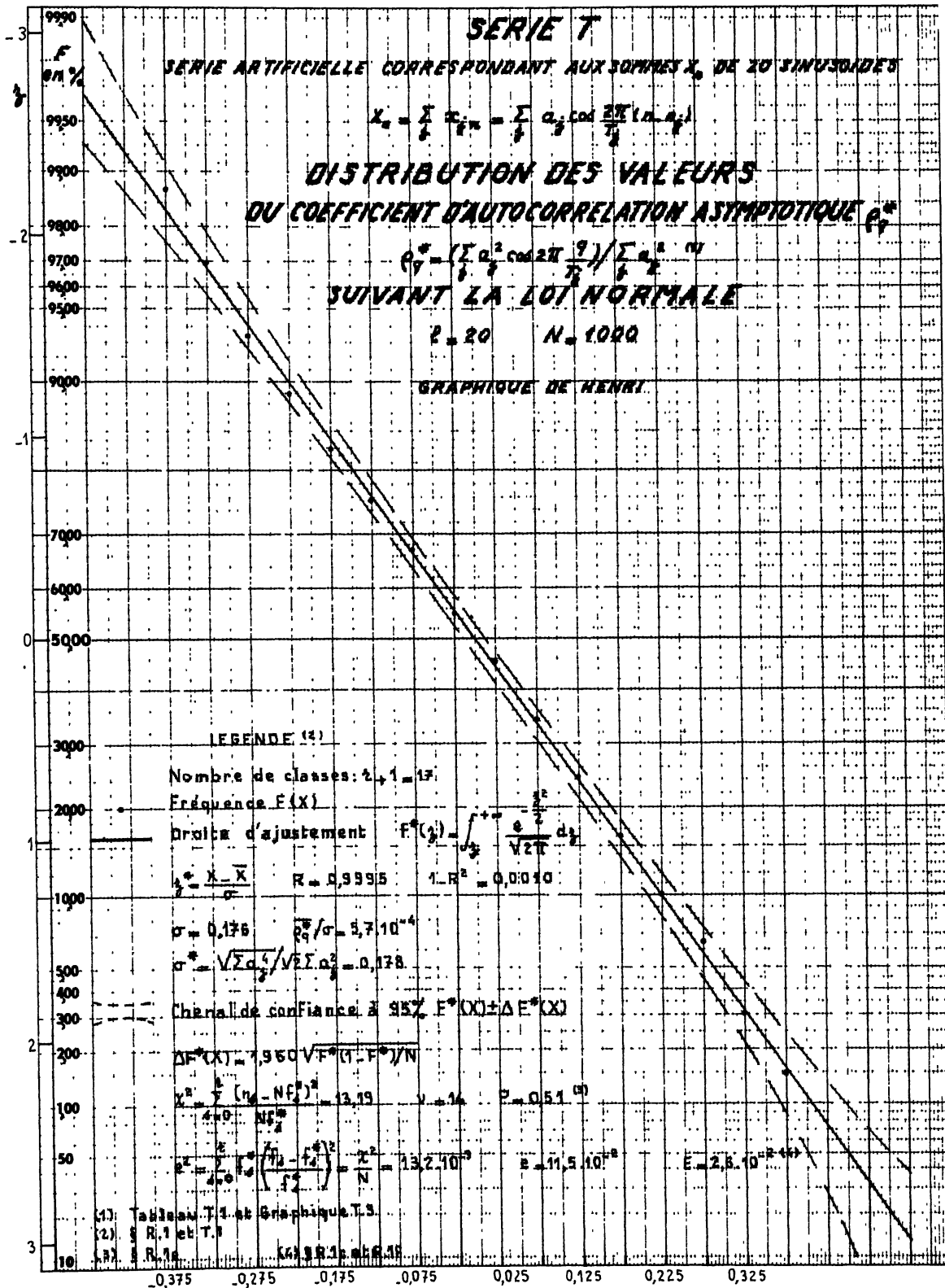


G.T.4. Corrélogramme calculé $r_q = r[X_n, X_{n+q}]$ ($1 \leq n \leq 700; 1 \leq q \leq 600$)



(1) Voir § Q.10, relation (56).

GRAPHIQUE T.5 (§ T.3)



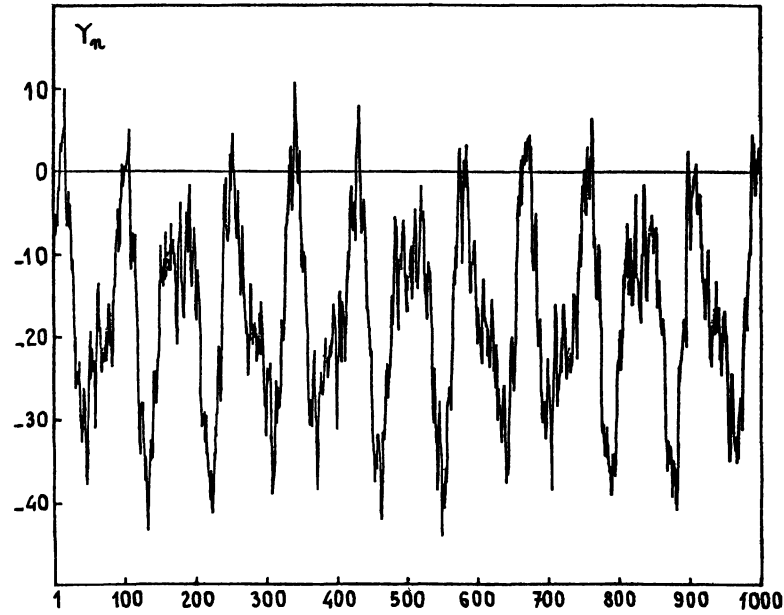
GRAPHIQUES T.6 ET T.7 (§ T.5)

SÉRIE T

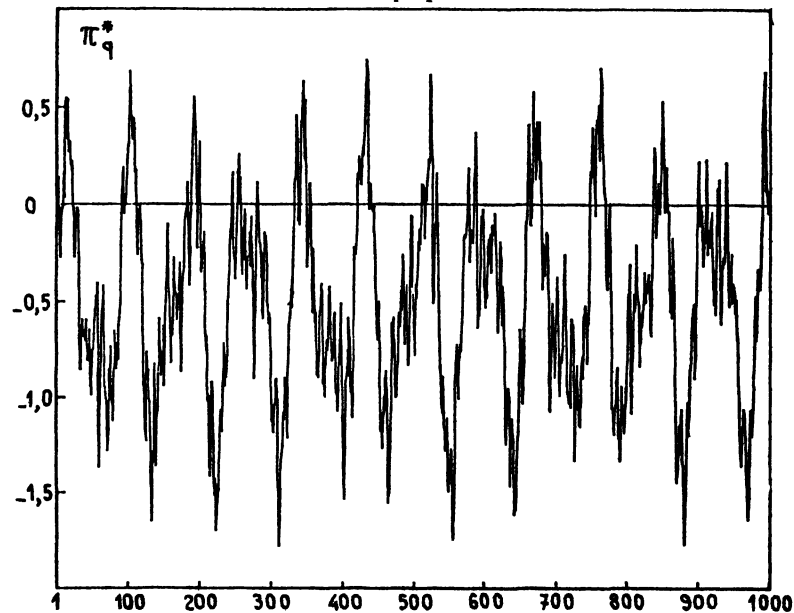
Série artificielle correspondant aux sommes $X(n)$ de 20 sinusôides

$$X(n) = \sum_j x_j(n) = \sum_j a_j \cos 2\pi f_j(n - n_j) \quad 1 \leq j \leq l = 20 \quad (1 \leq n \leq N = 1000)$$

$$G.T.6. \text{ Valeurs cumulées des } X(n) : Y(n) = \sum_{p=1}^n X(p) \quad (1 \leq n \leq N = 1000)$$



$$G.T.7. \text{ Valeurs cumulées des } \varrho_q^* \quad \pi^*(q) = \sum_{p=1}^q \varrho_p^* \quad (1 \leq q \leq 1000)$$



U. SÉRIE U

ANALYSE D'UNE SÉRIE TEMPORELLE ARTIFICIELLE
 CONSTITUÉE PAR LES 1.000 VALEURS SUCCESSIVES D'UNE SOMME DE 200 SINUSOÏDES
 ET SIMULANT UNE SUITE ALÉATOIRE DE TERMES INDÉPENDANTS

U.1. La série U

Comme il a été indiqué ⁽¹⁾, la série U est constituée par une série artificielle de $N = 1.000$ valeurs, sommes de $l = 200$ composantes sinusoïdales. Sa structure est indiquée sur le *Tableau U.1* et elle est représentée sur le *Graphique U.1*.

Tous les a_j ont été pris égaux à l'unité de manière à illustrer le cas où tous les a_j sont peu différents les uns des autres ⁽²⁾. On a alors

$$X(n) = \sum_j \cos 2 \pi f_j (n - n_j) \quad (1 \leq j \leq l = 200; 1 \leq n \leq N = 1\ 000) \quad (1)$$

On a pris

$$n_j = j \quad f_j = \alpha \sqrt{j} \quad T_j = 1/f_j \quad \alpha = \sqrt[4]{23}$$

Il résulte de là que les T_j satisfont aux conditions d'irrationalité du § Q.1 ci-dessus ⁽³⁾.

Avec cette définition des périodes T_j , on a

$$0,03 < T_j < 0,46 \quad \text{médiane des } T_j \sim 0,046$$

Il ne faudrait cependant pas en déduire que les résultats ci-dessous, et notamment l'indépendance apparente des $X(n)$, impliquent que les périodes soient petites. Comme les n_j ont des valeurs entières il résulte du Théorème d'équivalence du § Q.16 que les $x_j(n)$ et $X(n)$ conservent les mêmes valeurs en prenant par exemple les périodes

$$\theta_j = 1/[1/T_j]_d \quad (2)$$

où la notation $[1/T_j]_d$ représente la partie décimale de $1/T_j$. On a alors

$$1 < \theta_j < 246 \quad \text{médiane de } \theta_j \sim 2$$

et respectivement 24 %, 16 % et 10 % des valeurs de θ_j sont supérieures à 4 ; 5,5 ; et 10 ⁽⁴⁾.

Ici encore la définition de la série $X(n)$ est naturellement totalement arbitraire. Son objet essentiel est d'illustrer la génération par la somme d'un grand nombre de sinusoïdes d'une série de termes ayant toutes les apparences d'une suite de termes indépendants et normalement distribués.

