

TRAITEMENT DU SIGNAL ET SES APPLICATIONS

Nice 1^{er} au 5 juin 1971

PAR ECHANTILLONNAGE ALEATOIRE, ESTIMATION DE LA DENSITE
SPECTRALE ENERGETIQUE D'UNE FONCTION ALEATOIRE GAUSSIENNE.

R. H. MACPHIE
OBSERVATOIRE DE MEUDON

RESUME

Nous considérons un échantillonnage aléatoire d'une fonction aléatoire gaussienne à spectre borné. Nous trouvons que, dans certaines conditions, les estimations de la densité spectrale énergétique provenant de M échantillons aléatoires sont presque aussi bonnes que celles issues de N échantillons périodiques avec $N > M$. Donc en utilisant la méthode aléatoire nous obtenons un gain N/M par rapport à la méthode périodique, ce gain est d'autant plus grand que le nombre d'échantillons est important.

SUMMARY

We consider here the random sampling of a gaussian band-limited random function. Under certain general conditions we find that the estimates of the power spectrum obtained from M random samples are almost as good as those obtained from $N > M$ periodic samples. Consequently the random sampling has a gain N/M when compared to periodic sampling, a gain which becomes larger as the number of samples increases.



PAR ECHANTILLONNAGE ALEATOIRE
ESTIMATION DE LA DENSITE SPECTRALE
ENERGETIQUE D'UNE FONCTION
ALEATOIRE GAUSSIENNE

R - H. MACPHIE
OBSERVATOIRE DE MEUDON

I - INTRODUCTION

Il est bien connu que, grâce au théorème de Shannon, une fonction à spectre borné peut être, sans perte d'information, échantillonnée périodiquement avec un pas $\Delta t = \frac{1}{2B}$ où B (Hertz) est la largeur de bande. A une durée d'observation T correspond ainsi un nombre d'échantillons $N = 2BT$.

Lorsque la fonction est aléatoire, il s'agit bien sûr d'obtenir sa fonction d'autocorrélation $R(\tau)$, et par transformée de Fourier, sa densité spectrale énergétique (d.s.e.). Celle-ci, ayant une même bande passante B que celle de la fonction aléatoire, $R(\tau)$, est représentée par la série de valeurs $R(\tau_n)$, où $\tau_n = \frac{n}{2B}$ avec $n = 0, 1, 2, \dots, N-1$. A toute paire d'échantillons de la fonction aléatoire, ayant pour intervalle τ_n , résulte l'estimation de la valeur de $R(\tau_n)$.

Du fait qu'il y a plus de paires disponibles pour un écart τ_n petit que pour un écart grand il en résulte une certaine redondance d'information sur $R(\tau_n)$ d'autant plus grande que τ_n est petit. Or il s'avère ainsi que nous le montrerons, qu'une estimation de $R(\tau_n)$ tirée d'une vingtaine de paires d'échantillons



est en fait presque aussi bonne que celle qui résulterait de plus d'un millier de paires. De telles redondances sont d'autant plus inutiles que le rapport signal sur bruit de l'enregistrement est grand. Dans cet article nous supposons que le bruit est négligeable.

Afin de réduire la redondance nous avons employé un échantillonnage aléatoire, à partir d'une durée d'observation égale à celle de l'échantillonnage périodique. La densité de probabilité d'un instant d'échantillonnage est considérée ici comme étant uniforme. Avec M instants d'échantillonnage distribués au hasard nous aurons $\frac{1}{2} M(M-1)$ paires d'échantillons disponibles. Bien que les écarts de temps entre les paires soient aléatoires, il y a une très grande probabilité que plusieurs écarts soient très voisins de $\tau_n = \frac{n}{2B}$ (c'est à dire entre $\frac{n-\frac{1}{2}}{2B}$ et $\frac{n+\frac{1}{2}}{2B}$). De toutes les paires d'échantillons ayant de tels écarts résulte l'estimation de la valeur de $R(\tau_n)$.

Nous considérons par exemple le cas d'un échantillonnage périodique avec $N = 1000$ échantillons et un échantillonnage aléatoire auquel correspondent $M = 200$ échantillons. Il y a ainsi 1000 valeurs de $R(\tau)$ à évaluer et $\frac{1}{2} (200)(199) \approx 20\ 000$ paires d'échantillons aléatoires disponibles. Donc, en moyenne, le nombre de paires d'où résulte une estimation de $R(\tau_n)$ est vingt, l'avantage étant que la valeur moyenne du pas d'échantillonnage est cinq fois plus grande que la largeur du pas de Shannon.

Une comparaison théorique de l'échantillonnage aléatoire par rapport à l'échantillonnage périodique sera effectuée sous forme statistique. Nous obtiendrons la valeur moyenne et la variance des $R(\tau_n)$ à partir des deux types d'estimation. De tels résultats théoriques seront d'ailleurs confirmés numériquement par calcul machine.

Il sera ainsi montré que les estimations de $R(\tau)$ obtenues par l'échantillonnage aléatoire sont sensiblement



égales à celles provenant de l'échantillonnage périodique, avec pour avantage un nombre d'échantillons nettement inférieur.

2 - LES DEUX METHODES D'ECHANTILLONNAGE

2.1 Echantillonnage périodique (Shannon - Nyquist)

Il est bien connu [1] que lorsqu'une fonction $f(t)$ a une bande limitée de largeur B , l'on peut, après échantillonnage périodique ($\Delta t = \frac{1}{2B}$), la reconstituer sous la forme

$$f(t) = \sum_{k=-\infty}^{\infty} f_k \frac{\sin[(2Bt - k)\pi]}{(2Bt - k)\pi} \quad (1)$$

avec $f_k = f(\frac{k}{2B})$. Si $f(t)$ est une fonction aléatoire il est normal que l'on s'intéresse à sa fonction d'autocorrélation, définie par l'intégrale $R(\tau) = \lim_{T \rightarrow \infty} \frac{1}{2T} \int_{-T}^T f(t) f(t+\tau) dt$, (2)

dont la largeur de bande est aussi bornée; sa densité spectrale énergétique (d.s.e.), obtenue par transformée de Fourier, a pour limites $0 < \omega < 2\pi B$. Donc, l'on peut reconstituer $R(\tau)$ à partir d'une série comparable à celle de $f(t)$:

$$R(\tau) = \sum_{n=-\infty}^{\infty} R_n \frac{\sin[(2B\tau - n)\pi]}{(2B\tau - n)\pi} \quad (3)$$

avec $R_n = R(\tau_n) = R(\frac{n}{2B})$.

