Problèmes de l'analyse spectrale des séries temporelles stationnaires

M. S. Bartlett

▶ **To cite this version:**

M. S. Bartlett. Problèmes de l'analyse spectrale des séries temporelles stationnaires. Annales de l'ISUP, 1954, 3 (3), pp.119-135. hal-04093370

HAL Id: hal-04093370
https://hal.science/hal-04093370

Submitted on 10 May 2023

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

PROBLÈMES DE L'ANALYSE SPECTRALE DES SÉRIES TEMPORELLES STATIONNAIRES

par

M. S. BARTLETT

1. INTRODUCTION : QUELQUES REMARQUES

L'analyse statistique des séries temporelles, et des processus stochastiques en général, ne diffère pas en principe de l'analyse des autres types de calcul statistique. Toutefois, il est bien connu que la déduction statistique doit dépendre en premier lieu des spécifications théoriques appropriées; aussi n'est-il pas surprenant que de récents développements dans la spécification théorique des processus stochastiques, et en particulier des séries temporelles stationnaires, puissent avoir une répercussion considérable sur les méthodes de l'analyse statistique. Il est également bien apparent que, pour les processus stochastiques, ce n'est que ces dernières années qu'on a commencé à prêter une sérieuse attention aux rapports qui existent entre la théorie et la méthode. Il y a une publication fondamentale, bien que très abstraite, qui est de Grenander (1950). Le travail des mathématiciens anglais, avec la note de Yule qui a devancé les autres (1927) sur les séries temporelles, tend à être d'un caractère en quelque sorte plus utilitaire. Suivant la voie tracée par Yule, il y a eu une tendance à passer de l'analyse classique du type de "périodogramme" associée au spectre des séries, à une analyse d'autocorrélation et d'auto-régression associée au "corrélogramme" des séries (voir, par exemple, M.G. Kendall, 1946; Bartlett, 1946). Cela est en accord raisonnable avec l'expérience: bien des séries temporelles appartiennent au type qui présente un spectre continu dans le sens du théorème de Wiener-Khintchine (à voir dans Levy, 1948) et du théorème analogue sur les séries discrètes énoncé par Wold; et, de plus, ces mêmes séries ont une représentation de base autorégressive (Wold 1938) qui peut être finie avec quelques termes seulement, au moins en faisant une approximation. Pour ces modèles finis autorégressifs, et (pour les valeurs discrètes du temps) de la forme:

$$X_t + a_1 X_{t-1} + \ldots\ldots a_k X_{t-k} = Y_t$$

des solutions asymptotiques aux problèmes d'estimation et de la qualité d'ajustement existent pour les résidus normaux Y_t (en supposant qu'on a un écart σ^2 constant et une moyenne nulle), ou, plus généralement, pour les résidus indépendants. Ainsi Mann et Wold (1943) ont donné une solution très satisfaisante du problème d'estimation, aussi bien pour les séries simples que

pour les séries multiples; et des tests de "goodness of fit" ont été exposés par divers auteurs (Quenouille, 1947; Wold, 1949; Bartlett et Diananda, 1950 ; Whittle 1951, 1952, Walker, 1952; et pour les séries multiples, Bartlett et Rajalakshman, 1953; Whittle, 1953; voir pour une revue générale, Bartlett, 1954). Somme toute, la tendance à tout rapporter aux méthodes de corrélation a été jusqu'à un certain point une réaction contre l'insuccès de la technique classique du périodogramme pour ces séries, et contre la découverte, faite parfois empiriquement, que tout essai direct d'estimation de la fonction spectrale engendre des fluctuations qui augmentent, au lieu de diminuer, avec la longueur de la série. Avant de dire le mot final sur les valeurs relatives des méthodes de corrélation ou de périodogramme, il est prudent de revoir cette dernière avec plus de soin. On peut naturellement remarquer que du théorème de Wiener-Khintchine l'analyse de corrélation et l'analyse harmonique sont maintenant bien connues comme un double aspect de la même analyse théorique (on peut montrer qu'il en est de même pour l'analyse arithmétique de la série qui se réalise) et qu'alors la méthode employée est en fait peu importante, supposant toujours une interprétation juste. Il serait prématuré de chercher à porter dès maintenant un jugement définitif, mais il est peu probable que ce soit si simple; le problème de dégager les caractéristiques de données statistiques au moyen de quelques constantes, dans le cas de séries temporelles, se résout au mieux tantôt par les corrélations, tantôt par les périodogrammes. Ainsi, les constantes d'un modèle fini autorégressif seront mieux déterminées par les corrélations; et les termes discrets harmoniques dans une série harmonique classique par le périodogramme. Je me propose ici de discuter de certains problèmes restés en suspens, associés à des périodogrammes ou à des analyses spectrales et j'indiquerai des solutions qui sont plutôt possibles que finales

2. PÉRIODOGRAMMES ET SPECTRES CONTINUS

Pour une série temporelle réelle strictement stationnaire en temps discret[1] X_t , le périodogramme classique est obtenu en calculant (soit pour l'intégrale p) la quantité:

$$G_p \equiv A_p + iB_p = \sqrt{\frac{2}{n}} \sum_{r=1}^{n} X_r \, e^{2\pi i p r/n}$$

(on calcule A_p et B_p séparément), et l'intensité $I_p = G_p G_p^* \equiv A_p^2 + B_p^2$ Le facteur $\sqrt{(2/n)}$ a été introduit pour que la moyenne $E\{I_p\}$ soit égale à $2\sigma^2$ pour une série à termes complètement indépendants. Nous supposons pour simplifier, que $E\{X_t\} = 0$, bien que cela ne soit pas essentiel pour $p \neq 0$. Si maintenant nous définissons la fonction de covariance

$$C_s = \frac{1}{n-s} \sum_{r=1}^{n-s} X_r \, X_{r+s} \, , \qquad (s > 0 \, , \; C_{-s} = C_s)$$

(1) L'analyse d'une série temporelle en temps continu est naturellement semblable en principe: pour une discussion plus serrée de ce sujet, voir Bavenport, Johnson and Middleton (1952)

un calcul assez simple nous montre que:

$$I_p = 2 \sum_{s=-n+1}^{n-1} \left(1 - \frac{|s|}{n}\right) C_s \cos(s\omega_p), \qquad \left(\omega_p = \frac{2\pi p}{n}\right)$$

C'est la relation entre I_p et le corrélogramme C_s/C_o dont on a déjà parlé; c'est cependant la moyenne de cette relation, soit

$$E\{I_p\} = 2\sigma_X^2 \sum_{s=-n+1}^{n-1} \left(1 - \frac{|s|}{n}\right) \rho_s \cos(s\omega_p)$$

qui correspond pour $n \to \infty$ à la relation de Wold entre ρ_s et la fonction spectrale. Quand cette dernière est absolument continue avec, pour la densité $f(\omega)$, $0 \leqslant \omega \leqslant \pi$, cette relation devient

$$2\pi\sigma_X^2 f(\omega_p) = 2\sigma_X^2 \sum_{s=-\infty}^{\infty} \rho_s \cos(s\omega_p)$$

Considérons maintenant les fluctuations de I_p.