(U.1) Voir ci dessus § P.5.

(U.2) § Q.12 ci dessus.

(U.3) A l'approximation naturellement des calculs effectués sur l'ordinateur. (Voir note Q.33 ci-dessus).

(U.4) Tableau U.1 ci-dessous.

Voici à titre d'illustration les valeurs de T_j et θ_j de vingt en vingt avec 3 chiffres significatifs

j	T_j	θ_j	j	T_j	θ_j
1	0,457	5,26	101	0,0454	116
21	0,0996	28,1	121	0,0415	11,2
41	0,0713	44,5	141	0,0385	245
61	0,0585	9,62	161	0,0360	1,27
81	0,0507	1,41	181	0,0339	2,16

U.2. Ajustement de la distribution des valeurs de $X(n)$ suivant la loi normale

Le *Graphique U.2* représente l'ajustement de la distribution des $X(n)$ suivant la loi normale, ajustement qui ici encore est *très remarquable*.

Les valeurs de l'écart type $\sigma_x = 9,73$ et de la moyenne relative $m_x/\sigma_x = 1,76 \cdot 10^{-2}$ de $X(n)$ déduites de l'ajustement de Henri sont *très voisines* des valeurs asymptotiques $\sigma_x^* = \sqrt{l/2} = \sqrt{200/2} = 10$ et $\mu_x^* = 0$ correspondant à la formulation des $X(n)$ ⁽⁵⁾.

Les valeurs de σ_x et m_x/σ_x calculées à partir des 1.000 valeurs de $X(n)$ sont respectivement 9,59 et $3,43 \cdot 10^{-3}$ (*Graphique U.1*).

Alors que la valeur maximale susceptible d'être atteinte par $X(n)$ est égale à $A = \sum a_j = 200$, 95,6 % environ des valeurs de $X(n)$ correspondent à l'intervalle $(-19,5, +19,5)$. Leurs 14 points représentatifs sont pratiquement alignés sur le *Graphique de Henri*. On a $A/\sigma^* = 200/10 = 20$. Si les valeurs $X(n)$ se distribuait rigoureusement suivant la loi normale, une telle valeur devrait être considérée comme ne se rencontrant jamais ⁽⁶⁾.

A titre indicatif, j'ai indiqué dans les Tableaux R.1 et R.2 ci-dessus les résultats de l'ajustement à la loi normale de la distribution de la série des $X(n)$ lorsque l'on considère cette série dans l'intervalle $(1 \leq n \leq 3\,000)$. C'est le meilleur ajustement à la loi normale qui ait été obtenu. On vérifie ainsi que l'applicabilité approximative du *Théorème (T)* à des séries de longueur finie est d'autant meilleure que le nombre N des valeurs successives considéré est plus élevé.

U.3. Coefficients d'autocorrélation asymptotiques ρ_q^* de la série des $X(n)$

A l'expression (1) des $X(n)$ correspond l'expression asymptotique ⁽⁷⁾

$$\rho_q^* = \frac{1}{l} \sum_{j=1}^l \cos 2\pi f_j q \quad 1 \leq j \leq l = 200 \quad (3)$$

des coefficients d'autocorrélation ρ_q^* . Le coefficient ρ_q^* est la limite du coefficient r_q lorsque N augmente indéfiniment.

a) *Graphique représentatif*

Le *Graphique U.3* représente le *corrélogramme asymptotique* ρ_q^* pour $1 \leq q \leq 600$. Les vingt premières valeurs de ρ_q^* sont représentées à une échelle agrandie.

L'écart type et la moyenne des 600 premières valeurs de ρ_q^* sont respectivement 0,0498 et $-1,3 \cdot 10^{-3}$. Pour les 20 premières valeurs l'écart type et la moyenne sont respectivement 0,045 et $-3,6 \cdot 10^{-3}$. Les valeurs asymptotiques sont respectivement $\sigma(\rho_q^*) = 0,05$ et $\rho_q^* = 0$.

b) *Ajustement de la distribution des coefficients d'autocorrélation asymptotiques suivant la loi normale*

Le *Graphique U.5* représente l'ajustement suivant la loi normale de la distribution des $Q = 1.000$ premières valeurs de ρ_q^* . Compte tenu de l'expression (3) de ρ_q^* on a ici une nouvelle illustration de la validité du *Théorème (T)* ⁽⁸⁾.

Les valeurs de l'écart type et de la moyenne de ρ_q^* déduites de l'ajustement de Henri pour $1 \leq q \leq Q = 1.000$ soit 0,050 et $1,9 \cdot 10^{-4}$ sont très voisines ou du même ordre de grandeur que leurs

(U.5) § Q.4, relation (44), et Q.12, relation (68) ci-dessus.

(U.6) Sur la comparaison de cet ajustement avec les autres ajustements obtenus, voir les tableaux R.1 et R.2 des § R.1.c et R.1.e ci-dessus.

(U.7) § Q.12, relation (66).

(U.8) § Q.3 et Q.10 ci-dessus.

valeurs *asymptotiques* correspondant à une valeur infinie de Q égales à $1/\sqrt{2l} = 1/\sqrt{400} = 0,050$ et $\mu_q^* = 0$ ⁽⁹⁾.

Les valeurs de l'écart type et de la moyenne de q_q^* calculées à partir des 1.000 premières valeurs de q_q^* sont respectivement 0,0509 et $-1,12 \cdot 10^{-3}$. La concordance des valeurs des écarts types est *particulièrement remarquable* puisque la valeur asymptotique de l'écart type de q_q^* suppose que les valeurs de N et Q sont infiniment grandes ⁽¹⁰⁾.

On voit sur le *Graphique U.5* que 97 % des valeurs de q_q^* sont comprises dans l'intervalle $(-0,11, +0,11)$, les 12 points représentatifs prenant place très près de la droite d'ajustement (voir les Tableaux R.1 et R.2 du § R.1 ci-dessus).

D'après la relation (3) la valeur maximale de q_q^* susceptible d'être atteinte est égale à l'unité et son rapport à son écart type asymptotique est $1/0,05 = 20$. Si les valeurs de q_q^* se distribuaient rigoureusement suivant la loi normale, une telle valeur devrait être considérée comme ne se rencontrant *jamais* ⁽¹¹⁾.

Ici encore et à titre indicatif j'ai indiqué sur les *Tableaux R.1 et R.2* les résultats de l'ajustement à la loi normale des valeurs de q_q^* dans l'intervalle $(1 \leq q \leq 3\ 000)$. Ici encore on vérifie que la qualité de l'ajustement est améliorée par l'accroissement de la valeur du nombre Q de termes considérés.

U.4. Coefficients d'autocorrélation r_q de la série des $X(n)$

Les coefficients d'autocorrélation r_q sont calculés à partir des valeurs de la série des $X(n)$.

a) *Graphique représentatif des coefficients d'autocorrélation r_q*

Le *Graphique U.4* représente les 600 premières valeurs du coefficient d'autocorrélation r_q calculées à partir des 700 premières valeurs de $X(n)$ ⁽¹²⁾.

L'écart type et la moyenne des 600 premières valeurs de r_q sont respectivement 0,072 et $-1,1 \cdot 10^{-4}$. L'écart type et la moyenne des 20 premières valeurs de r_q sont respectivement 0,052 et $-1,4 \cdot 10^{-2}$.

En fait, on constate sur les *Graphiques U.3** et *U.4** qu'il y a une très bonne correspondance entre les 100 premières valeurs de r_q et les 100 premières valeurs de q_q^* , particulièrement nette pour $q \leq 50$. Au-delà les valeurs de r_q s'écartent sensiblement des valeurs de q_q^* . Cette divergence s'explique par le fait que q_q^* est une valeur *asymptotique* correspondant à une suite comprenant *un nombre N infini* de valeurs de X_n ($1 \leq n \leq N$). Il résulte de là par exemple que le calcul de la 600^e valeur de r_q ne tient compte que de 100 couples de valeurs $X(n)$, $X(n+q)$ alors que q_q^* tient compte d'un nombre *infini* de tels couples.

Il n'en est que plus remarquable que les écarts types de r_q et q_q^* pour les 20 premières valeurs soient respectivement 0,052 et 0,045; pour les 50 premières valeurs 0,058 et 0,052; pour les 100 premières valeurs 0,060 et 0,053, leurs différences étant ainsi relativement faibles. Les 20 premières valeurs de r_q sont toutes comprises entre $-0,1$ et $+0,1$ et leur écart type est égal à 0,052. Pour interpréter ces valeurs il suffit de les rapprocher de l'écart type $\sigma_T = 1/\sqrt{700} = 0,038$ du coefficient d'autocorrélation d'une série de 700 valeurs *indépendantes et normalement distribuées* ⁽¹³⁾. On voit que *la série des $X(n)$ se présente sous les apparences d'une série de termes indépendants*.

(U.9) § Q.12 ci dessus, relation (68).

(U.10) § Q.10 ci dessus.

(U.11) Sur la comparaison de cet ajustement avec les autres ajustements obtenus, voir les tableaux R.1 et R.2 des § R.1 c et R.1.e ci dessus.

(U.12) Cette limitation a été imposée par la capacité de l'ordinateur utilisé.

(U.13) Voir les § R.2.a et R.2.b ci dessus.

σ_T correspond à l'écart type des coefficients d'autocorrélation d'une suite de termes *indépendants et normalement distribués*. En fait et ainsi qu'il a été indiqué (§ R.2.a) il y a lieu de tenir compte d'une correction du fait que l'écart type des r_q est calculé à partir des $Q = 20$ premières valeurs de r_q . Cette correction est faite dans le Tableau U.2. Elle est négligeable pour $Q = 20$.

En considérant également les 20, les 50, les 100, les 200 et les 600 premières valeurs de r_q on a le tableau suivant pour les écarts types ⁽¹⁰⁾.

TABLEAU U.2

Q	20	50	100	200	600
$\sigma(e_q^*) = 1/\sqrt{2l} \quad (l=200)$	0,050	0,050	0,050	0,050	0,050
$\sigma(r_q)$	0,052	0,058	0,060	0,057	0,072
$\sigma_T \sim 1/\sqrt{700 - (Q/2)}$	0,038	0,038	0,039	0,041	0,050

On voit que tous les coefficients d'autocorrélation des $X(n)$ sont bien *du même ordre de grandeur* que ceux correspondant à une suite de valeurs normalement distribuées et indépendantes et que par suite ils en sont *pratiquement indiscernables*.

b) *Ajustement de la distribution des coefficients de corrélation r_q suivant la loi normale*

La distribution des 600 coefficients de corrélation r_q calculés est bien représentée par la loi normale. On trouve par la méthode de Henri un écart type 0,071 et une moyenne $5,8 \cdot 10^{-4}$ de r_q ⁽¹⁴⁾.