En pratique le nombre d'échantillons disponibles est limité, disons N , et l'on ne peut donc en tirer qu'une estimation de $f(t)$ et $R(\tau)$. A partir des relations (1), (2) et (3)

l'estimation de R_n est

$$\tilde{R}_{Nn} = \frac{1}{N^2} \sum_{k=1}^N \sum_{l=1}^N f_k f_l \delta_{n,k-l}, \quad (4)$$

où $\delta_{n,m} = 1$, si $n = m$,

$\delta_{n,m} = 0$, si $n \neq m$. (le symbole de Kronecker). (5)

Grâce à la propriété du symbole de Kronecker, la relation (4) se



réduit à

$$\tilde{R}_{Nn} = \frac{1}{N} \sum_{k=n+1}^N f_k f_{k-n}, \quad n = 0, 1, 2, \dots, N-1. \quad (6)$$

Dans cette dernière relation on voit qu'il y a plus de paires $f_k f_{k-n}$ pour n petit que pour n grand (tendant vers N). Il s'avère que la distribution des paires est triangulaire diminuant de N paires pour $n = 0$ à une paire pour $n = N-1$. Afin de compenser cette espèce d'apodisation, chacun des \tilde{R}_{Nn} est corrigé sous la forme

$$R_{Nn} = \frac{\tilde{R}_{Nn}}{K_{Nn}} \quad (7)$$

$$\text{avec } K_{Nn} = 1 - \frac{n}{N}, \quad n = 0, 1, 2, \dots, N-1. \quad (8)$$

Enfin, la série, limitée à N termes, qui représente l'estimation de $R(\tau)$ est

$$R_N(\tau) = \sum_{n=0}^{N-1} R_{Nn} \frac{\sin[(2Bt - n)\pi]}{(2Bt - n)\pi} \approx R(\tau), \quad \text{pour } 0 < \tau < \frac{N-1}{2B},$$

$$\approx 0, \quad \text{pour } \frac{N-1}{2B} < \tau. \quad (9)$$

De la transformée de Fourier de $R(\tau)$ résulte la densité spectrale énergétique $S(\omega)$, une fonction réelle et paire. Son estimation $S_N(\omega)$ est alors

$$S_N(\omega) = \sum_{n=0}^{N-1} \epsilon_n R_{Nn} \cos\left(n \frac{\omega}{2B}\right), \quad \text{pour } 0 < \omega < 2\pi B, \quad (10)$$

$$= 0, \quad \text{pour } \omega > 2\pi B.$$

$$\text{avec } \epsilon_n = \begin{cases} 1, & \text{si } n = 0, \\ 2, & \text{si } n > 0. \end{cases}$$

Bien sûr cette série $S_N(\omega)$, limitée à N termes, n'est qu'un lissage du vrai spectre $S(\omega)$; les «fréquences» qui correspondent à $\tau_n > \frac{N-1}{2B}$ étant éliminées. Mais si N tend vers l'infini $S_N(\omega)$ converge vers $S(\omega)$ au sens des moindres carrés.



2.2 Echantillonnage aléatoire

Considérons maintenant un échantillonnage aléatoire dont la durée d'observation $T = \frac{N-1}{2B}$ est la même que celle correspondant à l'échantillonnage périodique. Nous supposons la densité de probabilité des instants d'échantillonnage t_m comme étant uniforme :

$$p(t_m) = \frac{1}{T}, \text{ pour } 0 < t_m < T, \\ = 0, \text{ ailleurs,} \quad (11)$$

avec $m = 1, 2, 3, \dots, M$.

Le nombre d'échantillons aléatoires sera, dans les cas d'intérêt, inférieur au nombre N d'échantillons périodiques.

Les M valeurs de t_m sont facilement obtenues au moyen d'un ordinateur, et ordonnées ensuite en fonction des accroissements des t_m sur l'intervalle borné $[0, T]$. Un exemple d'une telle série est présenté sur la Figure 1, le nombre M des instants t_m étant égal à 25. Bien entendu, à chaque t_m correspond un échantillon $f(t_m) = f_m$ de la fonction aléatoire. De plus, l'écart d'un instant d'échantillonnage t_m par rapport à n'importe quel autre instant t_p est $t_{mp} = t_m - t_p$ et il en résulte M^2 écarts accompagnés par M^2 paires d'échantillons (f_m, f_p) . Cependant, puisque $t_{mp} = -t_{pm}$ il y a $\frac{1}{2}M(M-1)$ paires disponibles.

Par analogie avec (4) et (7) une estimation de $R(\frac{n}{2B})$, sous forme aléatoire, est obtenue au moyen de la relation

$$R_{mn}^a = \frac{\tilde{R}_{mn}}{Q_{mn}} = \frac{1}{Q_{mn}} \sum_{m=1}^M \sum_{l=1}^M f_m f_l \pi_n(t_{ml}) \quad (12)$$

$$n = 1, 2, 3, \dots, N-1.$$

Q_{mn} représente le nombre des $\frac{1}{2}M(M-1)$ paires d'échantillons dont les écarts t_{ml} se trouve dans la cellule de largeur $\frac{1}{2B}$ centrée sur $\frac{n}{2B}$. $\pi_n(t_{ml})$, analogue du symbole de Kronecker, est définie par la relation



$$\pi_n(t_{m\bar{t}}) = \begin{cases} 1, & \text{pour } |t_{m\bar{t}} - \frac{n}{2B}| \leq \frac{1}{4B}, \\ 0, & \text{ailleurs.} \end{cases} \quad (13)$$

$$\text{Si } n = 0, \quad R'_{M0} = \frac{1}{M} \sum_{m=1}^M f_m^2. \quad (12a)$$

La relation (12) nous montre que toutes les paires d'échantillons tombantes dans la cellule centrée sur $\frac{n}{2B}$ contribuent à l'estimation de $R(\frac{n}{2B})$ tandis que les autres n'y contribuent pas.

Il est intéressant de considérer une distribution typique des Q_{Mn} . Un tel histogramme est présenté sur la figure (2) avec $M = 90$ et $N = 200$; pour simplifier la présentation nous ne considérons que des cellules qui sont multiples de 5 (5, 10, 15, 20, ...). Il y a $\frac{1}{2}99(89) = 4005$ valeurs de $t_{m\bar{t}}$ distribuées dans l'intervalle $[0, \frac{199}{2B}]$, et 200 cellules pour les recevoir. Leur distribution est aléatoire mais approximativement triangulaire. En effet elle est analogue de K_{Nn} du cas périodique (voir relation (8)). Cependant, R'_{Mn} indiqué dans (12) n'est pas complètement analogue au R_{Nn} dans la relation (7) et cette différence sera précisée dans la section suivante.

3 - ANALYSE PROBABILISTE

Les estimations R_{Nn} et R'_{Mn} introduites dans la section 2 résultent des deux types d'échantillonnage effectués sur la fonction aléatoire $f(t)$. Lorsque le nombre d'échantillons est limité les estimations elles-mêmes sont des variables aléatoires. Afin de comparer l'efficacité de l'une et l'autre des méthodes nous allons évaluer la valeur moyenne et la variance des estimations.

3.1 - Valeur moyenne et variance de R_{Nn} (échantillonnage périodique)

La valeur moyenne ou espérance mathématique de l'estimation R_{Nn} est



$$E \left\{ R_{Nn} \right\} = \bar{R}_{Nn} = \frac{1}{NK_{Nn}} \sum_{k=n+1}^N E \left\{ f_k f_{k-n} \right\} = \frac{N-n}{NK_{Nn}} R\left(\frac{n}{2B}\right) \\ = R\left(\frac{n}{2B}\right), \quad n = 0, 1, 2, \dots, N-1. \quad (14)$$

La valeur moyenne de R_{Nn} est donc égale à la vraie valeur de $R(\tau)$ avec $\tau = \frac{n}{2B}$.

