ONZIEME COLLOQUE GRETSI - NICE DU 1er AU 5 JUIN 1987

E V A L U A T I O N   D E   T E S T S   D ' H Y P O T H E S E

Hocine CHERIFI

CEPHAG, INPG/IEG, UA 346, BP 46 - 38402 ST-MARTIN-D'HERES Cédex (FRANCE)

Dans cet article on étudie quelques tests non paramètriques utiles en traitemement du signal.On traite les problémes classiques de test : échantillon aléatoire, homogénéité, adéquation à une loi. Aprés une présentation succinte des tests utilisés,on commente les résultats d'une étude expérimentale de puissance de ces procédures pour diverses alternatives .Le but recherché est de cerner les limites respectives de ces tests dans des situations fréquemment rencontrées lors d'analyse de données en acoustique sous-marine.

Statistical tests useful in data analysis are discussed.Following a brief tutorial discussion of hyothese testing,selected test for randomness,normality and homogeneity are presented. Using experimental power evaluation we give results of their comparative usefulness in different situations that can arise in data analysis. The idea is to give a practical guide to hypothese testing applied to sonar data.

1) INTRODUCTION

L'étude des propriétés statistiques des signaux est fondamentale dans le traitement du signal. La plupart des systémes de traitement sont en effet conçus pour une modélisation probabilistique donnée. Les tests statistiques d'hypothése permettent de déterminer certaines de ces propriétés.On s'intéresse ici aux trois propriétés de base les plus fréquemment utilisées.On étudie ainsi quelques tests du caractére aléatoire d'un échantillon,cette hypothése étant à la base de tout traitement statistique,puis on aborde le probléme d'adéquation d'un échantillon à une loi de probabilité donnée,en l'occurence la loi gaussienne car elle occupe une place prépondérante dans la majeure partie des problémes de détection et d'estimation. En tout dernier lieu, on étudie des tests qui permettent de décider de l'homogénéite de deux échantillons cette dernière propriété étant la garante de la reproductibilité des résultats lorsqu'on effectue une analyse de données.

Ce travail est plus particulièrement axé sur l'étude de signaux en acoustique sous-marine .Il a été motivé par une analyse de bruit acoustique ambiant [3].

Aprés avoir rappelé quelques données élémentaires sur les tests d'hypothése, on présente de façon succinte les procédures utilisées pour chacune des trois propriétés sous investigation.On donne ensuite des résultats sur leur facultés à mettre en évidence des hypothéses alternatives que l'on peut rencontrer fréquemment en analyse de données acoustiques.

2) TEST STATISTIQUE D'HYPOTHESE

Ces tests ont pour but de déterminer, au vu d'un échantillon, si une certaine hypothése H0 (Hypothése nulle) concernant cet échantillon est vérifiée. La méthodologie des tests d'hypothése demande de construire une statistique,c'est-à-dire une fonction des données que l'on comparera à un seuil (valeur critique).La distribution du test est connue au moins de façon approximative lorsque l'hypothése nulle est vérifiée,ce qui permet de définir le seuil de telle maniére que la probabilité de déclarer H0 fausse lorsqu'elle est vraie soit égale à une certaine valeur α appelée le niveau de signification du test.

On peut distinguer deux grandes familles de tests d'hypothése:
- Les tests paramétriques qui supposent une connaissance préalable de la distribution des données à quelques paramétres près.Dans ce cas les techniques statistiques recherchent les tests sur lesquels on peut fonder des stratégies présentant certains caractéres d'optimalité (Bayes,Neymann-Pearson, ect...).
- Les tests non paramétriques qui abandonnent l'idée d'un paramétrage et se placent alors dans un ensemble plus grand de lois de probabilité.On n'utilise par conséquent aucune hypothése à priori quant à la distribution des données mais plutôt certaines de leurs propriétés.Dans ce cas on ne peut définir un critére d'optimalité.Le choix d'un test, pour une alternative donnée, repose sur sa puissance. La puissance d'un test pour une hypothèse contraire à l'hypothèse nulle est la probabilité de détection de cette alternative lorsqu'elle se produit effectivement.

Il existe dans la littérature un grand nombre de tests non paramétriques pour une hypothèse donnée. Cela s'explique par le manque de critére d'optimalité et aussi par le nombre d'alternatives que l'on peut envisager.

En général, il est préferable d'utiliser des tests non paramètriques.En effet, dans la pratique, les problémes statistiques de base se posent naturellement en terme non paramétrique.De plus, la connaissance que l'on a des lois est souvent insuffisante. Les tests paramétriques risquent alors d'être très "mauvais".Ceci est particulièrement vrai pour l'étude des données d'acoustique sous-marine caractérisées par une grande variabilité.
Nous avons donc considéré essentiellement des tests non paramétriques.Pour une recension assez complète de ces tests on pourra consulter [2,7,8].