Pour apprécier la valeur de l'écart type 0,071 de r_q il convient de la rapprocher de la valeur 0,05 correspondant au calcul des 600 premiers coefficients de corrélation dans une série de 700 termes indépendants ⁽¹⁵⁾. On voit qu'au regard du nombre de termes considérés *la série X(n) se présente avec toutes les apparences d'une série de termes indépendants* ⁽¹⁶⁾.

c) *Considération de tranches de la série des X(n)*

Si on considère les M premières valeurs de la série des $X(n)$ et que l'on calcule les $S = M/2$ premiers coefficients d'autocorrélation r_q correspondants, leur écart type σ_r , *dans l'hypothèse d'une série dénuée de toute autocorrélation*, a pour valeur approximative $\sigma_T = \sim 1/\sqrt{M - S/2} = \sqrt{4/3} M$ ⁽¹⁷⁾. Pour chaque valeur de M l'écart type σ_r calculé peut être confronté à sa valeur théorique *dans l'hypothèse de l'indépendance*. Voici pour différentes valeurs de M les valeurs théoriques σ_T et calculées σ_r , de l'écart type de r_q .

TABLEAU U.3

M	20	40	60	100	150	200	300	400	700
$\sigma_T \sim \sqrt{4/3} M$	0,258	0,183	0,149	0,115	0,094	0,082	0,067	0,058	0,044
σ_r	0,249	0,169	0,133	0,112	0,112	0,093	0,078	0,074	0,061

(U.14) Pour abrégé, ce graphique n'est pas représenté. Le graphique de Henri se caractérise par la valeur $1 - R^2 = 1,36 \cdot 10^{-3}$.

(U.15) Voir le tableau ci-dessus. On a dans l'hypothèse de l'indépendance $\sigma_T = \sqrt{700 - (600/2)} = 0,05$.

En fait en raison des valeurs particulières considérées $l = 200$, $N' = 700$, où N' est le nombre des valeurs de $X(n)$ utilisées pour le calcul des r_q , on a :

$\sigma(e_q^*) = \sigma_T$ pour $Q = 600$ puisque (§ R.2.a ci-dessus) :

$\sigma(e_q^*) = 1/\sqrt{2l} = 0,05 \quad \sigma_T = \sqrt{N' - Q/2} = \sqrt{700 - 300} = 0,05$

(U.16) Il convient en tout cas de préciser ici que les écarts types de e_q^* et r_q ne sont pas rigoureusement comparables. Le premier correspond à une valeur infinie de N et à une suite infinie de valeurs de q . Le second correspond à une valeur finie de N et à une borne supérieure Q également finie (inférieure en tout cas à N) des valeurs de q .

(U.17) Voir § R.2.a et R.2.b ci-dessus.

On voit que les ordres de grandeur des coefficients de corrélation calculés diffèrent très peu des ordres de grandeur correspondant à ce que l'on constaterait si les $X(n)$ étaient des variables aléatoires indépendantes normalement distribuées ⁽¹⁸⁾. Ils en diffèrent d'autant moins que les σ , sont calculés à partir d'un nombre plus faible de valeurs successives de $X(n)$ ⁽¹⁹⁾.

Si on considère maintenant les 25 tranches de $M = 40$ valeurs de la série des 1.000 valeurs de $X(n)$ et si pour chacune de ces tranches on calcule les $S = M/2 = 20$ premiers coefficients d'autocorrélation des $X(n)$ correspondants, on trouve pour la moyenne des écarts types la valeur 0,190 à rapprocher de la valeur approximative $1/\sqrt{40 - 20/2} = 0,183$ correspondant à l'hypothèse de normalité et d'indépendance.

U.5. Valeurs cumulées $Y(n)$ des $X(n)$ et $\pi^*(q)$ des q_q^*

Les séries des $X(n)$ et des q_q^* se présentant avec toutes les apparences de suites de termes indépendants, il est intéressant d'examiner le comportement de leurs valeurs cumulées.

a) Valeurs cumulées $Y(n)$ des $X(n)$

Le *Graphique U.6* représente les valeurs cumulées

$$Y(n) = \sum_{p=1}^n X(p) \quad 1 \leq n \leq 1000 \quad (4)$$

Trois critères permettent d'apprécier dans quelle mesure les $X(n)$ peuvent être considérés comme indépendants et normalement distribués ⁽²⁰⁾.

a.1) Chenaux correspondant à des valeurs données des écarts

Si les $X(n)$ étaient indépendants et normalement distribués on aurait

$$\sigma_Y^*(n) = \sigma_X^* \sqrt{n} \quad \sigma_X^* = \sqrt{1/2} = 10 \quad (5)$$

Sur le *Graphique U.6* sont dessinées les courbes ⁽²¹⁾

$$\alpha_n = \pm \sigma_X^* \sqrt{n} \quad \beta_n = \pm 2 \sigma_X^* \sqrt{n} \quad (6)$$

(U. 18) Naturellement pour des valeurs extrêmement grandes de M la valeur de $\sigma_T = \sqrt{4/3 M}$ tendrait vers zéro alors que la valeur de σ tendrait vers $1/\sqrt{2} = 1/\sqrt{400} = 0,05$. Des écarts significatifs apparaîtraient alors entre les deux séries démontrant la nature non aléatoire effective de la série des $X(n)$, mais les séries observées dans la nature ne sont jamais suffisamment longues pour que de telles différences systématiques puissent apparaître dans le cas où leur structure serait presque périodique.

En tout état de cause, pour des valeurs très grandes à la fois du nombre l de sinusoides et du nombre N de valeurs considérées, les deux écarts types σ_T et σ , deviendraient réellement indiscernables. Il en serait notamment ainsi pour des valeurs de l supérieures à 5 000 et des valeurs de N supérieures à 10 000.

(U.19) Si on considère les coefficients d'autocorrélation r'_q de la série des q_q^* on obtient le tableau suivant :

Tableau U.7

M	20	40	60	100	150	200	300	400	700
$\sigma_T \sim \sqrt{4/3 M}$	0,258	0,183	0,149	0,115	0,094	0,082	0,067	0,058	0,044
σ'_r	0,269	0,173	0,136	0,111	0,091	0,087	0,077	0,068	0,063

Ici encore tout se passe comme si les termes successifs des sommes de sinusoides qui représentent les coefficients d'autocorrélation asymptotiques q_q^* étaient normalement distribués et indépendants.

(U.20) § R.4 ci-dessus.

(U.21) § R.4.a ci-dessus.

De la table de la loi normale il résulte que 68 % des points de la trajectoire $Y(n)$ sont à l'intérieur du chenal des courbes α et que 95,4 % des points de la trajectoire $Y(n)$ sont à l'intérieur des courbes β . On voit ainsi que la courbe représentative de $Y(n)$ est *tout à fait compatible avec l'hypothèse d'indépendance pour quiconque ignorerait la structure effective de la série $Y(n)$* .

a.2) *Maximum de $Y(n)$ dans l'intervalle $1 \leq n \leq N = 1.000$*

Soit M_Y la valeur maximale de $Y(n)$ dans l'intervalle $1 \leq n \leq N = 1.000$

On a (*Graphique U.6*)

$$M_Y = 469 \quad \text{pour} \quad n = 495 \quad (7)$$

On a ainsi

$$\lambda_1 = M_Y / \sigma_x^* \sqrt{N} = 469 / 10 \sqrt{1000} = 1,483 \quad (8)$$

On en déduit que *dans l'hypothèse de normalité et d'indépendance des $X(n)$ la fréquence d'une valeur maximale supérieure ou égale à 469 dans l'intervalle $1 \leq n \leq 1.000$ est égale à (22)*

$$F_1 = \text{Fréquence} \left\{ \frac{M_Y}{\sigma_x^* \sqrt{N}} \geq u \right\} = 2 \int_u^{+\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = 13,8\% \quad (9)$$

Le maximum constaté n'aurait donc rien d'anormal dans l'hypothèse où les $X(n)$ seraient considérées comme des variables aléatoires normales indépendantes d'écart type égal à $\sigma_x^ = 10$.*

a.3) *Distance entre deux passages successifs par zéro de $Y(n)$*

On voit sur le *Graphique U.6* que $Y(n)$, nul pour une valeur de n comprise entre 41 et 42, reste constamment positif pour $42 \leq n \leq 1.000$. *Une telle circonstance devrait-elle être considérée comme anormale au regard de l'hypothèse de normalité et d'indépendance des $X(n)$?*

On a

$$Y(n) = 32,9 \quad \text{pour} \quad n = N = 1.000$$

Si on considère l'horizontale d'ordonnée $Y = 32,9$ elle coupe la courbe $Y(n)$ pour $n \sim 99$. Il résulte de là que si on considère les valeurs cumulées $Y(n)$, *non pas à partir de la valeur $n = 1$, mais à partir de la valeur $n'_1 = 99$* , la courbe des valeurs cumulées Y'_n reste constamment au-dessus du nouvel axe des n' de $n' = 1$ à $n' = 901$. On peut alors appliquer la formule (23)

$$F(N') \sim 2 \sqrt{2/\pi} / \sqrt{N'} \quad (10)$$

On trouve ainsi pour la fréquence des trajectoires Y'_n dont le premier retour à zéro a lieu pour une valeur N' de n supérieure ou égale à 900

$$F_2 = \text{Fréquence} \{N' \geq 900\} \sim 2 \sqrt{2/\pi} / \sqrt{900} \sim 5,3\% \quad (11)$$

valeur tout à fait compatible avec l'hypothèse de normalité et d'indépendance des $X(n)$ suivant les seuils habituellement admis de signification.

Comme $Y(n)$ s'annule pour $n = 41$, la courbe $Y(n)$ appartient au faisceau des trajectoires qui coupent l'axe des n pour $n \geq 1.000$. La fréquence de ces trajectoires est suivant la même formule

$$F_3 \sim 2 \sqrt{2/\pi} / \sqrt{1000 - 41} = 5,16\% \quad (12)$$

(U.22) § R.4.b ci-dessus.

(U.23) Voir ci-dessus § R.4.c., relation (45).