On sait que, pour une suite normale indépendante Y_t, ces fluctuations sont représentées par la loi de probabilité:

$$P\{I_p \geqslant Z\} = \exp(-1/2\, Z/\sigma^2)$$

et, en particulier, $\mathrm{var}(I_p) = 4\sigma^4 = [E\{I_p\}]^2$. Même si Y_t n'est pas normale, nous avons encore

$$E\{H_p H_q^*\} = \frac{2\sigma_Y^2}{n} \frac{\left(1 - e^{in(\omega_1 \pm \omega_2)}\right) e^{i(\omega_1 \pm \omega_2)}}{1 - e^{i(\omega_1 \pm \omega_2)}}$$

où $\omega_1 = 2\pi p/n$, $\omega_2 = 2\pi q/n$, le signe $\pm$ correspond à H_q, H_q^*, et $H_p = \sqrt{(2/n)} \sum Y_r e^{2\pi i p r/n}$. Celle-ci disparait avec p et q entiers, ce qui montre que H_p n'est en corrélation ni avec H_q ni avec H_q^*, ou ce qui revient au même, que les parties réelles et imaginaires de H_p et H_q ne sont pas en corrélation. $E\{H_p^2\}$ est aussi nulle, donc les parties réelles et imaginaires de H_p ne sont pas en corrélation et apportent le même carré moyen dans $E\{H_p H_p^*\}$.

De plus, pour les quantités du 2° ordre $J_p = H_p H_p^*$ on trouve immédiatement que $E\{J_p J_q\} - 4\sigma_Y^4$

$$= \frac{4K_4}{n} + \frac{4\sigma_Y^4}{n^2}\left[\frac{1-\cos n(\omega_1+\omega_2)}{1-\cos(\omega_1+\omega_2)} + \frac{1-\cos n(\omega_1-\omega_2)}{1-\cos(\omega_1-\omega_2)}\right]$$

où $K_4 = E\{Y_t^4\} - 3\sigma_Y^4$

Ceci conduit aux conclusions suivantes:

(i) $\mathrm{var}\{J_p\} \sim 4\sigma_Y^4$ ce résultat étant toujours vrai si Y_t n'est pas normale

(ii) $\mathrm{covar}\{J_p\ J_q\}$, pour $n\,|\omega_1 - \omega_2| \gg 1$, est:

a) $O(1/n^2)$ pour Y_t normale (exactement zéro si p, q sont entiers)

b) $O(1/n)$ si $K_4 \neq 0$.

Supposons maintenant que X_t soit ce que j'ai appelé (Bartlett 1946) un processus linéaire

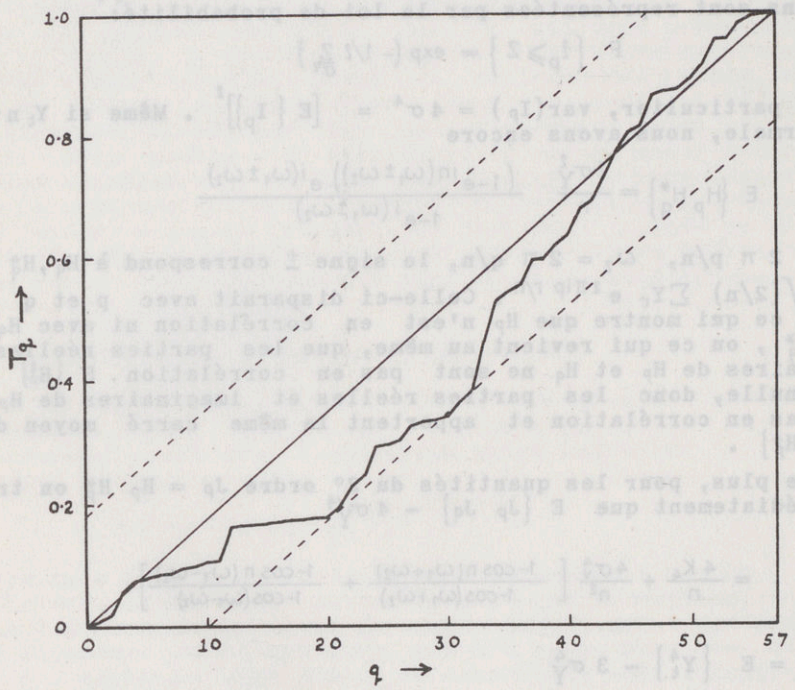
$$X_t = \sum_{u=-\infty}^{\infty} b_u Y_{t-u} \qquad (b_u = 0 \text{ pour } u < 0),$$

Cette représentation et la représentation autorégressive pour Y_t orthogonale (sans corrélation) sont équivalentes, mais l'hypothèse supplémentaire c'est que Y_t est une suite normale ou au moins indépendante. Alors

$$G_p = \sqrt{\frac{2}{n}} \sum_{u=-\infty}^{\infty} \sum_{r=1}^{n} e^{ir\omega} b_u Y_{r-u}, \quad (\omega = 2\pi p/n)$$

$$= \beta^*(\omega) H_p \{1 + O(1/\sqrt{n})\}$$

où $\beta(\omega) = \sum_{u=-\infty}^{\infty} b_u e^{-iu\omega}$



On suppose que $b_u \to 0$, exponentiellement, lorsque u croit, la différence entre G_p et $\beta^*(\omega) H_p$ n'étant dûe qu'à "l'effet de l'extrémité" (end effect). En fait on montre que ceci affecte la distribution de G_p (ou I_p) jusqu'à seulement $O(1/n)$ (cf. Sargan, 1953), mais le résultat plus faible pour le résidu découle plus directement, et il est suffisant pour notre étude.