La variance de R_{Nn} est considérée dans l'appendice A où nous montrons qu'elle se réduit à l'expression

$$\text{var} \left\{ R_{Nn} \right\} = \sigma_{Nn}^2 = \frac{f}{(NK_{Nn})^2} \sum_{k=0}^{N-n-1} \epsilon_k [N-n-k] \left[R^2\left(\frac{k}{2B}\right) + R\left(\frac{k+n}{2B}\right) R\left(\frac{k-n}{2B}\right) \right] \quad (15)$$

avec $n = 0, 1, 2, \dots, N-1$.

Bien sûr cette variance est une fonctionnelle des $R(\tau)$; cependant, si $R(\tau)$ tend vers zéro quand τ tend vers l'infini, comme c'est souvent le cas, on trouve que

$$\lim_{N \rightarrow \infty} \sigma_{Nn}^2 = 0, \quad n = 0, 1, 2, \dots$$

Donc, plus grand est le nombre d'échantillons meilleure est l'estimation.

3.2 - Valeur moyenne et variance de R'_{Mn} (échantillonnage aléatoire)

La valeur moyenne ou espérance mathématique de l'estimation R'_{Mn} est

$$E \left\{ R'_{Mn} \right\} = \bar{R}'_{Mn} = \frac{1}{Q_{Mn}} \sum_{m=1}^M \sum_{l=1}^M E \left\{ f_m f_l \pi_n(t_{ml}) \right\} \quad (16)$$

Dans ce cas il y a deux fonctions aléatoires indépendantes :

1. la fonction, $f(t)$;
2. la série d'instants, t_{ml} .

Par conséquent nous écrivons

$$E \left\{ f_m f_l \pi_n(t_{ml}) \right\} = E^{(s)} \left\{ E^{(f)} \left[f_m f_l \right] \pi_n(t_{ml}) \right\} \quad (17)$$



où $E^{(s)}\{\dots\}$ et $E^{(t)}\{\dots\}$ sont les opérateurs d'espérance pour la série d'instants aléatoires t_m et pour la fonction aléatoire $f(t)$ respectivement. Mais $E^{(t)}\{f_m f_1\} = R(t_{m1})$ et la relation (17) devient

$$E\{f_m f_1 \pi_n(t_{m1})\} = E^{(s)}\{R(t_{m1}) \pi_n(t_{m1})\}. \quad (18)$$

A partir de la densité de probabilité $p(t_m)$ indiquée dans (11) nous obtenons

$$E^{(s)}\{R(t_{m1}) \pi_n(t_{m1})\} = \int_0^T dt_m \int_0^T dt_1 R(t_{m1}) \pi_n(t_{m1}) p(t_m, t_1 | t_{m1} \in \pi_n). \quad (19)$$

Il n'est pas difficile à montrer que la densité de probabilité conditionnelle dans l'intégrale est

$$p(t_m, t_1 | t_{m1} \in \pi_n) = \frac{2B}{T-t_n}, \text{ pour } t_{m1} \in \pi_n, \quad (20)$$

$$= 0, \text{ ailleurs.}$$

Il est indiqué sur la Figure 3 que $t_n = \frac{n}{2B}$ et que $t_{m1} \in \pi_n$ (t_{m1} appartient à π_n) si $|t_{m1} - \frac{n}{2B}| < \frac{1}{4B}$. La région π_n est hachurée sur la figure.

Après combinaison des relations (19) et (20)

nous obtenons

$$E^{(s)}\{R(t_{m1}) \pi_n(t_{m1})\} = 2B R\left(\frac{n}{2B}\right) * \pi_0\left(\frac{n}{2B}\right) \quad (21)$$

où le symbole * représente l'opération de convolution, et $\pi_0(t_{m1})$ est défini par (13) avec $n = 0$. Puisque cette expression (21) est indépendante de t_m ou t_1 et puisque dans la relation (16) il y a Q_{mn} espérances de la double série qui sont non nulles, nous pouvons récrire (16) comme $\bar{R}'_{mn} = 2B R\left(\frac{n}{2B}\right) * \pi_0\left(\frac{n}{2B}\right)$ (22)

Pour simplifier la notation nous définissons cette convolution comme $R_c(\tau)$, et par conséquent $\bar{R}'_{mn} = R_c\left(\frac{n}{2B}\right)$. (22a)

La relation (22) nous indique que la valeur moyenne de R'_{mn} n'est pas égale à $R\left(\frac{n}{2B}\right)$; elle représente la valeur moyenne de $R(\tau)$ pour une cellule, centrée sur $\frac{n}{2B}$ dont la largeur est le pas de Shannon.



Il sera néanmoins établi dans l'appendice B qu'à partir des N estimations R'_{Mn} nous pourrons obtenir une autre série d'estimations R_{Mn} :

$$R_{Mn} = \sum_{q=0}^{N-1} A_{Mnq} R'_{Mq}, \quad n = 0, 1, 2, \dots, N-1; \quad (23)$$

les éléments A_{Mnq} sont précisés dans l'appendice. Il en résulte alors que la valeur moyenne est

$$\begin{aligned} E \{ R_{Mn} \} &= \bar{R}_{Mn} = \sum_{q=0}^{N-1} A_{Mnq} \bar{R}'_{Mq} \\ &= R\left(\frac{n}{2B}\right), \quad n = 0, 1, 2, \dots, N-1. \end{aligned} \quad (24)$$

Les relations ci-dessus s'écrivent facilement sous forme matricielle :

$$[R_M] = [A_M] [R'_M]$$

ou $[R_M]$ et $[R'_M]$ sont des matrices colonnes et $[A_M]$ est une matrice carrée.

La variance de R'_{Mn} est considérée dans l'appendice B où l'on montre que dans certaines conditions elle se simplifie à

$$\text{var} \left\{ R'_{Mn} \right\} = \sigma_{Mn}^2 = \frac{2}{(T-t_n)^2} \int_0^{T-t_n} [R_c^2(t) + R_c(t+t_n)R_c(t-t_n)] dt \quad (25)$$

$Q_{Mn} > 20$
 $N > M$

Une comparaison des deux expressions (15) et (25) pour les variances révèle qu'elles sont bien similaires à ceci près que l'intégration dans les cas aléatoires, (25), est remplacée par une sommation dans le cas d'échantillonnage périodique (15).

La variance de l'estimation transformée R_{Mn} peut être facilement présentée sous la forme matricielle :

$$[\sigma_M^2] = [A_M] [\sigma_M'^2] [A_M]^+ \quad (26)$$

avec le symbole $+$ pour indiquer la transposée de la matrice. $[\sigma_M'^2]$ et $[\sigma_M^2]$ sont des matrices colonnes, chacune ayant N éléments.



4 - DISCUSSION.