Au regard de l'hypothèse de normalité et d'interdépendance des $X(n)$ la configuration de la courbe $Y(n)$ du Graphique U.6 ne présenterait donc aucun caractère anormal pour tout analyste qui ne connaîtrait pas la structure effective des $Y(n)$.

En fait c'est là une circonstance générale que les valeurs cumulées d'une variable aléatoire indépendante tendent à se trouver en général d'un même côté de l'axe des n pendant des durées relativement longues.

Il convient de souligner en effet que d'après la formulation ci-dessus 10 % des trajectoires $Y(n)$ correspondant à l'hypothèse de normalité et d'indépendance restent constamment au-dessus ou au-dessous de l'axe des n pendant une durée

$$N \geq (2\sqrt{2/\pi}/0,1)^2 = 255 \quad (13)$$

En fait on a le Tableau suivant pour la distribution des distances $n_2 - n_1$ entre deux retours à zéro consécutifs de $Y(n)$ dans l'hypothèse d'indépendance et de normalité.

TABLEAU U.4

$F = \% \text{ des valeurs de } n_2 - n_1 \geq d$	75 %	50 %	25 %	10 %	7,5 %	5 %	2 %	1 %
$d = 8/\pi F^2$	4,53	10,2	40,7	255	453	1.018	6.366	25.463

On voit que pour 1 % des trajectoires le premier retour à zéro a lieu pour une valeur de n supérieure à 25 463.

On voit ainsi que le retour à zéro de $Y(n)$ se fait généralement pour des valeurs de n relativement petites, mais que pour des pourcentages importants des trajectoires le retour à la valeur zéro n'a lieu que pour des valeurs de n très élevées. En fait la moyenne de la valeur de n pour laquelle a lieu le premier retour à zéro est infinie ⁽²⁴⁾.

b) Valeurs cumulées $\pi^*(q)$ des coefficients d'autocorrélation asymptotiques ρ_q^*

Le Graphique U.7 représente la valeur cumulée

$$\pi^*(q) = \sum_{p=1}^q \rho_p^* \quad 1 \leq q \leq 1.000 \quad (14)$$

Au regard des trois critères considérés et des commentaires qui viennent d'être donnés pour la courbe représentative des $Y(n)$ du Graphique U.6 la courbe représentative de $\pi^*(q)$ pour $1 \leq q \leq Q = 1.000$ doit être considérée *a fortiori* comme tout à fait compatible avec l'hypothèse de normalité et d'indépendance des ρ_q^* .

On peut donc en conclure qu'au regard de ces trois critères et pour tout analyste qui ignorerait leur structure effective les coefficients d'autocorrélation asymptotiques ρ_q^* peuvent être considérés dans l'intervalle considéré $1 \leq n \leq N$ comme des variables indépendantes et normalement distribuées.

c) Limitation des intervalles ($1 \leq n \leq N = 1.000$) et ($1 \leq q \leq Q = 1.000$) considérés.

Naturellement la simulation de séries de termes indépendants et normalement distribués par les valeurs de $X(n)$ et de ρ_q^* au regard des tests des valeurs cumulées ne vaut qu'en raison de la limitation des

(U.24) Ce qui confirme ce résultat général qu'il y a effectivement « compensation » mais que cette compensation prend en moyenne un temps infini (voir ALLAIS, *Fréquence, Probabilité et Hasard*, J.S.S.P., tome 124, n° 2, 1983, § 7, p. 83).

C'est ce que confirment encore les résultats empiriques de Wold et Feller (note R.21 ci dessus).

valeurs de N et Q considérées. Pour des valeurs plus élevées, la structure presque périodique de ces séries finirait par apparaître pour quiconque ne connaîtrait pas leur structure presque périodique effective ⁽²⁵⁾.

U.6. Distribution des suites de valeurs positives ou négatives pour $X(n)$ et q_q^*

Les suites de valeurs positives et négatives de $X(n)$ et q_q^* se distribuent comme si les variables correspondantes pouvaient être considérées valablement comme des variables aléatoires indépendantes et normalement distribuées ⁽²⁶⁾. Voici en effet pour $N = 1.000$ et $Q = 1.000$ les nombres r_s , constatés rapprochés de leurs valeurs q_s^* théoriques :

TABLEAU U.5

Distribution des suites de valeurs positives ou négatives

$$1 \leq n \leq N = 1.000 \quad 1 \leq q \leq Q = 1.000 \quad v = 6$$

Séries	s	1	2	3	4	5	$6 \leq s$	$1 \leq s$	χ^2	P
		r_s^*	250,5	125,12	62,50	31,22	15,59	15,56	500,5	6
$X(n)$	r_s	259	121	75	26	18	12	511	4,98	0,54
q_r^*	r_s	262	127	62	32	16	11	510	1,93	0,91

Les valeurs de χ^2 déduites de ce tableau pour les $X(n)$ et les q_q^* sont respectivement de 4,98 et 1,93, auxquelles correspondent, pour $v = 6$ degrés de liberté, les probabilités $P = 0,54$ et $P = 0,91$, valeurs tout à fait « vraisemblables » au regard de l'hypothèse de normalité et d'indépendance.

Comme on ne dispose que de 600 valeurs pour la série des r_q et que seules sont réellement significatives les 100 premières valeurs de r_q ⁽²⁷⁾, voici le tableau correspondant de la distribution des suites de valeurs positives et négatives de r_q correspondant à $1 \leq q \leq Q = 100$ où sont également indiquées pour faciliter les comparaisons, les distributions des suites correspondant aux 100 premières valeurs de $X(n)$ et q_q^* ($1 \leq n \leq 100$). Il y a ici $v = 3$ degrés de liberté ⁽²⁸⁾.

(U.25) Pour des valeurs de N supérieures à 3 000 les graphiques représentatifs de $Y(n)$ et de $\pi^*(q)$ mettent en évidence une oscillation de longue période et de grande amplitude

Dans le cas de la série $Y(n)$ cette oscillation résulte du fait que parmi les $l = 200$ fréquences f_j il y en a une, la fréquence correspondant à $j = 152$, qui est très voisine de l'unité. Il en résulte que la composante sinusoidale correspondante de $Y(n)$ a à la fois une très grande période d'environ 1613 et une très grande amplitude d'environ 257. Trois autres fréquences sont également voisines de l'unité tout en l'étant nettement moins. Les valeurs de j correspondantes sont 30, 92 et 141.

Pour $j = 152$ on a

$$f_j = 27 - 0,00062 \quad T_j^* = 1/0,00062 = 1612,9 \quad a_j^* = a_j/2 \sin \pi f_j = 256,7$$

Pour les trois autres fréquences, on a :

$$\begin{array}{llll} j = 30 & f_j = 11,9948 & T_j^* = 191,9 & a_j^* = 30,55 \\ j = 92 & f_j = 21,0052 & T_j^* = 194,0 & a_j^* = 30,91 \\ j = 141 & f_j = 26,0041 & T_j^* = 245,0 & a_j^* = 39,01 \end{array}$$

L'influence de ces périodes apparaît sur le Graphique U.6

Des observations analogues peuvent être présentées pour la série des $\pi^*(q)$. Sur ces circonstances voir § P.13 et T.5 ci dessus.

(U.26) Voir ci dessus § R.3. Le cas examiné ici correspond exactement à l'exemple traité dans le § R.3.

(U.27) Voir § U.4 ci dessus.

(U.28) D'après les relations (30) et (32) du § R.3 on a ici

$$r_1^* = (100 + 3 - 1) 2^2 = 25,5 \quad r_2^* = (100 + 3 - 2) 2^3 = 12,625$$

$$R_{3,100}^* = (100 - 3 + 2) 2^3 = 12,375$$

Le calcul des χ^2 est effectué d'après la relation (33) du § R.3.

TABLEAU U.6

Distribution des suites de valeurs positives ou négatives

$$1 \leq n \leq 100 \quad 1 \leq q \leq 100 \quad \nu = 3$$

Séries	s	1	2	$3 \leq s$	$1 \leq s$	χ^2	P
		r_s^*	25,5	12,625	12,375	50,5	3
X_n	r_s	25	13	14	52	0,234	0,97
q_q^*	r_s	25	22	10	57	7,43	0,06
r_q	r_s	24	13	15	52	0,656	0,88

Ici encore les valeurs de χ^2 doivent être considérées comme tout à fait « vraisemblables » au regard de l'hypothèse de normalité et d'indépendance des $X(n)$.

U.7. *Commentaire général*

Il résulte ainsi de l'ensemble des résultats qui précèdent que pour tout analyste *qui ignorerait la structure effective* des séries considérées de 1.000 valeurs des $X(n)$ et q_q^* , ces séries apparaîtraient *comme présentant tous les caractères de séries aléatoires constituées de séries indépendantes et normalement distribuées*.

On peut ainsi en conclure qu'une série de longueur limitée constituée par les valeurs successives d'une somme de sinusoides peut présenter, au regard de tous les critères de l'analyse statistique, toutes les apparences d'une suite constituée par des termes aléatoires indépendants et normalement distribués. Il ne s'agit d'ailleurs là en aucune façon d'un cas exceptionnel ⁽²⁹⁾.

ERRATUM

Il convient de rectifier ainsi qu'il suit le Tableau de la page 95 de la Première Partie de ce mémoire (*J.S.S.P.*, 1983, n° 2) : *il faut lire 6 au lieu de 5 au-dessous du point M_{42} de la Figure 2.*

POST-SCRIPTUM

Dès maintenant je remercie vivement tous ceux qui à la suite de ma Communication du 16 mars 1983 et de la publication de la Première Partie de ce Mémoire dans le *J.S.S.P.* (Tome 124, n° 2, 1983) m'ont fait part de leurs observations. Ces dernières seront publiées prochainement avec mes commentaires dans le *J.S.S.P.*, ainsi que quelques compléments que je n'ai pu insérer dans ce mémoire faute de place.

(U.29) Le fait qu'il s'agisse de séries de *longueur limitée* est naturellement tout à fait essentiel (voir § P.11 et P.12 ci-dessus).