Le résultat asymptotique

$$G_p \sim \beta^*(\omega) H_p$$

a une importance considérable. Il nous donne tout de suite la relation asymptotique

$$I_p \sim \beta(\omega)\beta^*(\omega) J_p \qquad (\beta(\omega)) \text{ supposé non nul)}$$

Si nous prenons la moyenne de cette relation, nous avons pour $n \to \infty$, le résultat théorique

$$2\pi\sigma_X^2 F(\omega) = 2\sigma_Y^2 \beta(\omega)\beta^*(\omega)$$

mais ici l'on a besoin de la relation sous la forme stochastique Cela nous donne la relation asymptotique entre le périodogramme d'un processus linéaire et le périodogramme d'une suite indépendante avec un spectre uniforme et nous montre que la distribution de I_p est encore asymptotiquement valable (au moins pour Y_t normale) à condition de remplacer $2\sigma^2$ par $\lim_{n\to\infty} E\{I_p\}$ .Ce résultat a été indiqué d'abord par Slutsky (1927) (bien que pas très clairement).

Il est évident d'après les résultats énoncés ci-dessus, que I_p ne saurait fournir une estimation pour $2\pi\sigma_X^2 f(\omega)$ bien que C_s/C_o soit une estimation de ρ_s. On a été amené ainsi à suggérer certaines modifications pour avoir des estimations"consistantes" (cf. Bartlett, 1950; cf. Tukey, 1949), par exemple,

(i) $$I'(\omega) = \frac{1}{2\varepsilon} \int_{\omega-\varepsilon}^{\omega+\varepsilon} I(\omega)\, d\omega, \qquad \text{(P.J. Daniell)}$$

(ii) $$\overline{I}(\omega) = 2 \sum_{s=-n'+1}^{n'-1} \left(1 - \frac{|s|}{n'} C_s \cos(s\omega)\right) \qquad \text{(Bartlett)}$$

où n = n'm. Cette dernière formule présente la propriété utile de permettre le calcul à partir des n' - 1 premières autocorrélations de la série réalisée et ne nécessite pas une évaluation complète de I_p.

Grenander (1951) a ensuite considéré la formule plus générale, avec une fonction pondérale arbitraire $\alpha(\omega)$,

(iii) $$I''(\omega) = \int I(\omega)\, \alpha(\omega)\, d(\omega)$$

qui comprend (i) et (ii) comme cas particuliers. On ne peut faire un bilan final de l'efficacité pratique des différentes méthodes.

(Un étudiant de Manchester, J. Medhi, dans un travail en cours, a entrepris la recherche de la fonction optimum $\alpha(\omega)$ de $I''(\omega)$ dans des conditions déterminées)

3. BANDE ASYMPTOTIQUE DE CONFIANCE DE LA FONCTION SPECTRALE

Grenander a aussi attaqué cette intéressante question: la construction de la bande asymptotique de confiance de la fonc-

tion spectrale entière, et, (en collaboration avec Rosenblatt, 1952,1953), il a proposé une solution dans certaines conditions. J'ai l'impression qu'on peut avoir des solutions plus satisfaisantes si l'on accorde une considération adéquate aux résultats précédents sur les réalisations et je me propose donc d'étudier ce problème

La transformation linéaire de X en $A_p B_p$ pour $p=1,\ldots\frac{1}{2}(n-1)$, et A_o, pour n impair; où $p=1,\ldots\frac{1}{2}(n-2)$, et $A_o\ A_{1/2n}$, pour n pair, est orthogonale (en admettant que le facteur 2 dans $\sqrt{2/n}$ est supprimé pour p=0 ou 1/2 n). Nous remarquons que de plus:

$$\sum_{r=1}^{n} X_r^2 = \sum_p \left(A_p^2 + B_p^2\right)$$

$$\sim \sum_{p=o}^{Y_2 n} I_p \qquad (\tfrac{1}{2}\ (n-1) \text{ au lieu de } \tfrac{1}{2}\ n \text{ si n est impair}$$

maintenant pour une suite normale indépendante X_r, il s'ensuit que la somme cumulative

$$s_q = \sum_{p=o}^{q} I_p$$

représente une promenade au hasard. Si nous imposons la condition que:

$$\sum_{r=1}^{n} X_r^2 \sim \sum_{p=0}^{Y_2 n} I_p = S_{1/2n}$$

soit donné, il ressort de la théorie des distributions laplaciennes que la distribution de S est équivalente à celle de

$$T_q = S_q / S_{1/2n} \qquad \text{(indépendamment de } S_{1/2n})$$

T_q variant de 0 à 1 lorsque q varie de 0 à $\frac{1}{2}$ n. Les T_q représentent une promenade au hasard conditionnelle qui peut être testée contre son espérance mathématique 2q/n selon la théorie de distribution des fonctions de répartition énoncée d'abord par Kolmogoroff (pour son interprétation selon la théorie de marche au hasard, voir par exemple Bartlett, 1949). Ainsi l'écart de T_q à 2q/n n'excède jamais $\pm\lambda/\sqrt{(\frac{1}{2}n)}$ avec la probabilité

$$\sum_{s=-\infty}^{\infty} (-1)^s e^{-2\lambda^2 s^2}$$

Ce test est asymptotiquement valable pour X_r indépendante et normale. Mais là-dessus nous avons édifié un critère qui est un rapport et donc une fonction des autocorrélations réalisées $R_s = C_s/C_o$; il ne sera pas donc sensible à la non normalité (cf. Bartlett, 1946); en particulier la variance asymptotique de T_q

sera indépendante de K_4, le cumulant du quatrième ordre de X_r. En fait on peut vérifier par nos résultats précédents sur les réalisations que

$$\text{var}\left\{T_q\right\} \sim 4q\left(1-2q/n\right)/n^2$$

indépendamment de K_4, et en conformité avec une marche au hasard restreinte à

$$T_{1/2n} = 1$$

Pour adapter ce test au processus linéaire X_r, nous remarquons que:

$$\sum_{p=o}^{q} I_p \Big/ \left[\beta(\omega)\beta^*(\omega)\right] = \sum_{p=o}^{q} J_p\left(1+0\left(1/n\right)\right) \qquad \left(\omega = \frac{2\pi p}{n}\right)$$

(l'ordre de grandeur du résidu reste le même que précédemment puisque $J_p > 0$ et que son espérance mathématique est la même pour des valeurs différentes de ω). Par suite, finalement, comme $\beta(\omega)\ \beta^*(\omega)$ est proportionnel à $f(\omega)$, nous pouvons prendre

$$T_q \sim \sum_{p=o}^{q} \frac{Ip}{f(\omega)} \Big/ \sum_{p=o}^{1/2n} \frac{Ip}{f(\omega)}$$