Au dessous de la Figure 3 se trouve un sommaire des relations obtenues sur la valeur moyenne et la variance des estimations dans les deux cas, échantillonnage périodique et échantillonnage aléatoire. Il y a aussi les transformées matricielles qui s'avèrent nécessaires dans le dernier cas. Il est bien entendu que les conditions $Q_{Mm} > 20$ et $N > M$ sont satisfaites dans la relation pour la variance σ_{Mm}^2 . Ces deux conditions sont très importantes du point de vue pratique, pour les raisons suivantes :

- 1) Lorsque $Q_{Mm} > 20$ le nombre de paires aléatoires qui tombent dans la $n^{\text{ième}}$ cellule, est plus grand que 20, l'estimation ainsi obtenue s'avère presque toujours acceptable.
- 2) Lorsque $N > M$ la valeur moyenne du pas d'échantillonnage est augmentée de $\frac{1}{2B}$ dans le cas périodique à $\frac{N-1}{M-2B}$ dans le cas aléatoire. Néanmoins on obtient le même nombre d'estimations N dans les deux cas. Peut-être le plus remarquable est-il que la variance σ_{Mm}^2 est indépendante de M, le nombre d'échantillons aléatoires. C'est ainsi qu'à condition que l'on ait $Q_{Mm} > 20$ le rapport N/M peut tendre vers l'infini sans qu'il y ait modification de σ_{Mm}^2 .

Considérons, par exemple, le cas où $M = 10^3$ et $N = 10^4$. Alors que la valeur moyenne du pas d'échantillonnage aléatoire est dix fois plus grande que celle qu'elle aurait été pour un échantillonnage périodique, il y a, malgré tout, $\frac{1}{2}M(M-1) \approx 0,5$ million paires d'échantillons qui se trouvent réparties parmi les $N = 10^4$ cellules. Il en résulte donc, en moyenne une cinquantaine de paires d'échantillons aboutissant à une cellule donnée ($\bar{Q} = 50$). Ceci est montré par le petit triangle de la Figure 4, M et N étant les axes de coordonnées. Les facteurs de gain N/M ainsi que les moyennes \bar{Q} aboutissent alors à des lignes droites de pentes différentes. Quant au petit triangle, il se trouve à droite de l'aire hachurée et au-dessus de l'aire ombrée. Alors que la condition $\bar{Q} > 20$ n'est pas satisfaite dans la première région (sous-



échantillonnage), dans la seconde région le pas moyen d'échantillonnage est inférieur à celui de Shannon-Nyquist (sur-échantillonnage).

La Figure (4) nous montre à quel point l'échantillonnage aléatoire est d'autant plus efficace que M et N sont plus grands. Par exemple lorsque $\bar{Q} = 20$ le facteur de gain N/M est, pour M = 100, égal à 2.47. Il est, par contre, égal à 24.7 lorsque M = 1000, (\bar{Q} étant toujours égal à 20).

Sur la Figure (5) le facteur de gain N/M, est indiqué en fonction du nombre d'échantillons aléatoires avec pour paramètre \bar{Q} . A $\bar{Q} = 20$ et $\bar{Q} = 50$ correspondent deux lignes droites, le dernier cas étant plus pessimiste. Cependant, même pour $\bar{Q} = 50$ le gain s'avère tout aussi grande quand le nombre M est élevé. Par exemple nous avons un gain N/M = 100 lorsque $\bar{Q} = 50$ et M = 100. Le pas moyen de l'échantillonnage aléatoire est alors cent fois plus grand que le pas de Shannon.

5 - CONFIRMATION PAR ORDINATEUR.

Afin de confirmer les résultats théoriques nous avons utilisé un ordinateur simulant les deux méthodes d'échantillonnage et les répétant cinquante fois pour obtenir les valeurs moyennes dans chaque cas.

La fonction particulière considérée était gaussienne et possédait deux composantes :

- 1) un <<signal>> ayant une largeur de bande étroite,
- 2) un bruit blanc dont la largeur de bande était beaucoup plus grande.

La fonction d'autocorrélation (normalisée) est alors

$$R(\tau) = \frac{9}{49} \left\{ \frac{40}{9} \frac{\sin\left(\frac{\tau}{20}\right)}{\tau} \cos(\sqrt{2}\tau) + \frac{\sin 2\tau}{2\tau} \right\}, \quad (27)$$

avec $B = \frac{1}{2}$ (simplification non-restrictive des calculs). Le nombre d'échantillons périodiques était N = 200 et M = 90 était le nombre



d'échantillons aléatoires. Donc, le gain est $N/M = 2.22$.

Malheureusement le temps de calcul pour obtenir les résultats était long et nous n'avons obtenu que les premières onze valeurs moyennes de \bar{R}'_n et \bar{P}_n tirées de cinquante répétitions de l'échantillonnage; \bar{R}'_n est la valeur dans le cas aléatoire et \bar{P}_n indique la valeur provenant de l'échantillonnage périodique. Cependant les résultats montrés sur la Figure (6) sont bons. On y trouve trois symboles, représentant \bar{R}'_n , \bar{P}_n et $R(n)$, le dernier, obtenu à partir de la relation (27), est la valeur théorique. Nous voyons que les \bar{R}'_n ont des amplitudes inférieures aux deux autres. Ces phénomènes s'expliquent par des transformations matricielles. Si \bar{R}'_n est transformée en \bar{R}_n comme indiqué par la relation (23) les trois valeurs seront toutes presque identiques.

6 - CONCLUSION

Il y a une certaine redondance à échantillonner périodiquement une fonction aléatoire à spectre borné si l'on utilise ces échantillons pour estimer la fonction d'autocorrélation.

Afin de réduire cette redondance nous avons considéré un échantillonnage aléatoire. La densité de probabilité des instants d'échantillonnage était supposée uniforme et le nombre M d'échantillons aléatoires était inférieur au nombre N d'échantillons périodiques; c'est à dire que la valeur moyenne d'un pas d'échantillonnage aléatoire était plus grande que le pas de Shannon.

Néanmoins, nous avons montré que les estimations de la fonction d'autocorrélation obtenue à partir des deux méthodes d'échantillonnage sont pratiquement identiques, si l'on respecte certaines conditions.

Par conséquent si $M < N$ la méthode aléatoire conduit à un gain N/M par rapport à la méthode périodique. De plus, nous avons trouvé que le gain devient d'autant plus grand que le nombre d'échantillons est important.



BIBLIOGRAPHIE

- [1] C.E. Shannon - "Communication in the presence of noise", Proc. I.R.E., Vol. 37, pp 10-21, January 1949.
- [2] I.A. Reed - "On a moment theorem for complex Gaussian processes", I R E Trans. on Information Theory, vol. I T - 8, pp 194-195, April 1962.