TABLEAU U.1 (§ U.1)

SÉRIE U

Série artificielle constituée par les sommes de 200 sinusoides

$$l = 200 \quad N = 1\,000$$

Données structurelles ⁽¹⁾ ⁽²⁾

$$X(n) = \sum_{j=1}^l x_j(n) \quad x_j(n) = a_j \cos 2\pi f_j(n - n_j) \quad 1 \leq n \leq N = 1\,000$$

$$1 \leq j \leq l = 200 \quad a_j = 1 \quad n_j = j$$

$$f_j = \alpha \sqrt{j} \quad \alpha = \sqrt[4]{23}$$

$$T_j = 1/f_j \quad 0,0323 < T_j < 0,456 \quad \text{Médiane de } T_j = 0,0457$$

Coefficients caractéristiques ⁽³⁾

$$A = \sum_j a_j = l = 200 \quad a = \sum a_j / l = 1 \quad A / \sigma_x^* = 20$$

$$\sigma_x^* = \sqrt{\sum_j a_j^2 / 2} = \sqrt{l/2} = 10 \quad \mathcal{L} / \delta = \sqrt[3]{\sum_j a_j^3} / \sqrt[2]{\sum_j a_j^2} = 1 / \sqrt[6]{200} = 0,414$$

$$\sigma(\varrho_q^*) = \sqrt{\sum_j a_j^4 / 2} / \sum_j a_j^2 = 1 / \sqrt{2l} = 0,05 \quad \mathcal{L}_1 / \delta = \sqrt[3]{\sum_j a_j^6} / \sqrt[2]{\sum_j a_j^4} = 1 / \sqrt[6]{l} = 0,414$$

Analyse de la Série U

Graphique U.1 : Représentation des $X(n)$ Graphique U.2 : Distribution des $X(n)$ Graphiques U.3 et U.4 : Corrélogrammes asymptotique ϱ_q^* et calculé r_q des $X(n)$ Graphique U.5 : Distribution des coefficients de corrélation ϱ_q^* Graphique U.6 : Représentation des valeurs cumulées $Y(n) = \sum_{p=1}^n X(p)$ Graphique U.7 : Représentation des valeurs cumulées $\pi^*(q) = \sum_{p=1}^q \varrho_q^*$

1. Voir § P.4 et U.1 ci-dessus.

2. D'après le théorème d'équivalence (§ Q.16 ci-dessus) les valeurs $x_{i,n}$ et $X(n)$ restent inchangées si on prend par exemple

$$\theta_j = 1/[1/T_j]_d$$

où $[1/T_j]_d$ représente la partie décimale de $1/T_j$. On a alors

$$1 < \theta_j < 246 \quad \text{Médiane de } \theta_j \sim 2$$

16 % des valeurs de θ_j sont supérieures à 5,5, et 10 % sont supérieures à 10.

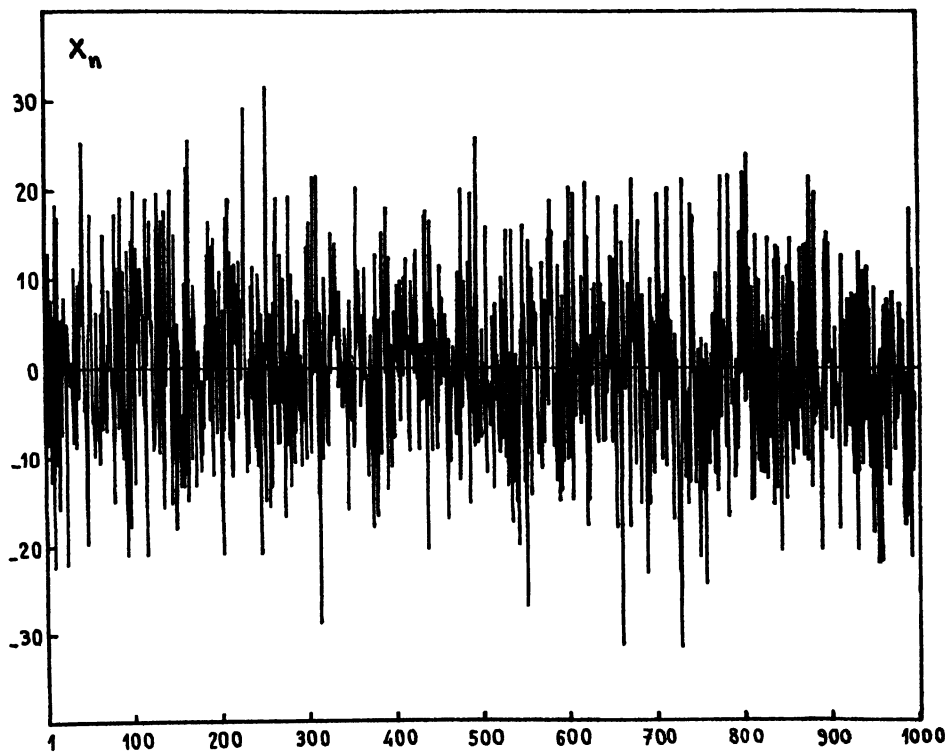
3. § Q.9 et Q.10 ci-dessus.

GRAPHIQUE U.1 (§ U.1)

SÉRIE U

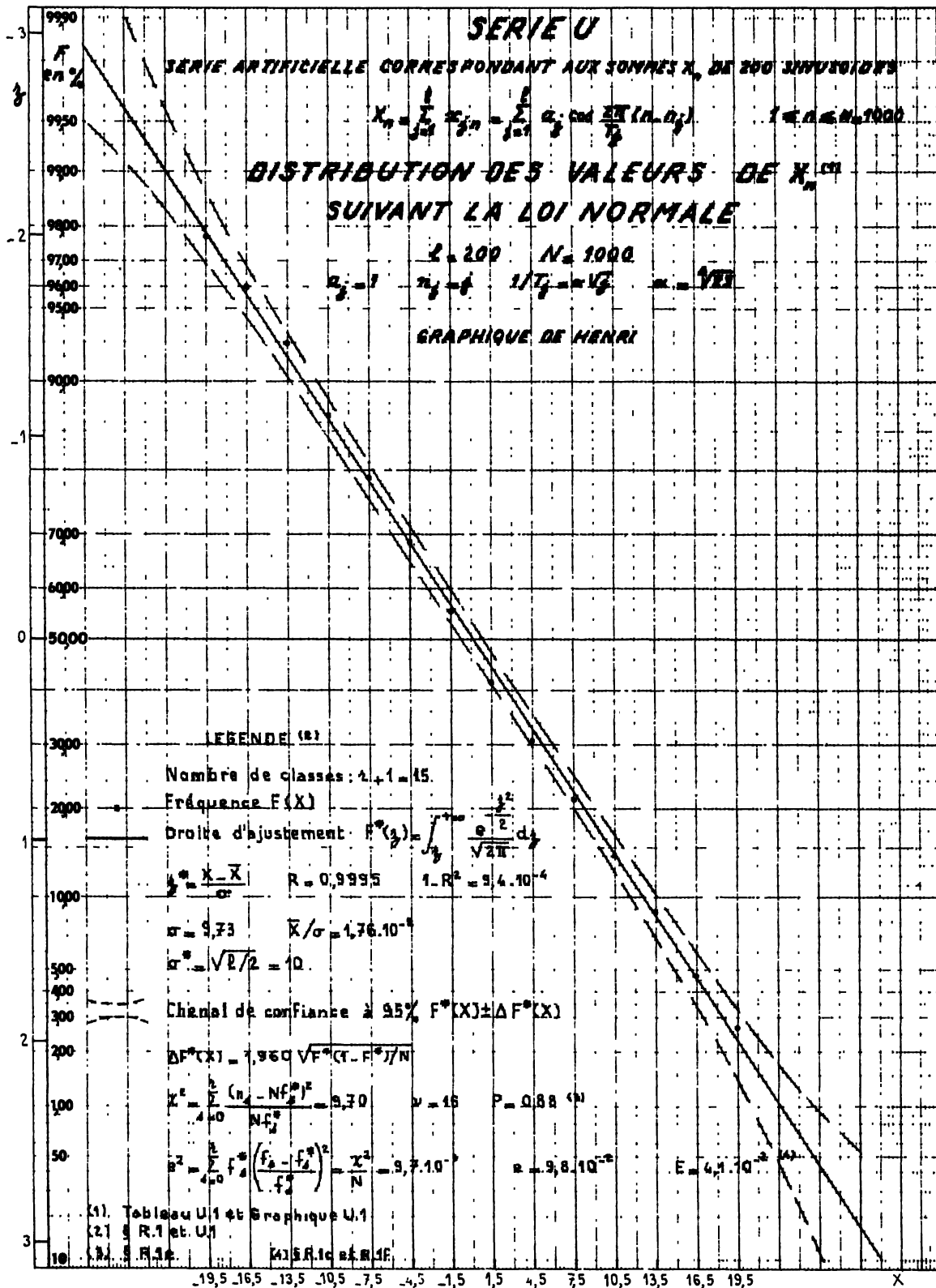
Série artificielle correspondant
aux $N = 1.000$ sommes $X(n)$ de $l = 200$ sinusôides

$$X(n) = \sum_{j=1}^l x_j(n) = \sum_{j=1}^l a_j \cos 2 \pi f_j (n - n_j)$$
$$1 \leq j \leq l \quad 1 \leq n \leq N \quad l = 200 \quad N = 1000$$



LÉGENDE : voir Tableau U.1
(valeurs des $X(n)$: Graphique 8332).