Cette fonction est immédiatement utilisable pour tester la qualité de l'ajustement d'une fonction spectrale quelconque donnée a priori. Elle est en principe également utilisable pour construire une bande de confiance pour un spectre inconnu; mais puisque cette fonction dépend de $f(\omega)$ d'une façon compliquée au point de vue du calcul arithmétique, elle est au moins commode pour cette construction. Pour éviter la pondération de chaque I_n on peut réciproquement considérer la fonction non pondérée

$$U_q = \sum_{p=o}^{q} I_p \Big| \sum_{p=o}^{1/2n} I_p$$

Malheureusement les propriétés de la répartition de U_q ne sont pas simples, car U_q n'est plus (même pour les processus laplaciens) équivalente à une quantité associée à une promenade au hasard (à moins que son dénominateur soit considéré comme fixe). Sa structure suppose, cependant, que sa distribution est insensible à la non-normalité; sa variance asymptotique, en particulier, est obtenue comme

$$\frac{4\pi^2}{n^2}\left[\left(1-F(\omega_q)\right)^2 \sum_{p=o}^{q} f^2(\omega) + F^2(\omega_q) \sum_{p=q+1}^{1/2n} f^2(\omega)\right]$$

où nous avons écrit

$$\sum_{p=o}^{q} f(\omega) \Big/ \sum_{p=o}^{1/2n} f(\omega) \sim F(\omega_q)$$

Si, ensuite, nous réduisons U_q en divisant son écart de $F(\omega_q)$ par

$$2\pi\sqrt{\left[\frac{1}{2n}\sum_{p=o}^{1/2n} f^2(\omega)\right]} \sim \sqrt{\left[2\pi\int_o^{\pi} f^2(\omega)\, d(\omega)\right]} = \sqrt{v}\ ,$$

nous suggérons que la quantité $\left[U_q - F(\omega_q)\right]/\sqrt{v}$, pourrait (comme approximation temporaire) en attendant de nouvelles recherches sur sa répartition(1) être testée de la même manière que $T_q - 2q/n$

Un tel critère alternatif pour T_q ne dépend toujours que de la fonction spectrale $f(\omega)$, mais il est peut être commode d'introduire une estimation consistante fonction de l'échantillon, et cela ne changera pas le test de U_q lorsque n est grand. Une telle estimation est, par exemple, obtenue par (1)

$$\sum_{p=0}^{1/2n} I^2(\omega) \Big/ \left[\sum_{p=0}^{1/2n} I(\omega)\right]^2$$

à condition de se rappeler que la valeur moyenne de $I^2(\omega)$ est asymptotiquement le double du carré de $E\{I(\omega)\}$, de sorte que la quantité ci-dessus a la valeur asymptotique

$$\frac{n}{2\pi}\int_0^{\pi} 2f^2(\omega)\, d\omega \Big/ \left(\frac{n}{2\pi}\right)^2 = 4v/n$$

La fonction proposée et étudiée par Grenander et Rosenblatt est

$$V_q = \int_0^{\omega_q} I(\omega)\, d\omega \sim \frac{2\pi}{n} \sum_{p=0}^{q} L(\omega) ;$$

mais elle me paraît avoir le désavantage, sa dépendance de K_4 mise à part, d'être une estimation de la "fonction spectrale" $\sigma_X^2 F(\omega_q)$ correspondant à l'autocovariance $\sigma_X^2 \rho_s$; alors qu'elle est généralement moins intéressante que la fonction spectrale définie ici, $F(\omega_q)$, correspondant directement au corrélogramme ρ_s. Théoriquement, T_q semble supérieur à U_q ou à V_q. Une démonstration rigoureuse des propriétés asymptotiques de V_q a été donnée par Grenander et Rosenblatt ; une démonstration analogue doit être bien plus simple pour T_q.

Il peut parfois être plus utile d'avoir des tests qui n'utilisent pas les "sommes cumulatives", mais utilisent toujours les propriétés asymptotiques de I_p. Un désavantage des tests précédents semble être que si I_p s'échelonne sur un spectre uniforme comme dans T_q, alors ω a le même poids dans tout son domaine de variation; bien que (pour employer la terminologie de la physique) "l'énergie" du spectre peut être largement concentrée dans un domaine particulier. D'autre part, si I_p ne s'échelonne pas,

---

(1) La vraie variance asymptotique de U_q peut, pour toute fonction particulière $f(\omega)$ être comparée, s'il en est besoin, avec la variance entraînée par l'approximation (une telle comparaison pour l'exemple numérique du § 5 a montré que l'approximation était dans ce cas tout à fait acceptable). La validité de l'emploi de U_q au lieu de T_q peut, cependant, être beaucoup accrue en modifiant $f(\omega)$ pour la rapprocher de l'uniformité au moyen d'un changement d'échelle en divisant par une densité provisoire $f_0(\omega)$. Ce procédé donne une statistique effectivement équivalente à T_q, mais pour laquelle une bande de confiance peut être immédiatement obtenue.

Cette estimation ne serait pas très juste, si la densité du spectre présentait une pointe importante pour quelque ω car son erreur d'échantillonnage serait alors grande sauf pour n très grand.

comme c'est le cas avec U_q et V_q , la moyenne supérieure de I_p dans les domaines plus importants de ω tendra à réduire la vitesse d'approche vers ces propriétés asymptotiques.

On a les deux alternatives:

i) Tester le plus grand I_p après avoir pondéré à l'uniformité au moyen d'un $f(\omega)$ hypothétique; cependant, ce test là semblera être plus sensible à un élément spectral anormal discret, et naturellement il est employé dans le cas "classique" de $f(\omega)$ déjà uniforme (Fisher, 1929; cf. voir aussi le § 4 de cette note).

ii) Tester l'homogénéité de I_p au moyen d'un χ^2 pour l'homogénéité des variances (Bartlett, 1937; voir aussi Bartlett, 1950, Whittle 1952 a, dans le présent contexte). Si les I_p sont employés dans sommation, chacun d'eux est asymptotiquement équivalent à une estimation de variance avec deux degrés de liberté. Après la pondération, ces estimations doivent être homogènes. Ce test est très souple, car avant de faire le test on peut faire un groupement de I_p pour un domaine de ω peu important, afin de le rendre plus sensible et tout aussi valable pour le domaine plus important de ω . Une conséquence intéressante des propriétés d'échantillonnage de I_p déjà indiquées c'est que ce test doit être insensible à l'absence de normalité dans le présent contexte (I) .