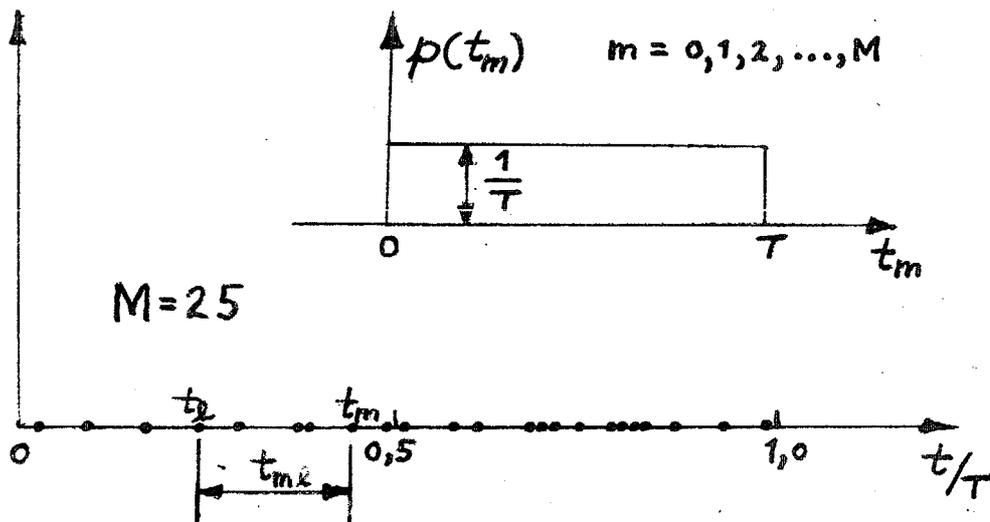


Fig. 1 - Distribution typique d'instants d'échantillonnage aléatoire dont la densité de probabilité est uniforme.

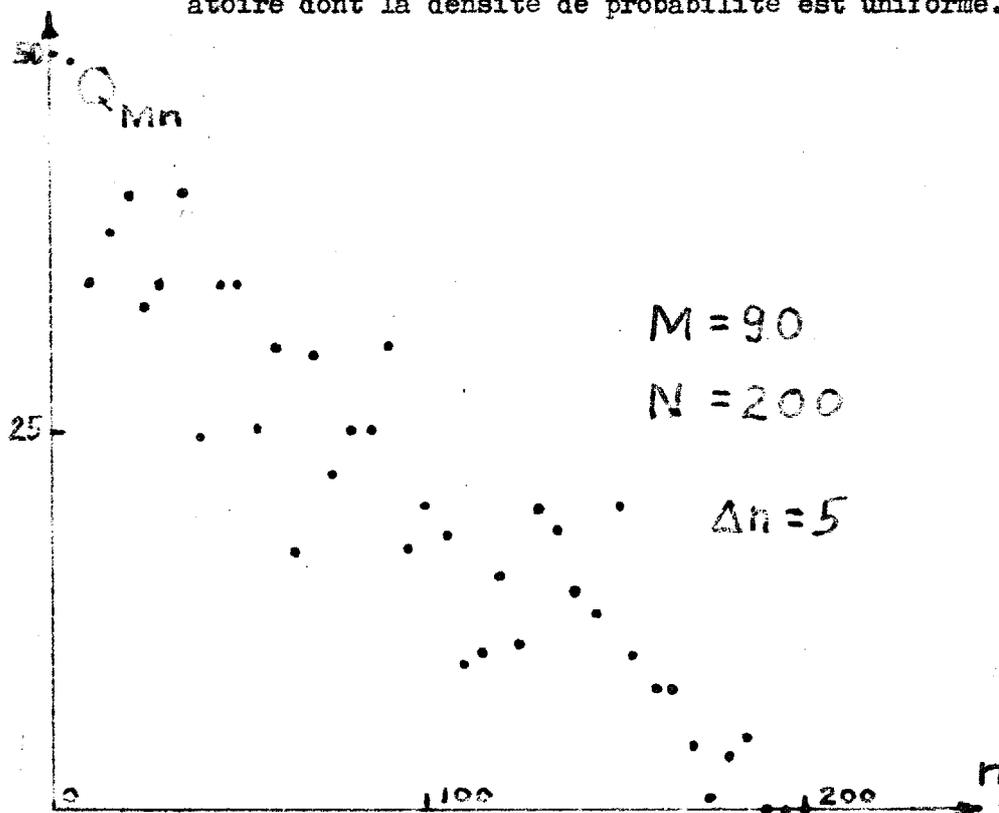


Fig. 2 - Distribution typique des écarts aléatoires t_{m1} entre paires d'instants d'échantillonnage.

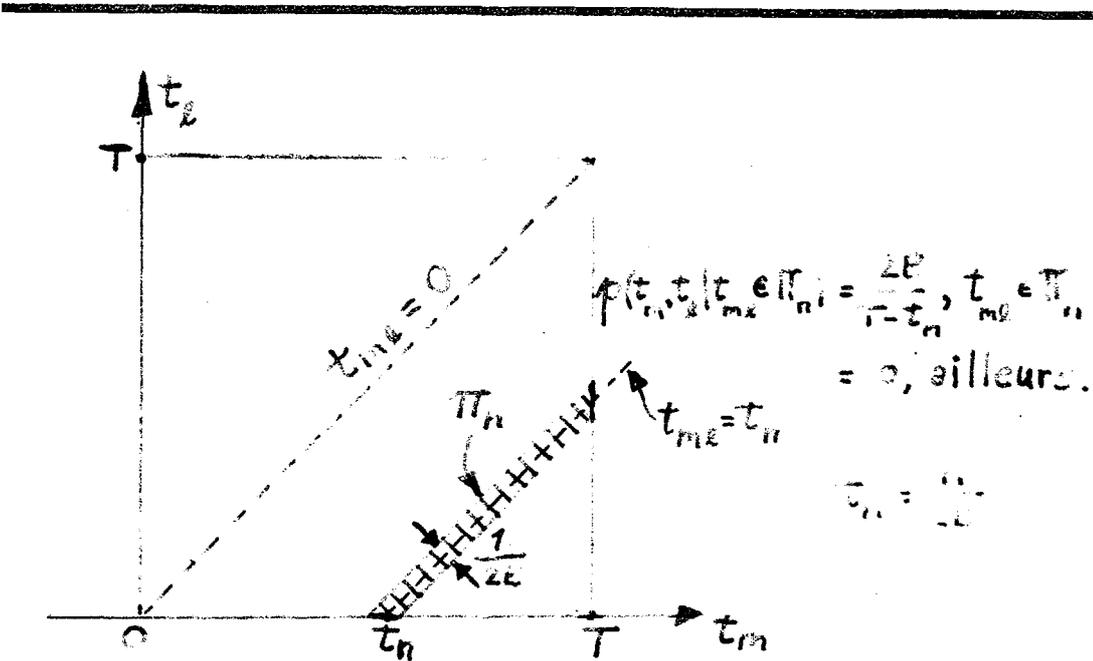


Fig. 3 - Distribution de la densité de probabilité conditionnelle $p(t_m, t_l | t_{ml} \in \pi_n)$ en fonction de t_n .

LES VALEURS MOYENNES

$\bar{R}_{Nn} = R(\frac{1}{2E})$ (périodique)

$R_{Mn}^i = R(\frac{1}{2E}) = 2ER(\frac{1}{2E}) * \pi_n(\frac{1}{2E})$ (aléatoire)

LES VARIANCES

$\sigma_{Nn}^2 = \frac{1}{(N-n)} \sum_{k=0}^{n-n-1} \epsilon_{k, N-n-k} [R(\frac{k}{2E}) + R(\frac{k+n}{2E}) - R(\frac{k-n}{2E})]$ (péri.)