GRAPHIQUE U.2 (§ U.2)

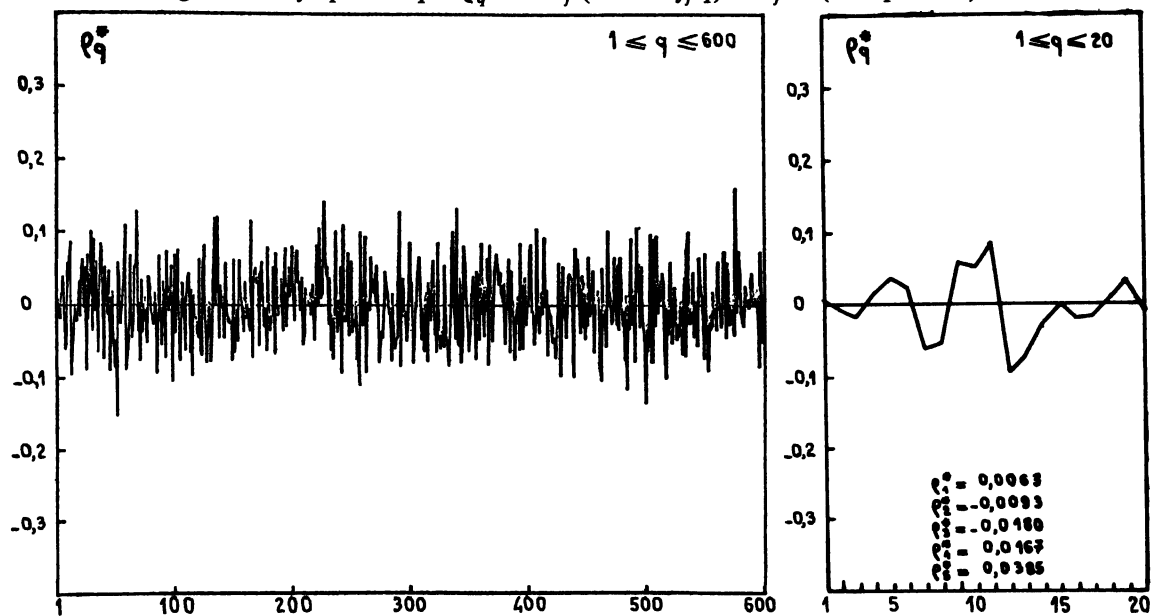


GRAPHIQUES U.3 ET U.4 (§ U.3-4)

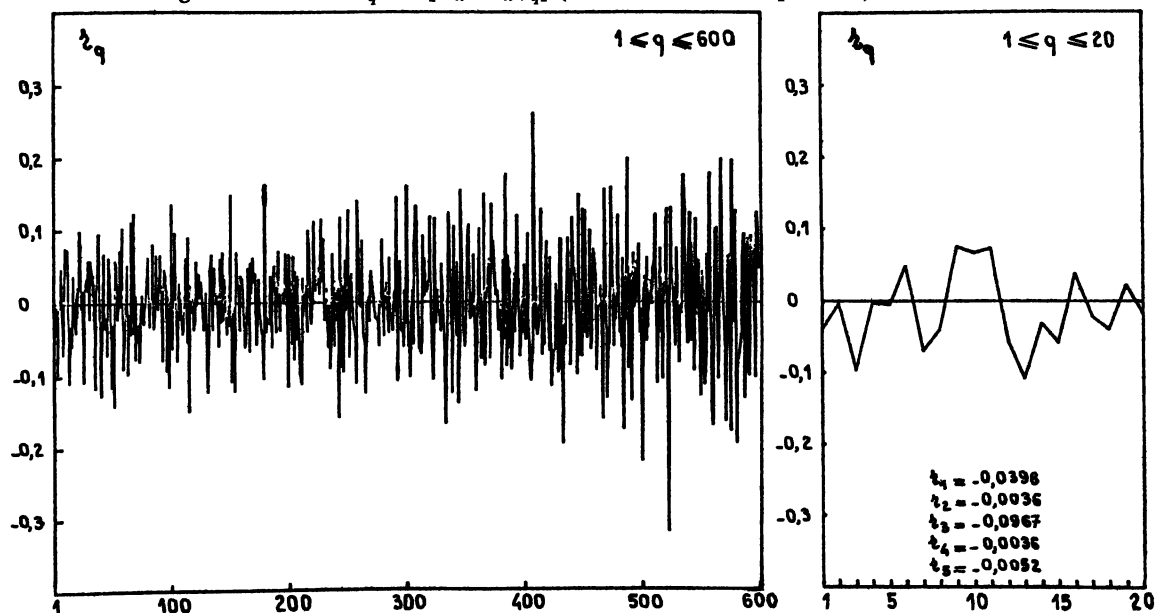
SÉRIE U

Série artificielle correspondant
aux $N = 1\ 000$ sommes $X(n)$ de $I = 200$ sinusoides
Corrélogrammes asymptotique et calculé

G.U.3. Corrélogramme asymptotique $\rho_q^* = \sum a_j^2 (\cos 2\pi f_j q) / \sum a_j^2$ (1) ($1 \leq q \leq 600$)



G.U.4. Corrélogramme calculé $r_q = r[X_n, X_{n+q}]$ ($1 \leq n \leq 700$; $1 \leq q \leq 600$)

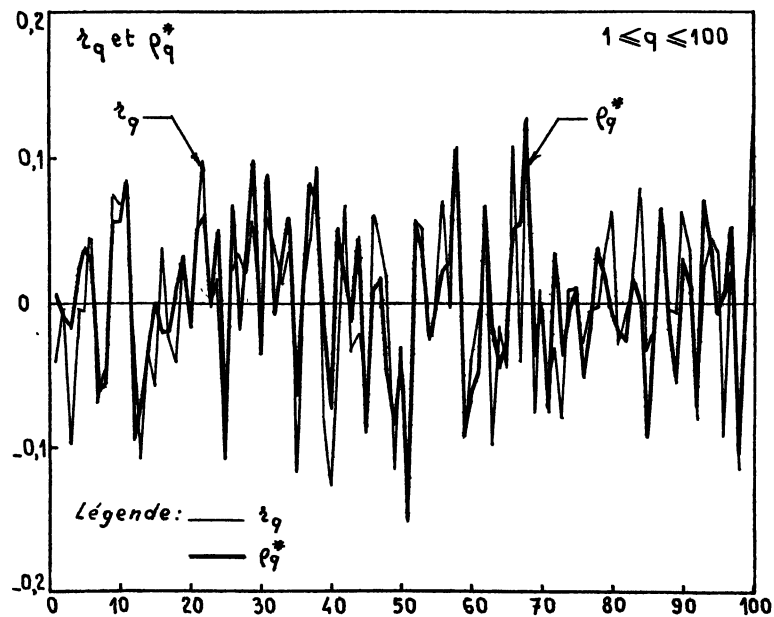
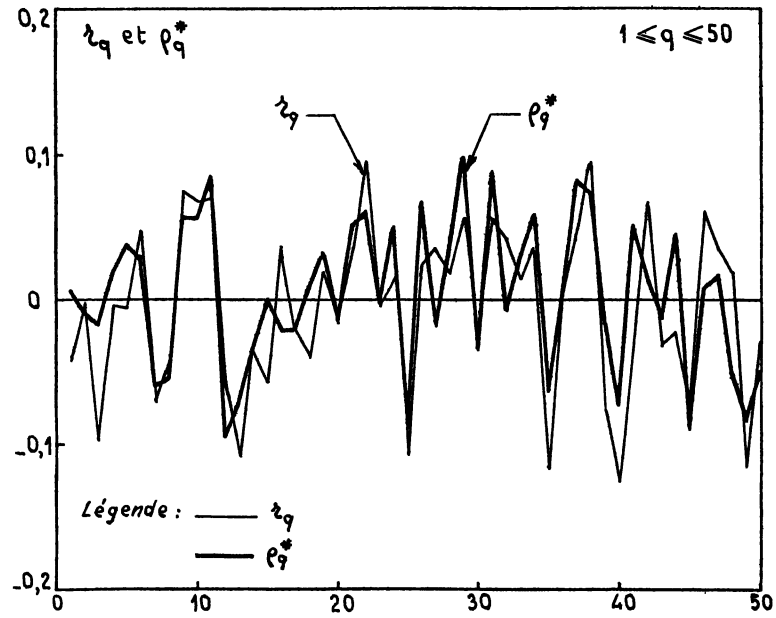


(1) Voir § Q.10, relation (56).

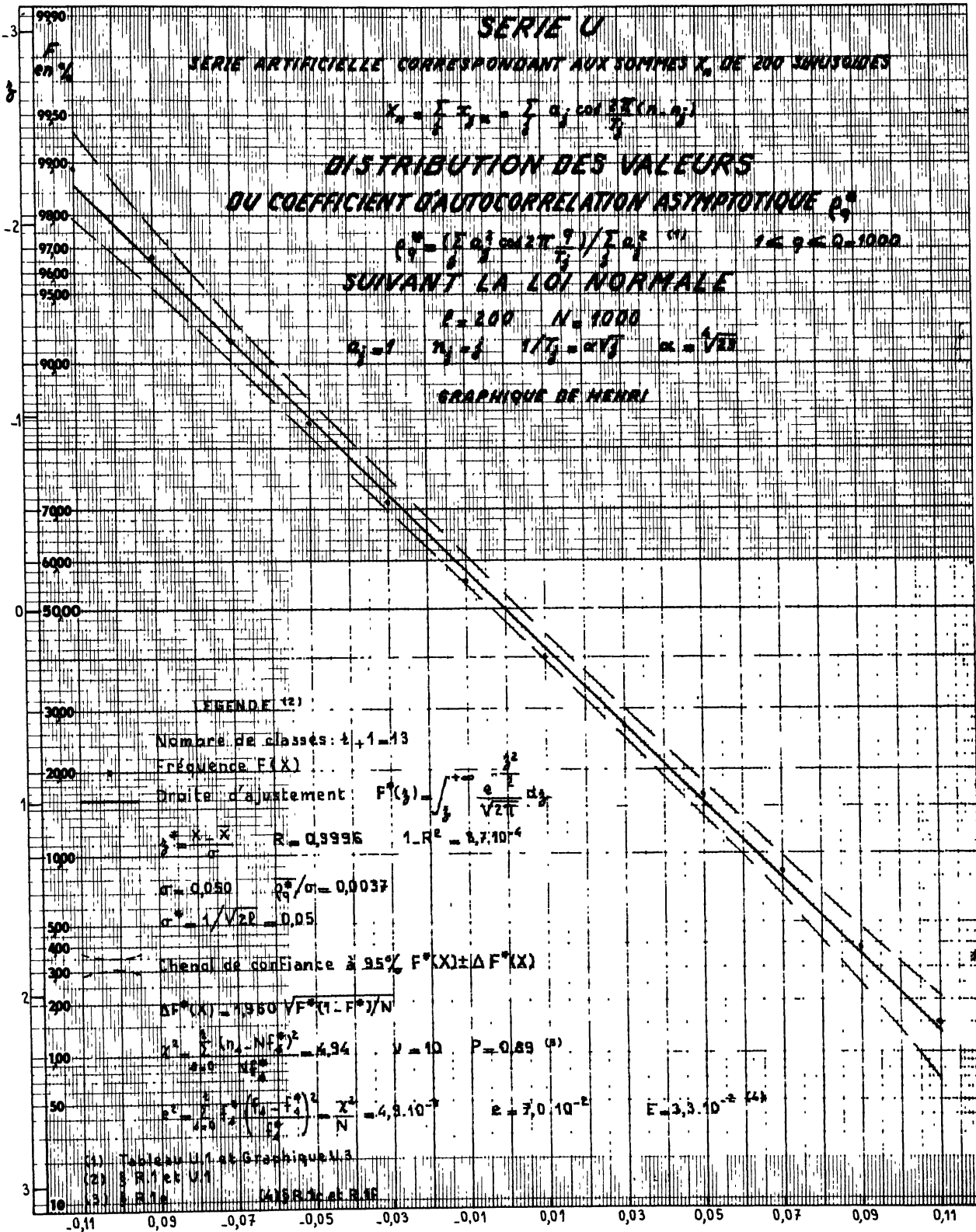
GRAPHIQUES U.3* ET U.4* (§ U.4)