4. SPECTRES MIXTES

Jusqu'à présent nous avons considéré des séries temporelles ayant des spectres continus, - en contraste avec des séries harmoniques classiques avec quelques fréquences discrètes, à part peut-être un élément "d'erreur" au spectre uniforme. Dans de nombreuses applications il est raisonnable de supposer que l'on sait a priori à laquelle de ces deux classes principales appartient la fonction spectrale. Il faut remarquer que tout élément harmonique discret à fréquences connues peut être éliminé par la méthode classique des moindres carrés; une variation annuelle ou saisonnière tomberait dans cette catégorie. La présence de tout spectre continu complique l'erreur d'estimation d'amplitude et de phase, mais pas plus que cela ne se produit dans l'estimation d'une moyenne générale non nulle.

En somme, il reste encore à discuter deux problèmes qui ne se présentent heureusement pas très souvent. Dans le premier nous pouvons avoir, ou bien un spectre discret à "bruit" uniforme ou bien un spectre continu; dans le second nous pouvons avoir les deux simultanément. Pour le premier la difficulté réside en ce que pour une longueur finie de séries, l'intensité du périodogramme près de la fréquence d'un élément discret peut être confondu avec le maximum d'un élément spectral continu et vice-versa. C'est un problème de discrimination où, si nous connaissons la distribution des résidus, le critère convenable est le

---

(1) On suppose remplie la condition que le nombre de degrés de liberté de χ^2 est restreint, disons, à un nombre fixé n_0 en groupant les I_p . (L'emploi du nombre maximum $\frac{1}{2}$ n a été en effet considéré par Wittle, 1952 a; mais, même dans le cas normal, sa discussion m'apparaît avoir besoin de développements supplémentaires).

quotient de vraisemblance p_o/p_1 ; où p_o et p_1 sont les vraisemblances des deux hypothèses respectives. La difficulté réside en ce que ni l'une ni l'autre hypothèse alternative n'est exactement connue, mais renferme plusieurs constantes non déterminées. On peut toutefois tenir compte approximativement de l'effet produit par l'utilisation des estimations de ces constantes. Ainsi si nous voulons comparer les deux hypothèses H_o, H_1 le processus normal consiste à voir dans:

H_o: un élément spectral uniforme continu avec un seul élément harmonique discret surajouté

H_1: une série autorégressive du second ordre, nous avons alors:

$$(\log p_o)_{max} = C - \tfrac{1}{2} n \log \sum\nolimits_o ,$$

où $\sum_o$ désigne la somme des carrés des résidus après avoir ajusté l'élément discret harmonique (on suppose que la variance résiduelle σ^2 est inconnue). De même

$$(\log p_o)_{max} = C - \tfrac{1}{2} n \log \sum\nolimits_1 ,$$

où $\sum_1$ est la somme des carrés des résidus après avoir ajusté la relation autorégressive. Comme $\sum_o$ perd deux degrés de liberté quand on ajuste un élément discret de fréquence connue (l'erreur supplémentaire provenant de l'estimation de fréquence peut être négligée) et, qu'il en est de même pour $\sum_1$, nous pouvons provisoirement admettre l'effet d'estimation en écrivant

$$\log p_o - \log p_1 = -\tfrac{1}{2}(n-2) \log \left(\sum\nolimits_o / \sum\nolimits_1 \right)$$

(on suppose que , à l'origine , le nombre d'observations dans la fonction de vraisemblance est n dans chaque cas; souvent, par commodité, les premières K observations ne sont pas employées dans l'ajustement d'une série autorégressive d'ordre K; alors il est utile de faire un ajustement de plus pour rendre p_o et p_1 aussi comparables que possible).

La manière la plus simple de tester le quotient précédent consiste à employer l'adaptation suggérée par G.A. Barnard de la théorie de Abraham Wald de tests séquentiels. Nous regardons notre échantillon de n observations comme le premier d'une suite hypothétique d'échantillons indépendants et adoptons la règle suivante;

a) si $p_o/p_1 \geqslant (1-\varepsilon)/\varepsilon$ on adoptera H_o

b) si $p_o/p_1 \leqslant \varepsilon/(1-\varepsilon)$ on adoptera H_1

c) si $(1-\varepsilon)/\varepsilon > p_o/p_1 > \varepsilon/(1-\varepsilon)$ on ne décidera rien sans de nouvelles données.

Ainsi le risque d'erreur d'accepter H_o, si H_1 est vrai ou vice-versa, n'est au moins pas supérieur à ε (1)

---

(1) Cette règle obligera à de trop prudentes évaluations des risques si (comme nous l'espérons) la probabilité d'une décision avec un seul échantillon disponible n'est pas petite. Quelques exemples explicites ont été étudiés par J. Medhi.

Le dernier problème que nous avons cité, et où l'on a à la fois le spectre discret et le spectre continu est la question la plus difficile à résoudre. Wittle (1952) suggère comme méthode possible une standardisation approximative du périodogramme d'échantillon I_p par $f(\omega)$ (comme dans notre fonction précédente T_q).

Nous avons vu que ceci réduit asymptotiquement le périodogramme d'une série à spectre continu à un périodogramme d'une série à spectre uniforme, de sorte qu'on peut employer la méthode "classique" de détection des éléments discrets sur des bruits uniformes du fond. La difficulté réside naturellement dans ce fait que nous ne connaissons pas $f(\omega)$; mais Whittle suggère qu'elle peut être estimée par les procédés de régularisation que nous avons déjà étudiés dans ces conférences. Comme cette estimation peut être influencée à son tour par des éléments discrets non encore éliminés, une ou plusieurs itérations de ce procédé peuvent être nécessaires. Même alors, il n'est pas très clair jusqu'à quel point cette itération converge pour des séries finies. Comme je l'ai déjà souligné, de nouvelles recherches sur tous ces problèmes que nous venons de discuter sont encore nécessaires avant qu'on puisse enfin imposer leur valeur pratique.