$\sigma_{Mn}^i = \frac{2}{(T-t_n)} \int_0^{T-t_n} [T-t_n-t] [R_c^2(t) + R_c(t+t_n)R_c(t-t_n)] dt$ (al.)

TRANSFORMATIONS MATRICIELLES

$\bar{R}_M = A_M \bar{R}'_M$ $\sigma_M^2 = A_M \sigma_M'^2 A_M^+$

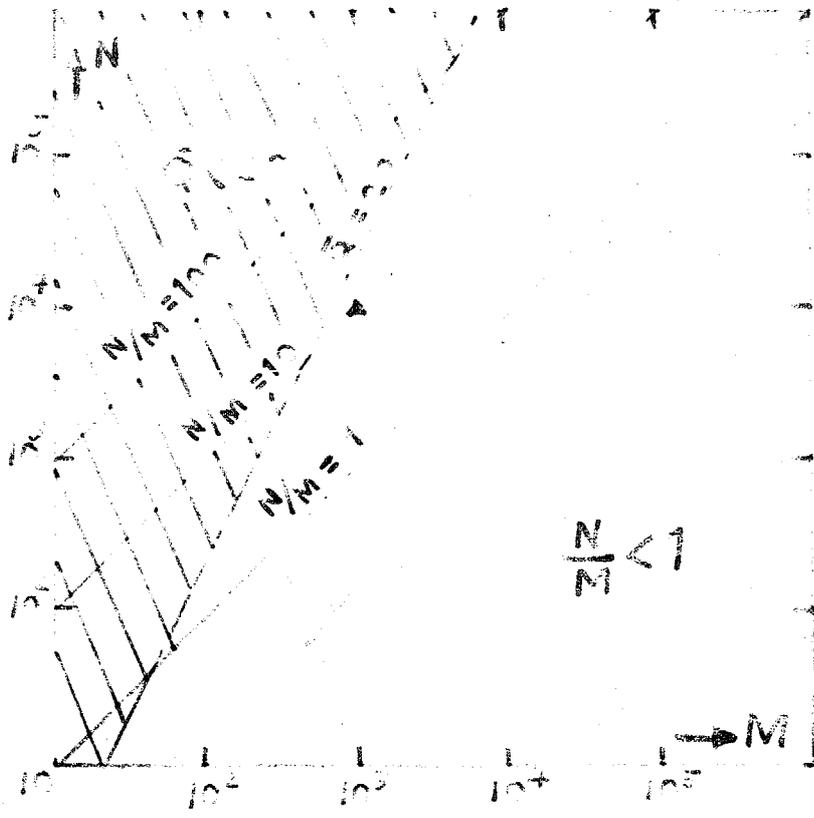


Fig. 4 - Relation entre N et M, avec comme paramètres N/M et \bar{Q} .