SÉRIE U

Série artificielle correspondant
aux $N = 1000$ sommes $X(n)$ de $I = 200$ sinusôides
Corrélogrammes asymptotique et calculé (¹)

1. A partir des 700 premières valeurs de $X(n)$.

GRAPHIQUE U.5 (§ U.3)



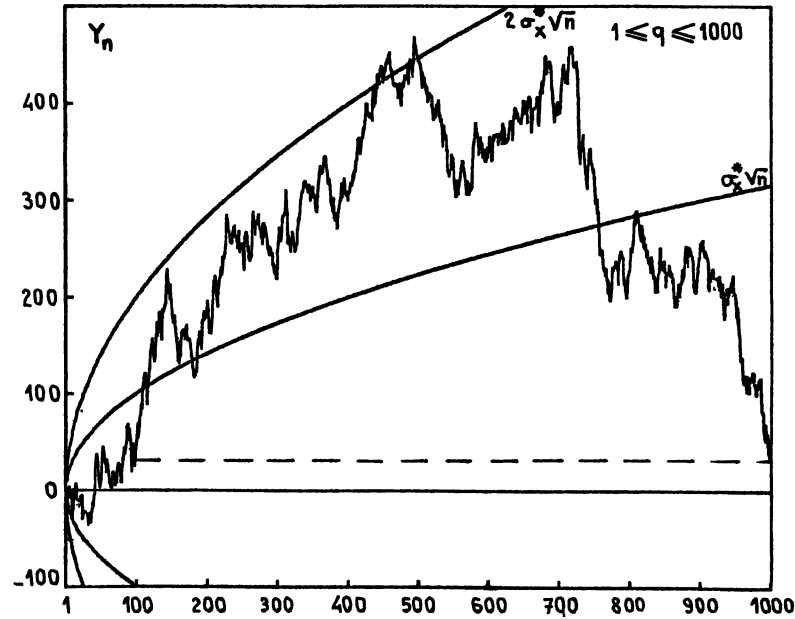
GRAPHIQUES U.6 ET U.7 (§ U.5)

SÉRIE U

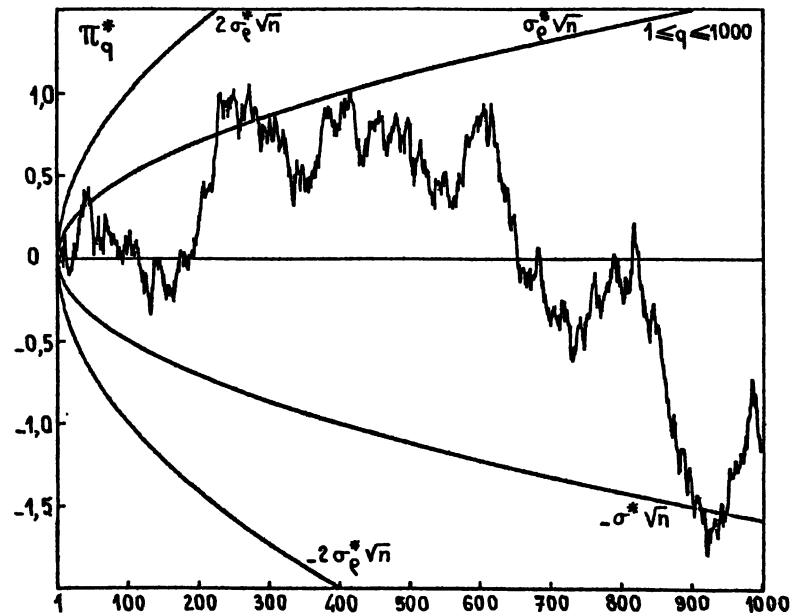
Série artificielle correspondant
aux $N = 1\,000$ sommes $X(n)$ de $l = 200$ sinusoides

$$X(n) = \sum_{j=1}^l \cos 2\pi f_j (n - n_j) \quad \rho^* = \frac{1}{l} \sum_{j=1}^l \cos 2\pi f_j q \quad 1 \leq n \leq N = 1000, 1 \leq j \leq l = 200$$

G.U.6. Valeurs cumulées des $X(n)$: $Y(n) = \sum_{p=1}^n X_p \quad (1 \leq n \leq 1000, \sigma_X^* = \sqrt{l}/2 = 10)$



G.U.7. Valeurs cumulées des ϱ_q^* : $\pi^*(q) = \sum_{p=1}^q \varrho_p^* \quad (1 \leq q \leq Q = 1000, \sigma_\varrho^* = 1/\sqrt{2l} = 0,05)$



RÉFÉRENCES

- ALLAIS M. : (1954), Évaluation des perspectives économiques de la recherche minière sur de grands espaces. Application au Sahara algérien. *Bulletin de l'Institut International de Statistique*, tome XXXV, 4, Rio de Janeiro, 1957, p. 89-140. Traduction anglaise, Revue américaine *Management Science*, vol. 3, n° 4, juillet 1957.
- ALLAIS M. : (1957a), Observation des mouvements du pendule paraconique. *Comptes rendus de l'Académie des Sciences*, 13 novembre 1957, tome 245, n° 20, p. 1697-1700.
- ALLAIS M. : (1957b), Analyse harmonique des mouvements du pendule paraconique. *Comptes rendus de l'Académie des Sciences*, 25 novembre 1957, tome 245, n° 22, p. 1875-1878.
- ALLAIS M. : (1957c), Mouvements du pendule paraconique et éclipse totale de soleil du 20 juin 1954. *Comptes rendus de l'Académie des Sciences*, 4 décembre 1957, tome 245, n° 23, p. 2001-2003.
- ALLAIS M. : (1957d), Généralisation du test de Schuster au cas de séries temporelles autocorrélées. *Comptes rendus de l'Académie des Sciences*, 13 mai 1957, tome 244, p. 2469-2471.
- ALLAIS M. : (1957e), Application du test de Schuster généralisé à l'analyse harmonique des azimuts du pendule paraconique. *Comptes rendus de l'Académie des Sciences*, 23 décembre 1957, tome 245, p. 2467-2470.
- ALLAIS M. : (1957f), Théorie du pendule paraconique et influence lunisolaire. *Comptes rendus de l'Académie des Sciences*, 16 décembre 1957, tome 245, n° 25, p. 2170-2173.
- ALLAIS M. : (1958a), Nouvelles expériences sur le pendule paraconique à support anisotrope. *Comptes rendus de l'Académie des Sciences*, tome 247, 1958, n° 18, p. 1428-1431.
- ALLAIS M. : (1958b), Structure périodique des mouvements du pendule paraconique à support anisotrope à Bougival et Saint-Germain en juillet 1958. *Comptes rendus de l'Académie des Sciences*, tome 247, 1958, p. 2284-2287.
- ALLAIS M. : (1958c), Doit-on reconsidérer les lois de la gravitation? *Perspectives X*, 1958, p. 91-106.
- ALLAIS M. : (1959a), Détermination expérimentale de l'influence de l'anisotropie du support sur le mouvement du pendule paraconique. *Comptes rendus de l'Académie des Sciences*, tome 248, 1959, p. 764-767.
- ALLAIS M. : (1959b), Détermination expérimentale de l'influence de l'inclinaison de la surface portante sur le mouvement du pendule paraconique à support anisotrope. *Comptes rendus de l'Académie des Sciences*, tome 248, 1959, p. 359-362.
- ALLAIS M. : (1959c), Should the Laws of Gravitation be Reconsidered? *Aero-Space Engineering*, Institute of the Aeronautical Sciences, New York, vol. 18, n°s 9, 10, 11; sept., oct., nov. 1959, p. 46-52, 51-55, 55.
- ALLAIS M. : (1960), *Test de Périodicité. Généralisation du test de Schuster au cas de séries temporelles autocorrélées*. Volume collectif : « Studi in Onore di Corrado Gini », Istituto di Statistiche Demografiche et Attuariali. Università degli Studi di Roma.
- ALLAIS M. : (1961), Généralisation du test de Schuster au cas de séries temporelles autocorrélées dans l'hypothèse d'un processus de perturbations aléatoires d'un système stable. Communication au 33^e Congrès de l'Institut international de Statistique, Paris, 1961. *Bulletin de l'Institut International de Statistique*, 1962, tome 39, 2^e livraison, p. 143-194.
- ALLAIS M. : (1978), *Distribution des suites de s faces ou piles en n tirages de pile ou face*. Mémoire miméographié, Centre d'Analyse Économique, Centre National de la Recherche Scientifique, n° 3646, 136 p.
- ALLAIS M. : (1979), *The So-called Allais' Paradox and Rational Decisions under Uncertainty* dans l'ouvrage collectif : *Expected Utility Hypotheses and the Allais Paradox. Contemporary Discussions of Decisions under Uncertainty with Allais' Rejoinder* (714 p.) (M. Allais et O. Hagen, editors, Reidel), p. 434-699.
- ALLAIS M. : (1980), Analyse des séries temporelles. Facteur X. *Rapport d'activité scientifique du Centre d'Analyse Économique*, Centre National de la Recherche Scientifique, septembre 1980, p. 19-39.
- ALLAIS M. : (1982a), *Simulation déterministe du hasard par des fonctions presque périodiques*. Mémoire miméographié, Centre d'Analyse Économique, Centre National de la Recherche Scientifique, n° 4084.
- ALLAIS M. : (1982b), *Illustration des régularités présentées par les parties décimales de nombres irrationnels correspondant à la suite des nombres entiers*. Mémoire miméographié, Centre d'Analyse Économique, Centre National de la Recherche Scientifique, n° 4085, 27 p.