3.) PRESENTATION DES TESTS

3.1) Test du caractère aléatoire d'un échantillon

Un échantillon (x1,...,xn) est aléatoire s'il résulte d'une suite de n expériences indépendantes relatives à un caractère mesurable réel X. L'hypothèse nulle est que l'échantillon est formé de n variables aléatoires indépendantes et de même loi.
Les tests qui permettent de vérifier cette hypothèse, utilisent de diverses façons le fait que toutes les positions possibles de la valeur des observations sont alors équiprobables.

3.1.1)Test des séquences croissantes décroissantes

Soit (x1,...,xn) l'echantillon à tester.On forme la suite u1,....,un-1 définie par:

$$u_i = 1 \quad \text{si } x_i + 1 > x_i \qquad\qquad i = 1,\ n-1$$
$$u_i = 0 \quad\quad x_i + 1 < x_i$$

On s'intéresse alors à la statistique :

$$T1 = T_a + T_b$$

où Ta désigne le nombre de séquences croissantes (séquence de 1) et Tb désigne le nombre de séquences décroissantes (séquence de 0).
Sous l'hypothèse nulle la statistique est indépendante de la loi de l'echantillon (statistique libre) ; elle est asymptotiquement de loi normale :

$$N\ (\ (2n-1)/3\ ,(16n-39)/90\ )$$

3.1.2)Test des séquences croissantes décroissantes autour de la moyenne

On étudie, de même que pour le test précédent, la suite de signe 0 où 1, mais la règle de formation de cette suite est autre pour ce critère. Dans ce cas :

$$u_i = 1 \text{ si } x_i > m$$
$$u_i = 0 \text{ si } x_i < m$$

On étudie alors la statistique :

$$T2 = T_a + T_b$$

Sous l'hypothèse nulle, cette statistique est libre et asymptotiquement de loi normale :

$$N\ (\ (2n1n0)/n\ ) +1\ ,\ (2n1n0(2n1n0 - n)/(n^2(n-1))$$

où n1 désigne le nombre de 1 et n0 le nombre de 0.

3.1.3 Test de Man-Kendall

Ce test s'inspire du test de Kendall concernant l'indépendance mutuelle de deux échantillons aléatoires. En tant que test d'échantillon aléatoire, l'un des ensembles est l'ensemble des indices des données, l'autre étant l'ensemble des valeurs des données. L'hypothèse nulle est que l'ordre des indices est indépendant de l'ordre de valeurs des données. Afin d'effectuer ce test on définit :

I_i = nombre de valeurs telles que $j>i$ et $x_j < x_i$
T_i = " " " " " $j>i$ et $x_j > x_i$

On calcule la statistique

$$T3 = \sum_{i=1}^{n} T_i - \sum_{i=1}^{n} I_i$$

Ce test est libre et asymptotiquement normal :

$$N\ (0,\ 2(2n + 5)/9n(n-1)\ )$$

Test d'ajustement à la loi normale

Dans le cas général, le problème d'ajustement consiste à construire des tests qui permettent de juger si la réalisation d'un échantillon est compatible avec une répartition hypothétique déterminée, valable pour toute une population. Tous les tests ainsi formés adoptent comme hypothèse préalable que l'on dispose d'un échantillon aléatoire issu de la variable aléatoire parente.

Test de Kolmogorov Smirnov

Ce test est le plus classique des tests basés sur la fonction de répartition.Il est défini comme suit : Soit X une variable aléatoire de fonction de répartition F, et Fn la fonction de répartition empirique associée à l'échantillon $(x_1, \dots x_n)$. Le test statistique calcule :

$$S = \sqrt{n} \sup_{x \in R} |\ Fn(x) - F(x)|$$

Lorsque la loi théorique est entièrement spécifiée, cette statistique est libre sous l'hypothèse nulle et possède comme loi asymptotique :

$$\text{Prob } (S \quad y\ ) = \sum_{k=-\infty}^{+\infty} -1^k \ \exp\ (\ -2k^2y^2) \qquad y>0$$

3.2.2) Test du χ^2

Soit $(x_1 \dots x_n)$ l'échantillon aléatoire à tester. On forme une partition du domaine de variation des x_i en k intervalles I_1, I_k puis on calcule la statistique

$$\chi^2 = \sum_{j=1}^{k} (Nj - Gj)^2/Gj$$

où N_j est le nombre d'observations dans l'intervalle I_j.
G_j est la valeur moyenne du nombre d'observations dans l'intervalle I_j si les n variables aléatoires générant l'échantillon sont indépendantes et gaussiennes. Asymptotiquement le test est libre. Il suit une loi du ki-deux à k - (r-1) degrés de liberté (r désignant le nombre de paramètres estimés).