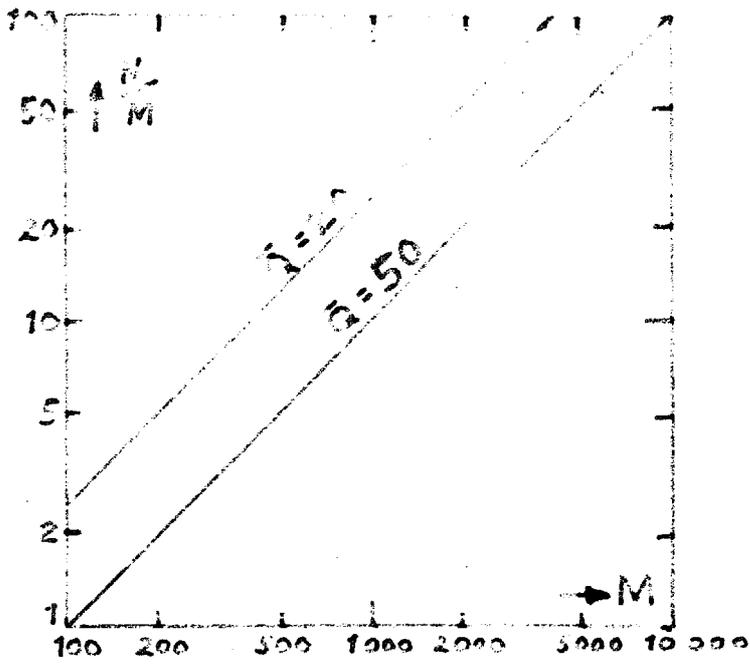


Fig. 5 - Facteur de gain N/M en fonction de M, avec \bar{Q} comme paramètre

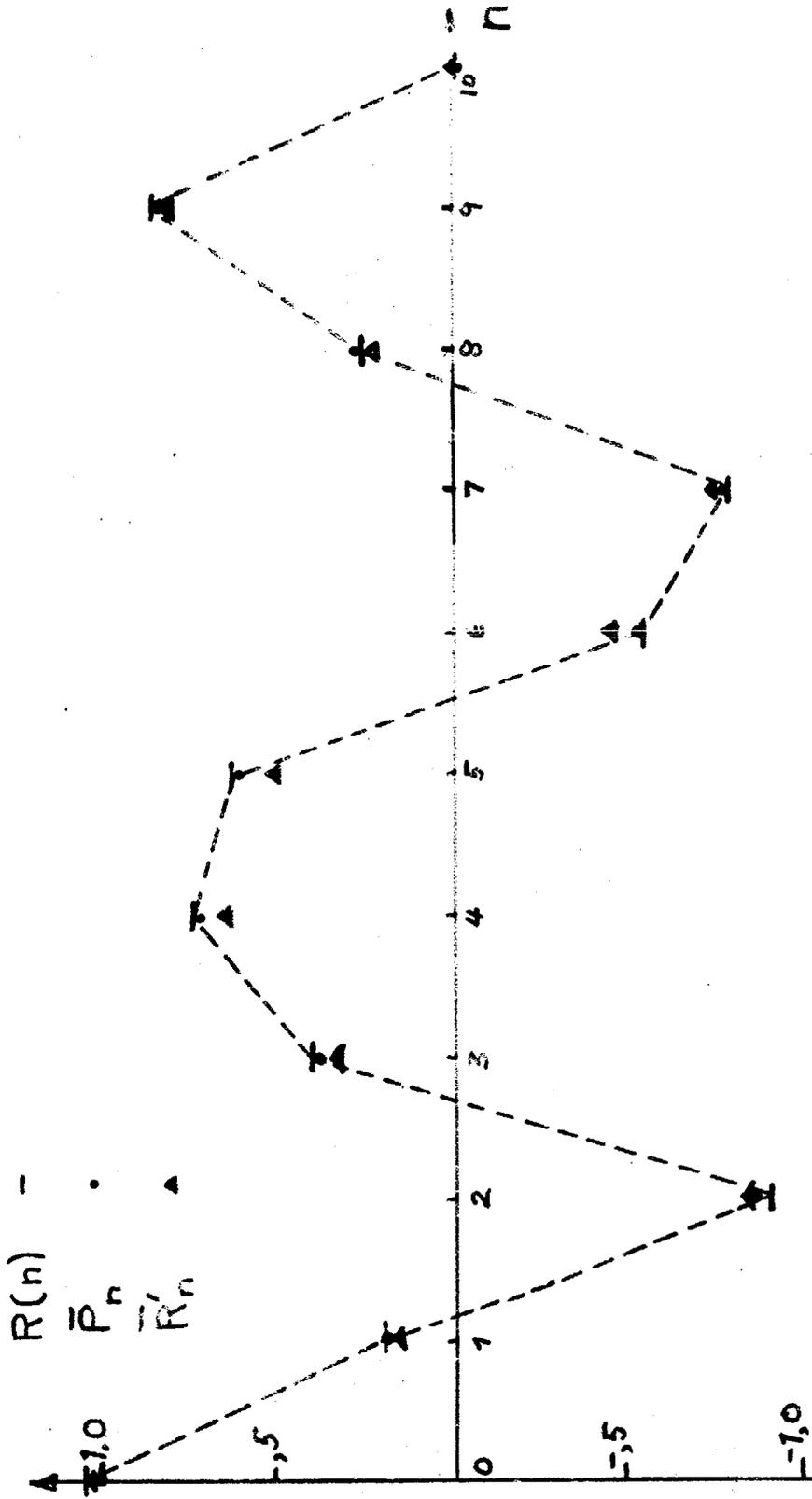


Fig. 6 - Confirmation par ordinateur : comparaison de $R(n)$, P_n et R'_n ($n=0,1,2,\dots,10$);
 $N = 200, M = 90$).



APPENDICE A

Variance de l'estimation R_{Nn} (échantillonnage périodique)

La variance de R_{Nn} est définie par

$$\text{var} \{R_{Nn}\} = \sigma_{Nn}^2 = E \{R_{Nn}^2\} - E^2 \{R_{Nn}\} = E \{R_{Nn}^2\} - \bar{R}_{Nn}^2$$

Si nous utilisons (6) et (7) nous pouvons écrire

$$\begin{aligned} E \{R_{Nn}^2\} &= \frac{1}{(KN)_n^2} E \left\{ \sum_{k=n+1}^N \sum_{l=n+1}^N f_k f_{k-n} f_l f_{l-n} \right\} \\ &= \frac{1}{(KN)_n^2} \sum_{k=n+1}^N \sum_{l=n+1}^N E \{f_k f_{k-n} f_l f_{l-n}\}. \end{aligned}$$

Mais lorsque la fonction $f(t)$ est Gaussienne l'espérance au-dessus peut s'écrire [2],

$$\begin{aligned} E \{f_k f_{k-n} f_l f_{l-n}\} &= E \{f_k f_{k-n}\} E \{f_l f_{l-n}\} + \\ &E \{f_k f_l\} E \{f_{k-n} f_{l-n}\} + E \{f_k f_{l-n}\} E \{f_{k-n} f_l\} \\ &= R^2 \left(\frac{n}{2B}\right) + R^2 \left(\frac{k-l}{2B}\right) + R \left(\frac{k-l+n}{2B}\right) R \left(\frac{k-l-n}{2B}\right). \end{aligned}$$

Donc

$$\begin{aligned} E \{R_{Nn}^2\} &= \frac{1}{(KN)_n^2} \left\{ (N-n)^2 R^2 \left(\frac{n}{2B}\right) + \sum_{k=n+1}^N \sum_{l=n+1}^N \right. \\ &\quad \left. \left[R^2 \left(\frac{k-l}{2B}\right) + R \left(\frac{k-l+n}{2B}\right) R \left(\frac{k-l-n}{2B}\right) \right] \right\} \end{aligned}$$

Si nous prenons $k-l=s$ nous trouvons que la double sommation devient

$$\sum_{k=n+1}^N \sum_{s=k-n-1}^{k-N} \left[R^2 \left(\frac{s}{2B}\right) + R \left(\frac{s+n}{2B}\right) R \left(\frac{s-n}{2B}\right) \right] = 2 \sum_{s=0}^{N-n-1} \epsilon_s [N-n-s] \left[R^2 \left(\frac{s}{2B}\right) + R \left(\frac{s+n}{2B}\right) R \left(\frac{s-n}{2B}\right) \right]$$



Enfin, la variance de R_{Nn} est

$$\begin{aligned} \sigma_{Nn}^2 &= \frac{1}{(K_n N)^2} \left\{ (N-n)^2 R^2\left(\frac{n}{2B}\right) + \sum_{k=0}^{N-n-1} \epsilon_s [N-n-s] \left[R^2\left(\frac{s}{2B}\right) + R\left(\frac{s+n}{2B}\right) R\left(\frac{s-n}{2B}\right) \right] \right\} - R^2\left(\frac{n}{2B}\right) \\ &= \frac{1}{(K_n N)^2} \sum_{s=0}^{N-n-1} \epsilon_s [(N-n)-s] \left[R^2\left(\frac{s}{2B}\right) + R\left(\frac{s+n}{2B}\right) R\left(\frac{s-n}{2B}\right) \right] \end{aligned}$$

APPENDICE B

Variance de l'estimation R'_{Mn} (échantillonnage aléatoire)

Dans ce cas nous voulons évaluer

$$E \left\{ R_{Mn}'^2 \right\} = \frac{1}{Q_{Mn}^2} \sum_{\substack{k=1 \\ k \neq l}}^M \sum_{\substack{l=1 \\ l \neq p}}^M \sum_{\substack{m=1 \\ m \neq r}}^M \sum_{p=1}^M E \left\{ f_k f_l f_m f_p \pi_n(t_{kl}) \pi_n(t_{mp}) \right\}$$

$n = 1, 2, \dots, N-1$.

Dans cette sommation il y a six cas que nous devons préciser :

- (1) le cas où $k \neq l \neq p \neq m$,
- (2) le cas où $k = m$ et $k \neq l \neq p$,
- (3) " " $k = p$ " $k \neq l \neq m$,
- (4) " " $l = m$ " $k \neq l \neq p$,
- (5) " " $l = p$ " $k \neq l \neq m$,
- (6) " " $k = m$ " $l = p$.

Cependant si M est grand le premier cas (tous les indices indépendants : $k \neq l \neq m \neq p$) domine les autres et nous approximons l'espérance par

$$E \left\{ R_{Mn}'^2 \right\} = \frac{1}{Q_{Mn}^2} \left[S_{Mn} E \left\{ f_k f_l f_m f_p \pi_n(t_{kl}) \pi_n(t_{mp}) \right\} \right]_{k \neq l \neq m \neq p}$$

où S_{Mn} est le nombre de termes dans la sommation quadruple dont $k \neq l \neq m \neq p$. Lorsque M est grand $S_{Mn} \approx Q_{Mn}^2$. Nous avons supprimé les signes de consommation parce que nous trouverons que l'espérance au-dessus est indépendante de k, l, m et p .