5. APPLICATIONS NUMÉRIQUES

Nous allons appliquer quelques uns des tests des paragraphes 3 et 4 à des records de chasse au lynx canadien (cf. Moran 1937); ces records montrent un "cycle" intéressant de 10 années. Les chiffres sont de 114 fréquences annuelles, qu'on exprimera pour des raisons de commodité en logarithmes. A partir de ces données, on calcule les valeurs I_p (voir le tableau I dans l'appendice). A cause de la valeur très forte de p = 12, on ne peut déduire immédiatement s'il n'y a pas lieu d'introduire un élément spectral discret. Cependant, si nous comparons les sommes des carrés des résidus dans les deux hypothèses H_o et H_1 du paragraphe 4, nous avons:

$$\sum\nolimits_o = 35{,}236 - 21{,}025 = 14{,}211 \qquad (111 \text{ degrés de liberté})$$

$$\sum\nolimits_1 = 5{,}788 \times 111/109 = 5{,}894^{(1)} \qquad 1/2(111)\log e(14{,}211/5{,}894) = 48{,}9 > \log e^{99} = 4{,}60$$

Aussi, entre H_o et H_1, c'est H_1 qu'il faut choisir. (Moran est arrivé à la même conclusion, sans cependant avoir établi un test définitif, par la seule inspection du corrélogramme). Les coefficients a_1 et a_2 de l'équation auto-régressive

$$X_t + a_1 X_{t-1} + a_2 X_{t-2} = Y_t$$

ou X_t est mesuré à partir de sa moyenne d'échantillon, sont $-1{,}4101$ et $0{,}77_{33}$

De la formule

$$f(\omega) = \left[\sigma_y^2 / \pi \sigma_x^2\right] \beta(\omega)\beta^*(\omega)$$

(1) Le chiffre 5,788 est obtenu à partir des 112 résidus actuels donnés par Moran après avoir ajusté un modèle autorégressif du 2° ordre; il a été ajusté pour compenser les 2 premiers résidus manquants.

où

$$1/\left[\beta(\omega)\,\beta^*(\omega)\right] = (1-a_2)^2 + a_1^2 + 2a_1(1+a_2)\cos\omega + 4a^2\cos^2\omega = b(\omega)$$

nous pouvons construire facilement T_q, dans l'hypothèse H_1.

$$T_q = \sum_{p=1}^{q} I_p\, b(\omega) \Big/ \sum_{p=1}^{1/2\,n} I_p\, p(\omega) \cdot \quad (\omega = 2\pi\, p/n)$$

I_o n'intervient pas, car elle correspond à la moyenne (ici non nulle). La différence absolue la plus grande de T_q par rapport à $q/57$ est 0,184 (pour q= 30) qui dépasse juste le niveau significatif de P= 0,05(1), de $1{,}36/\sqrt{57} = 0{,}180$.

Dans cet exemple l'effet dominant de la fréquence au voisinage de p = 12 est, ainsi qu'on l'a noté au § 3 perdu dans le test T_q, et l'emploi du test U_q non pondéré pourrait paraître plus approprié. Pour ce test alternatif, comme on cherche à tester une f(ω) particulière plutôt qu'à trouver une bande générale de confiance, il est préférable de calculer

$$v = \frac{(2\pi)^2}{2n} \sum_{p=o}^{1/2\,n} f^2(\omega) = \frac{2}{n}\left(\frac{\sigma_y^2}{\sigma_x^2}\right)^2 \sum_{p=o}^{1/2\,n} 1/b^2(\omega)$$

à partir des valeurs <u>théoriques</u> de b (ω), particulièrement en vue d'obtenir la valeur d'échantillon maximum de I_p pour p = 12.

Avec $\sigma_y^2/\sigma_x^2 = 0{,}147783$ (et la sommation comprend la moitié de la contribution de p= 0 et de p=½ n, comme pour f(ω))nous obtenons v = 4,430. Alternativement, on peut employer la formule équivalente

$$v \sim \pi \int_o^{\pi} f^2(\omega)\, d\omega = \sum_{-\infty}^{\infty} \rho_s^2 \cdot$$

qui donne la vérification suffisante, 4,417.

Une différence significative P=0,05 de U_q (<u>disposée dans</u> le tableau I)à F (ω)devra être de l'ordre de $1{,}36 \times \sqrt{4{,}43/\sqrt{57}} = 0{,}38$ et malgré le grand saut de U_q pour p=12, cette valeur n'est pas atteinte. Bien plus, la contribution dominante de I_{12} souligne combien toute théorie asymptotique est inadéquate pour U_q (ou V_q) dans cet exemple(2).

---

(1) Voir figure.

(2) En utilisant alternativement la Statistique modifiée suggérée à la note de la page , avec la fonction f (ω) utilisée dans T_q donnant la fonction spectrale provisoire, nous calculons:

$$\tfrac{1}{4}(114)\sum\left[I_p/f(\omega_p)\right]^2 \Big/ \left[\sum I_p/f(\omega_p)\right]^2 = 1{,}0304$$

de telle sorte que maintenant $\sqrt{v} = 1{,}015$. Comme on peut s'y attendre, cette valeur est voisine de l'unité, et montre qu'une bande de confiance approchée pour la vraie fonction f(ω) peut être obtenue en remplaçant les frontières parallèles a' de la distribution théorique cumulative 2 q/n du graphique T_q, par de nouvelles frontières déduites par des translations de la distribution cumulative observée T_q sur cette échelle provisoire.

Alors qu'on a eu une différence significative avec le test T_q, une comparaison de ces conclusions avec celles tirées des autres tests sur les corrélogrammes montre les limites des tests ci-dessus quand on les emploie conjointement avec une fonction $f(\omega)$ "pas très uniforme". (cf. mes remarques à la fin du § 3). Moran trouve que les coefficients de régression partielle du quatrième ordre et du cinquième ordre dépassent le niveau significatif P=0,05. Un calcul plus compréhensif des 27 premières autocorrélations d'échantillons (qui ont été données par Moran dans cette note) sur la base du test goodness of fit de Quenouille (1947), donne pour χ^2 total la forte valeur de 54,4 avec 25 degrés de liberté (1). Cette approximation par le corrélogramme semble la plus naturelle quand une hypothèse de structure autorégressive est sous examen. Néanmoins, quoique les tests des spectres ne soient pas aussi appropriés dans cet exemple, l'emploi du χ^2 du périodogramme apparaît comme suffisamment effectif.

Les nombres testés sont dans le tableau II; le groupement de quelques degrés de liberté a été décidé a priori par rapport aux contributions à la fonction de distribution théorique $F(\omega)$ et non par rapport aux valeurs I_p.