- ALLAIS M. : (1983a), Sur la distribution normale des valeurs à des instants régulièrement espacés d'une somme de sinusoides. *Comptes rendus de l'Académie des Sciences*, 30 mai 1983, tome 296, série I, p. 829-832.
- ALLAIS M. : (1983b), Généralisations et interprétation du Théorème (T). *Comptes rendus de l'Académie des Sciences*, octobre 1983.
- ALLAIS M. : (1983c), Frequency, Probability and Chance. Mémoire publié dans le volume collectif : *Foundations of Utility and Risk Theory with Applications*. B. Stigum and F. Wenstop, edit. Reidel, Dordrecht, 1983.
- ALLAIS M. : (1983d), Fréquence, Probabilité et Hasard. *Journal de la Société de Statistique de Paris*, tome 124, n° 2, 2^e trimestre, 1983, p. 70-102.
- ALLAIS M. : (1983e), *Les fondations de la théorie de l'utilité et du risque. Quelques points centraux des discussions de la Conférence d'Oslo, 1982*. Centre d'Analyse Économique, août 1983, 88 p. Ce mémoire sera publié en anglais dans le second volume de la Conférence, *Progress in Decision Theory* (Hagen and Wenstop, edit., Reidel, 1984 (voir ALLAIS, 1983d, note **, p. 71).
- ALLAIS M. : (1984), *Frequency, Probability and Chance*. Reidel, Dordrecht.
- BACHELIER Louis : (1915), La périodicité du hasard. *L'enseignement mathématique*, Paris, 1915, p. 5-11.
- BACHELIER Louis : (1938), *La spéculation et le calcul des probabilités*. Gauthier-Villars, Paris, 1938, 51 p.
- BACHELIER Louis : (1939), *Les nouvelles méthodes du calcul des probabilités*. Gauthier-Villars, Paris, 1939, 69 p.
- BASS Jean : (1971), *Cours de mathématiques*. Tome III, Masson, Paris, 405 p.
- BAUDEZ Gaston : (1952), *Le hasard et les causes rythmées*. Dunod, Paris, 1952, 170 p.
- BAUDEZ Gaston : (1953), Le hasard et les causes rythmées. *Journal de la Société de Statistique de Paris*, juillet-août 1953, n° 7, 8, 9, p. 169-187.
- BOHL P. : (1983), *Über die Darstellung von Funktionen einer Variablen durch trigonometrische Reihen mit mehrererer einer Variablen proportionalen Argumenten*. Dorpat, 1893.
- BOHR Harald : (1923), Sur les fonctions presque périodiques. *C.R.A.S.*, 22 octobre 1923.
- BOHR Harald et JESSEN Borge : (1949), Mean Motions and Almost Periodic Functions, *Colloques internationaux du Centre National de la Recherche Scientifique, XV : Analyse harmonique*, Nancy, 15-22 juin 1947, C.N.R.S., 1949, p. 75-84.
- BOREL Émile : (1926), *Applications de la théorie des probabilités à l'arithmétique et à la théorie des fonctions*. Fasc. I du tome II du *Traité du calcul des probabilités et de ses applications* d'Émile Borel, Gauthier-Villars, Paris, 1926, 102 p.
- COOTNER Paul H. (Ed.) : (1964), *The Random Character of Stock Market Prices*. The M.I.T. Press, Cambridge, 510 p.
- CORDUNEANU C. : (1968), *Almost Periodic Functions*. Interscience Publishers, Wiley, New York, 1968, 237 p. Cet ouvrage comporte une bibliographie très complète.
- CRAMER Harald : (1937), *Random Variables and Probability Distributions*. Cambridge, 1937 (second edit. 1962), 119 p.
- CRAMER Harald : (1945), *Mathematical Methods of Statistics*. Princeton University Press, 1946, 575 p.
- CRAMER Harald : (1959), *The Elements of Probability and some of its Applications*. Wiley, New York, 281 p.
- DAVIS Harold R. : (1941), *The Analysis of Economic Time Series*. The Principia Press, Bloomington, 1941.
- ESCLANGON Ernest : (1904), *Les fonctions quasi-périodiques*. Thèse présentée et soutenue le 25 juin 1904 pour l'obtention du grade de docteur ès sciences mathématiques. Gauthier-Villars, Paris, 1904, 281 p.
- ESCLANGON Ernest : (1917), Nouvelles recherches sur les fonctions quasi-périodiques. *Annales de l'Observatoire de Bordeaux*, tome XVI, p. 53-223, Gauthier-Villars, Paris.
- FAVARD J. : (1933), *Leçons sur les fonctions presque-périodiques*. Gauthier-Villars, Paris, 1933, 183 p.
- FELLER William : (1950-1971), *An Introduction to Probability Theory and its Applications*. 2 volumes, Wiley, New York, vol. I, 2^e Ed., 1957, 461 p. (third ed. 1968, 509 p.); vol. II, 1966, 626 p. (second ed. 1971, 669 p.).
- FRECHET Maurice : (1937), *Généralités sur les probabilités. Variables aléatoires*. Tome I, fascicule III, premier livre du *Traité du calcul des probabilités et de ses applications* d'Émile Borel, Gauthier-Villars, Paris, 1937, 308 p.
- GRANGER Clive W.J. et MORGENSTERN Oskar : (1970), *Predictability of Stock Market Prices*. Heath Lexington Books, 1970, 303 p.
- GRENDER Ulf et ROSENBLATT Murray : (1957), *Statistical Analysis of Stationary Time Series*. Wiley, 1957, 300 p.
- HALD A. : (1951), *Statistical Theory with Engineering Applications*. New York, John Wiley and sons, Inc., London, Chapman and Hall Ltd, 1952, 783 p.
- HALD A. : (1952), *Statistical Tables and Formulas*. Id., 97 p.

- HARDY G.H. et WRIGHT E.M. (1959), *An Introduction to the Theory of Numbers*, Clarendon Press, Oxford.
- HARTMAN P., van KAMPEN E.R. et WINTNER A : (1937), Mean Motions and Distribution Functions. *American Journal of Mathematics*, Vol. LIX, 1937, p. 261-267.
- JESSEN Borge et TORNEHAVE Hans : (1945), Mean Motions and Zeros of Almost Periodic Functions. *Acta Mathematica*, Vol. 77, p. 137-279, Uppsala, 1945.
- KENDALL Maurice : (1945), On the Analysis of Oscillatory Time Series. *J.R.S.S.*, vol. CVIII, Parts I, II, 1945, p. 93-141.
- KENDALL Maurice : (1946), *Contributions to the Study of Oscillatory Time Series*. Cambridge University Press, Cambridge.
- KENDALL Maurice : (1948), *The Advanced Theory of Statistics*. T. II, Chap. 29-30 : Time Series, p. 363-441, Griffin, London 1948, 521 p.
- KENDALL Maurice G. et SMITH Babington : (1938), Randomness and Random Sampling Numbers. *Journal of the Royal Statistical Society*, vol. 101, 1938, p. 147-166.
- KENDALL Maurice G. et SMITH Babington : (1939), Second Paper on Random Sampling Numbers, id., vol. 6, 1939, p. 51-61.
- KENDALL Maurice G. et SMITH Babington : (1976), *Times Series*. Griffin, London, 197 p.
- KENDALL Maurice G. et STUART Alan : (1966-1976), *The Advanced Theory of Statistics*. Vol. 3 : Design and Analysis, and Time-Series, Chapt. 45-50 : Time Series, p. 356-528. Charles Griffin, 1976, third edit., 585 p. Cet ouvrage comporte une bibliographie étendue.
- KERMACK W.O. et McKENDRICK A.G. : (1937), Tests for Randomness in a Series of Numerical Observations. *Proceedings of the Royal Society of Edinburgh*, Vol. LVII, 1936-1937, p. 228-240.
- KERMACK W.O. et McKENDRICK A.G. : (1937), Some Distributions Associated with a Randomly Arranged Set of Numbers. *Proceedings of the Royal Society of Edinburgh*, vol. LVII, 1936-1937, p. 332-376.
- KOOPMANS L.H. : (1974), *The Spectral Analysis of Time Series*. Academic Press, New York, 366 p.
- LEVY Paul : (1930), Le théorème fondamental de la théorie des erreurs. *Annales de l'Institut Henri-Poincaré*, Paris, 1930, fasc. II, vol. I, p. 163-177.
- LEVY Paul : (1931), Formules relatives au jeu de pile ou face. *Journal de l'École Polytechnique*, 2^e série, Cahier 29, 1931, p. 103-113.
- LEVY Paul : (1937), *Théorie de l'addition des variables aléatoires*. Chapitre I : Les fondements de la notion de probabilité, p. 1-10, Gauthier-Villars, Paris, 331 p.
- Von MISES Richard : (1920), *Probability, Statistics and Truth*. Second English edition, 1956, Allen, London, and Macmillan, New York, 244 p.
- MOOD M. : (1940), The Distribution Theory of Runs. *Annals of Mathematical Statistics*, 1940, p. 367-392.
- MORICE E. et CHARTIER F. : (1954), *Méthode Statistique*. 2 vol., Imprimerie Nationale, Paris, 1954, 187 et 555 p.
- NERLOVE Marc, GREYER DAVID M. et CARVALHO José L. : (1979), *Analysis of Economic Time Series*. Academic Press, New York, 1979, 468 p.
- POINCARÉ Henri : (1906), *La science et l'hypothèse*. Chapitre XI, Le calcul des probabilités, p. 213-244, Bibliothèque de Philosophie Scientifique, Flammarion, Paris, 1927, 292 p.
- POINCARÉ Henri : (1912), *Calcul des probabilités*. 2^e Éd., Gauthier-Villars, Paris, 1923, 336 p.
- RAUZY Gérard : (1976), *Propriétés statistiques de suites arithmétiques*. Presses Universitaires de France, Collection S U P, Paris, 1976, 133 p.
- SCHUREMAN Paul : (1941), *Manual of Harmonic Analysis and Prediction of Tides*. Special Publication n° 98, revised (1940) edition, United States Government Printing Office, Washington, 1941, 317 p.
- TINTNER Gerhard : (1952), *Econometrics*. John Wiley, New York et Chapman and Hall, London, 370 p.
- WOLD Herman : (1938), *A Study in the Analysis of Stationary Time Series*. Almqvist & Wiksell, Stockholm, 2^e édition, Uppsala, 1954, 236 p.
- WOLD Herman : (1953), *Demand Analysis*. Part III : Some Topics in the Theory of Stationary Random Processes, p. 149-186, Wiley, 1953, 358 p.
- WOLD Herman : (1965), A Graphic Introduction to Stochastic Process, p. 7-76 in *Bibliography on Times Series and Stochastic Processes*, Olivier and Boyd, London, 516 p. Cet ouvrage comporte une bibliographie très étendue pour la période 1938-1959 (p. 77-516).