Considérons cette espérance et n'oublions pas



qu'il y a deux fonctions aléatoires et indépendantes. (la fonction $f(t)$ et la série t_k). Donc nous avons

$$E \left\{ f_k f_l f_m f_p \pi_n(t_{kl}) \pi_n(t_{mp}) \right\} \\ = E^{(s)} \left\{ E^{(t)} \left\{ f_k f_l f_m f_p \right\} \pi_n(t_{kl}) \pi_n(t_{mp}) \right\}$$

mais, $f(t)$ étant Gaussienne, nous trouvons que

$$E^{(t)} \left\{ f_k f_l f_m f_p \right\} = R(t_{kl}) R(t_{mp}) + R(t_{km}) R(t_{lp}) + R(t_{kp}) R(t_{lm}).$$

Donc

$$E \left\{ f_k f_l f_m f_p \pi_n(t_{kl}) \pi_n(t_{mp}) \right\} = [2B] \\ E^{(s)} \left\{ [R(t_{kl}) R(t_{mp}) + R(t_{km}) R(t_{lp}) + R(t_{kp}) R(t_{lm})] \right. \\ \left. \pi_n(t_{kl}) \pi_n(t_{mp}) \right\}.$$

En considérant le premier terme nous avons

$$E^{(s)} \left\{ R(t_{kl}) R(t_{mp}) \pi_n(t_{kl}) \pi_n(t_{mp}) \right\} = \int dt_k \int dt_l \int dt_m \int dt_p R(t_{kl}) R(t_{mp}) \\ \pi_n(t_{kl}) \pi_n(t_{mp}) p(t_k, t_l, t_m, t_p | \left\{ \begin{matrix} t_{kl} \\ t_{mp} \end{matrix} \right\} \epsilon \pi_n)$$

mais puisque tous les t sont indépendants nous trouvons, grâce à (20), que

$$E^{(s)} \left\{ R(t_{kl}) R(t_{mp}) \pi_n(t_{kl}) \pi_n(t_{mp}) \right\} = \left[\int dt_k \int dt_l R(t_{kl}) p(t_k, t_l | t_{kl} \epsilon \pi_n) \right]^2 \\ = \left[2B \int_{t_n - \frac{1}{4B}}^{t_n + \frac{1}{4B}} R(t) \frac{T-t}{T-t_n} dt \right]^2$$

$$\text{où } t_n = \frac{n}{2B} \cdot \frac{T-t}{T-t}$$

Puisque $\frac{T-t}{T-t}$ reste presque égale à 1 à travers l'intervalle d'intégration nous pouvons écrire, et c'est une bonne approximation,

$$E^{(s)} \left\{ R(t_{kl}) R(t_{mp}) \pi_n(t_{kl}) \pi_n(t_{mp}) \right\} = \left[2B \int_{t_n - \frac{1}{4B}}^{t_n + \frac{1}{4B}} R(t) dt \right]^2 = [2BR(t_n) * \pi_0(t_n)]^2$$

où $\pi_n(t_n)$ est défini par (12). Pour simplifier la notation nous



précisons

$$R_c(t) = 2B R(t) * \pi_0(t)$$

et alors nous voyons que le premier terme est $R_c^2(t_n)$.

Une analyse un peu compliqué du second terme nous amène à l'expression

$$E^{(s)} \{ R(t_{km}) R(t_{lp}) \pi_n(t_{kl}) \pi_n(t_{mp}) \} = \frac{2}{(T-t_n)^2} \int_0^{T-t_n} (T-t_n-t) R_c^2(t) dt.$$

Le troisième terme se réduit, après quelques transformées de variable à

$$E^{(s)} \{ R(t_{kp}) R(t_{lm}) \pi_n(t_{kl}) \pi_n(t_{mp}) \} = \frac{2}{(T-t_n)^2} \int_0^{T-t_n} (T-t_n-t) R_c(t-t_n) R_c(t+t_n) dt$$

Puis en utilisant les relations et la définition de la variance nous obtenons enfin

$$\text{var} \{ R_{Mn}^i \} = \sigma_{Mn}^2 = \frac{2}{(T-t_n)^2} \int_0^{T-t_n} (T-t_n-t) [R_c^2(t) + R_c(t-t_n) R_c(t+t_n)] dt$$

sous la condition «M est grand» En pratique nous nous intéressons aux cas où $N > M$ et $Q_{Mn} > 20$. Ces deux conditions exigent automatiquement la condition que $M \gg 1$.

APPENDICE C.

Transformations matricielles

Nous avons déjà établi que l'estimation R_{Mn}^i a une valeur moyenne égale à $\bar{R}_{Mn}^i = R_c\left(\frac{n}{2B}\right) = 2B \pi_0\left(\frac{n}{2B}\right) * R\left(\frac{n}{2B}\right)$.

Le spectre de $R_c(\tau)$, obtenu par transformée de Fourier (T.F) s'écrit

$$S_c(\omega) = F\{R_c(\tau)\} = \frac{\sin\left(\frac{\omega}{4B}\right)}{\frac{\omega}{4B}} S(\omega)$$

grâce au théorème de convolution.

Donc,

$$S(\omega) = W(\omega) S_c(\omega) = W(\omega) F\{R_c(\tau)\}$$



où $F\{\dots\dots\dots\}$ indique la TF et $W(\omega) = \left[\frac{\sin(\frac{\omega}{4B})}{\frac{\omega}{4B}} \right]^{-1}$.

La fonction d'autocorrélation désirée $R(\tau)$ est facilement obtenue par l'inverse de la T.F.

$$R(\tau) = F^{-1} \{S(\omega)\} = F^{-1} \{W(\omega) F[R_c(\tau)]\}$$

Lorsque nous n'avons que N valeurs de $R_{\frac{m}{M}}$ dont le moyen est $R_c(\frac{n}{2B})$ il faut remplacer les T.F. par transformations matricielles. Donc la T.F. au-dessus est analogue à

$$[R_M] = \frac{[F][W][F][R'_M]}{2N-1}$$

où $[R_M]$ et $[R'_M]$ sont matrices colonnes et $[F]$ et $[W]$ sont matrices carrées :

$$F_{mn} = e_n \cos(mn\Delta\omega)$$

$$W_{mn} = W(m\Delta\omega B) \delta_{mn}$$

avec $\Delta\omega = \frac{2\pi}{2N-1}$ et $m, n = 0, 1, 2, \dots, N-1$.

Comme simplification nous définissons la matrice $[A_M] = \frac{[F][W][F]}{2N-1}$.

Donc $[R_M] = [A_M][R'_M]$.

Il est facile à montrer que $[\sigma_M^2]$ la variance de $[R_M]$, est obtenue à partir de $[A_M]$ et de $[\sigma_M'^2]$, la variance de $[R'_M]$.

$$\begin{aligned} [\sigma_M^2] &= E \{ [R_M][R_M]^+ \} - [R_M][R_M]^+ \\ &= E \{ [A_M][R'_M][R'_M]^+[A_M]^+ \} - [A_M][R'_M][R'_M]^+[A_M]^+ \\ &= [A_M] (E \{ [R'_M][R'_M]^+ \} - [R'_M][R'_M]^+) [A_M]^+ \\ &= [A_M][\sigma_M'^2][A_M]^+ \end{aligned}$$

Le symbole $+$ indique la transposée de la matrice.