Tableau II

p	$I_p b(\omega)$	D.F.	p	$I_p b(\omega)$	D.F.
1,2	0.14013	4	12	0.41208	2
3,4	0.40526	4	13	0.01589	2
5,6	0.05908	4	14	0.00179	2
7	0.02310	2	15,16	0.02989	4
8	0.07789	2	17,18,19	0.04600	6
9	0.01371	2	20,28	1.15438	18
10	0.05135	2	29,57	4.87392	57
11	0.01707	2	TOTAL:	7.32154	113

Le test d'homogénéité de ces sommes (pondérées) I_p donne

$$C\,\chi^2 = 37,70 \quad \text{où} \quad C = 1,1246$$

c'est-à-dire un χ^2 corrigé de 33,52 (avec 14 degrés de liberté) significatif au niveau P = 0,01. Ce résultat nous montre à partir du périodogramme avec une évidence plus convaincante combien l'hypothèse autorégressive est inadéquate.

On doit remarquer que, tandis que la valeur pondérée de I_p pour p = 12 n'apparaît plus particulièrement anormale , il indique sur l'échelle originale une hauteur suffisante et un maxi-

(1) Il y a une certaine évidence théorique (Bartlett et Rajalakshmann, 1953) que ce test exagère le niveau significatif véritable pour n borné. Cependant un calcul du prochain terme dans l'expansion asymptotique pour la variance d'échantillon, dans notre hypothèse que celui-là est comparable au terme correspondant du χ^2 basé sur les corrélations directes des résidus (Bartlett et Diananda, 1950) donne un "χ^2 réduit" de 38,5 avec 18 degrés de liberté, encore significatif pour le niveau P = 0,01.

mum assez proche pour rendre indésirable une régularisation excessive du périodogramme (avec seulement 114 observations); aucune régularisation, d'après les méthodes du § 2 n'a été effectuée. Il serait évidemment possible de chercher à trouver des fonctions de densité spectrale f (ω) plus compatibles avec les I_p observées que celle que l'on a testée: mais mon premier but a été d'avancer et de donner des exemples de divers tests, plutôt que de faire un examen approfondi de ces données particulières.

Remerciements - A M. A.M. Walker pour ses critiques au sujet de ces Conférences. A Mrs. A. Linnert pour l'analyse du périodogramme du tableau I, et à Mademoiselle Barbara Appleby qui m'a aidée dans l'analyse du corrélogramme. A MM. Adikhari et Joshi pour leur traduction.

BIBLIOGRAPHIE

Bartlett, M.S. (1937) "Properties of sufficiency and statistical test", Proc. Roy. Soc. Lond. A. 160, 268.

Bartlett (1946)"On the theoretical specification and sampling properties of autocorrelated time-series", J. Roy Statistical Soc. (Suppl.) 7, 211.

(1949)"Some evolutionary stochastic processes", J. Roy. Statist. Soc. (B), 11, 211.

(1950)"Periodogram analysis and continuous spectra" Biometrika, 37, 1.

(1954)An Introduction to Schochastic Processes with special reference to methods and applications (Cambridge, in Press).

Bartlett, M.S. and Diananda, P.H. (1950)"Extensions of Quenouille"s test for autoregressive schemes J. Roy. Statist. Soc. (B) 12, 108.

Bartlett, M.S., and Rajalaskshman, D.S. (1953) "Goodness of fit tests for simultaneous autoregressive series", J. Roy. Statist. Soc. (B), 15.

Davenport, Jr., W.B., Johnson, R. A., and Middleton, D. (1952) "Statistical errors in measurements on random time functions", J. App. Phys. 23, 377.

Fisher, R.A. (1929)"Test of significance in harmonic analysis" , Proc. Roy. Soc. Lond. A 125, 54.

Grenander, U (1950) "Stochastic processes and statistical inference", J. App. Phys. 23, 377. Ark.Mat. 1, 195.

(1951)"On empirical spectral analysis of stochastic processes", Ark. Mat. 1, 503.

Grenander, U., and Rosenblatt, M. (1952) "On spectral analysis of stationary time-series", Proc. Nat. Acad. Sci. 38, 519.

Grenander, U., and Rosenblatt, M. (1953) "Statistical spectral analysis of time series arising from stationary stochastic processes", Ann. Mat.Statist. 24, 537.

Kendall, M.G. (1946) Researches in Oscillatory Time-Series, Cambridge.

Kolmogorov A. (1953)"Sulla determinazione empirica di una legge di distribuzione", Inst. Ital. Attuari Giom. 4, 1

Levy, P. (1948) Processus Stochastiques et Mouvement Brownien (Paris).

Mann, H.B., and Wald, A. (1943)"On the statistical treatment of linear stochastic difference equations", Econometrica, 11, 173.

Moran, P.A.P. (1953)"The Statistical analysis of the Canadian lynx cycle. I. Structure and prediction", Austral J. Zool. 1, 163.

Quenouille, M.H. (1947)"A large-sample test for the goodness of fit of autoregressive schemes" J. Roy Statist.Soc 110, 123.

Rao, U.R. (1951)"The applicability of large sample tests for moving average and autoregressive schemes to series of short lenght-an experimental study.Part.3: The discriminant function approach in the classification of time series", Sankhyä, 11, 257.

Sargan, J.D. (1953)"An approximate treatment of the properties of the correlogram and periodogram" J.Roy.Statist Soc. (B), 15, 140.

Slutsky, E. (1927)"The summation of random causes as the source of cyclic processes", Problems of Economic Conditions, ed. by the conjuncture Institute, Moscow,3 Ne 1. (later reprinted in Econometrica, 5, 105).

Tukey, J.W. (1949)"Measuring noise colour", (unpublished memorandum).

Walker, A.M.(1952)"Some properties of the asymptotic power functions of goodness-of-fit tests for linear autoregressive schemes", J.Roy Statist.Soc.(B), 14,117.

Whittle P. (1951) Hypothesis testing in time series analysis Uppsala).

(1952a)"Tests of fit in time series" Biometrika 39 309.

(1952b)"The simultaneous estimation of a time series harmonic components and covariance structure", Trabajos Estadistica, 3, 43.

(1953)"The analysis of multiple stationary time series, J.Roy.Statist.Soc. (B) 15, 125.

Wold, H. (1938) Analysis of stationary Time-Series Uppsala.

(1949)"A large-sample test of moving averages", J. Roy Statist. Soc. (B), 11, 297.

Yule, G.U. (1927)"On the method of investigating periodicities in disturbed series, with special reference to Wolfer's sun-spotnumbers". Philos.Trans.A,226,267

Tableau I

p,q	I_p	U_q	$I_p b(\omega)$	T_q	$F(\omega)$	$\Delta F(\omega)$
0		0	0	0	0,00981	0,00981
1	0,3818	0,0108	0,05116	0,0070	0,02943	0,01962
2	0,7061	0,0309	0,08897	0,0192	0,04999	0,02056
3	2,2847	0,0957	0,27188	0,0563	0,07175	0,02176
4	1,2825	0,1321	0,13338	0,0745	0,09666	0,02491
5	0,4003	0,1435	0,03659	0,0795	0,12499	0,02833
6	0,2928	0,1518	0,02249	0,0826	0,15872	0,03373
7	0,3756	0,1625	0,02310	0,0858	0,20083	0,04211
8	1,6715	0,2099	0,07789	0,0964	0,25641	0,05558
9	0,4069	0,2215	0,01371	0,0983	0,33327	0,07686
10	2,1132	0,2814	0,05135	0,1053	0,43985	0,10658
11	0,9126	0,3074	0,01707	0,1076	0,57836	0,13851
12	21,0246	0,9041	0,41208	0,1639	0,71050	0,13214
13	0,5778	0,9205	0,01589	0,1661	0,80468	0,09418
14	0,0406	0,9217	0,00179	0,1664	0,86328	0,05860
15	0,1812	0,9268	0,01292	0,1681	0,89960	0,03632
16	0,1557	0,9312	0,01697	0,1704	0,92336	0,02376
17	0,0924	0,9339	0,01515	0,1725	0,93916	0,01580
18	0,0080	0,9341	0,00183	0,1728	0,95047	0,01131
19	0,0927	0,9367	0,02902	0,1767	0,95874	0,00827
20	0,0221	0,9374	0,00906	0,1780	0,96506	0,00632
21	0,2510	0,9445	0,13278	0,1961	0,96996	0,00490
22	0,3611	0,9547	0,24049	0,2290	0,97385	0,00389
23	0,1777	0,9598	0,14589	0,2489	0,97700	0,00315
24	0,3175	0,9686	0,31814	0,2924	0,97959	0,00259
25	0,0339	0,9698	0,04051	0,2979	0,98175	0,00216
26	0,0325	0,9707	0,04596	0,3042	0,98358	0,00183
27	0,1054	0,9737	0,17359	0,3279	0,98516	0,00158
28	0,0252	0,9744	0,04796	0,3345	0,98652	0,00136
29	0,00013	0,97438	0,00022	0,3345	0,987703	0,001187
30	0,02294	0,97503	0,05672	0,3422	0,988749	0,001046
31	0,05902	0,97671	0,16426	0,3647	0,989679	0,000930
32	0,04530	0,97800	0,14111	0,3840	0,990510	0,000831
33	0,16174	0,98259	0,55819	0,4608	0,991261	0,000751
34	0,13346	0,98637	0,50877	0,5298	0,991940	0,000679
35	0,02145	0,98698	0,08978	0,5420	0,992559	0,000619
36	0,01871	0,98751	0,08510	0,5536	0,993129	0,000570
37	0,06197	0,98927	0,30578	0,5954	0,993654	0,000525
38	0,00266	0,98935	0,01435	0,5974	0,994141	0,000487
39	0,02603	0,99009	0,14823	0,6176	0,994595	0,000454
40	0,02213	0,99072	0,13450	0,6360	0,995019	0,000424
41	0,04949	0,99212	0,32031	0,6798	0,995421	0,000402
42	0,01225	0,99247	0,08421	0,6913	0,995799	0,000378
43	0,04148	0,99365	0,29938	0,7322	0,996159	0,000360
44	0,04151	0,99482	0,31418	0,7751	0,996501	0,000342
45	0,02565	0,99555	0,20336	0,8029	0,996827	0,000326
46	0,01052	0,99585	0,08651	0,8147	0,997141	0,000314
47	0,04077	0,99701	0,34866	0,8624	0,997444	0,000303
48	0,00668	0,99720	0,05917	0,8705	0,997736	0,000292
49	0,00369	0,99730	0,03664	0,8751	0,998021	0,000285
50	0,01218	0,99765	0,11581	0,8906	0,998298	0,000277
51	0,01648	0,99812	0,15737	0,9121	0,998570	0,000272
52	0,03090	0,99899	0,30026	0,9531	0,998837	0,000267
53	0,00150	0,99904	0,01480	0,9552	0,999099	0,000262
54	0,01879	0,99957	0,18764	0,9808	0,999358	0,000259
55	0,01051	0,99987	0,10571	0,9952	0,999614	0,000256
56	0,00325	0,99996	0,03338	0,9998	0,999870	0,000256
57	0,00013	1,0000	0,00152	1,0000	1,000000	0,000130
TOTAL:	35,230		7,32154			1,000000

ERRATA
AU MÉMOIRE DE M. S. BARTLETT

PROBLÈMES DE L'ANALYSE SPECTRALE DES SÉRIES TEMPORELLES STATIONNAIRES

Vol. III - Fascicule 3

p.123. On doit lire comme suit la deuxième équation :

$$I_p \sim \beta(\omega)\,\beta^*(\omega)\,J_p \quad (\beta(\omega) \text{ supposé non nul})$$

p.125. Dans la troisième équation, on doit lire :

$O(1/\sqrt{n})$ au lieu de $O(1/n)$

p.126. au lieu de (1), lire (2)

au bas de la page, on doit lire (2) avant le dernier alinéa, commençant par "cette estimation"...

Dans la troisième équation, on doit lire :

$I(\omega)$ au lieu de $L(\omega)$

p.129. paragraphe 5, on doit lire :

$\log_e$ au lieu de $\log e$

p.130. Dans la première équation, on doit lire :

$a_2^2 \cos^2\omega$ au lieu de $a^2 \cos^2\omega$

La deuxième équation doit être lue comme suit :

$$T_q = \sum_{p=1}^{q} I_p\, b(\omega) \Big/ \sum_{p=1}^{\frac{1}{2}n} I_p\, b(\omega)$$

Dans la note (2) au bas de la page, il faut lire :

la Statistique modifiée suggérée à la note (1) de la page 126

p.131. Tableau II. Dans la colonne p lire :

20...28, au lieu de 20,28
et 29...57, " 29,57

p.132. dans la Bibliographie Bartlett 1954, lire :

Stochastic au lieu de Schochastic

p.134. Tableau I. Pour T_q et $q=33$, lire 0,4602 au lieu de 0,4608

Pour $I_p\, b(\omega)$ et $p=50$, lire 0,11381 au lieu de 0,11581