3.2.3) Test de Dagostino

Contrairement au test précédemment décrit, ce test n'est utilisable que pour la loi normale.
A partir de l'échantillon à tester $(x_1 \dots, x_n)$, on élabore l'échantillon $(u_1 \dots u_n)$ des valeurs des observations rangées dans l'ordre croissant. La statistique calcule alors :

$$D = T/n^2S$$

où S est l'estimateur classique de l'écart type :

$$S = \sqrt{\frac{(x_i - x)^2}{n}}$$

T est à une constante prés un estimateur de l'écart type de la distribution obtenu par la méthode des moindres carrés.

$$T = \sum_{i=1}^{n} (i-1/2\ (n+1))\ ui$$

Les valeurs critiques de cette statistique sont déterminées à partir de la variable centrée réduite Y, dont la distribution est tabulée [4].

$$Y = \frac{D - (2\sqrt{\pi})^{-1}}{(0.0299/\sqrt{n})}$$

3.2.4) Test du Kurtosis

Ce test est lui aussi employé spécifiquement lorsque l'on veut tester l'adéquation d'un échantillon $(x_1 \dots x_n)$ à la loi normale. La statistique s'écrit :

$$b2 = n \frac{(x_i - X)^4}{(x_i - X)^2}$$

Elle est approximativement normale :

$$N\ (\ 3(n-1)/(n+1)\ ,\ 24/n)$$

3.3) Test d'homogénéité

L'hypothése d'homogénéité peut se formuler ainsi : deux échantillons aléatoires, $(x_1 \dots, x_n)$ et $(y_1 \dots yk)$ sont homogènes si la distribution de la variable aléatoire parente de X est la même que celle de Y. Pour tester cette hypothèse, nous utilisons deux des tests les plus usuels.

3.3.1) Test de Wilcoxon

Pour élaborer ce test on détermine le rang de chacune des valeurs de l'échantillon pris dans son ensemble et on calcule la somme des rangs du premier échantillon c'est-à-dire :

$$W = \sum_{i=1}^{n} Ri \qquad \text{Ri est le rang de xi}$$

Si les échantillons sont indépendants la statistique est libre et asymptotiquement normale:

$$N\ (\ n(n+k+1)/2\ ,nk(n+k+1)/12)\ )$$

3.3.2) Test de Kolmogorov - Smirnov

Pour réaliser ce test on ordonne l'échantillon total dans l'ordre croissant des observations. Soit $z_1 \ldots z_{m+k}$ l'échantillon ainsi formé, le test statistique est alors :

$$KS = \max | F_M(zi) - G_K(zi) | \qquad i = 1 , m + k$$

où $F_M(zi)$ et $G_K(zi)$ représentent respectivement les fonctions de répartition empiriques des échantillons X et Y définis comme suit :

$$F_M(zi) = (\text{nombre de valeurs de } X \leq Z_i)/M$$
$$G_K(zi) = (\text{nombre de valeurs de } Y \leq Z_i)/K$$

4.) EVALUATION DE LA PUISSANCE DES TESTS

Il n'existe pas de solution universellement valide pour le choix d'un test parmi d'autres.Autremement dit il convient de définir le type d'alternative pour définir le test le "meilleur".L'ensemble d'éléments disponibles dans la littérature concernant les performances des tests présentés est plutôt limité. Les résultats théoriques existant ne concernent que des situations asymptotiques.Lorsqu'on a affaire à des échantillons de taille finie la solution consiste à mener une étude de puissance par simulation numérique.Tous les tests présentés ont donc été soumis à une étude expérimentale en simulant quelques alternatives choisies pour leur intérêt.Les tests sont toujours utilisés de façon bilatérale.

4.1) Puissance des tests d'échantillon aléatoire

Nous avons appliqué ces tests sur des échantillons gaussiens normalisés, obtenus par simulation, auxquels on fait subir une transformation décrivant l'alternative.Nous avons examiné trois types d'alternatives : alternative de position, d'échelle, de dépendance statistique. Cette étude a été menée pour deux valeurs de la taille de l'echantillon (n=100,500). Le nombre d'épreuves simulées M est suffisament grand pour que les résultats soient significatif au niveau de signification choisi (α = 10%). Les résultats complets sont exposés dans [4,5].

4.1.1 Alternative de position

Sous cette rubrique nous considérons trois alternatives qui modifient la valeur moyenne de l'échantillon.

.Tendance affine

On fait subir à chaque échantillon la transformation suivante:

$$xi = b\, i/n + gi \qquad i = 1,n \quad \text{avec } gi \quad N(0,1)$$

les résultats sont fournis sous forme de tableau qui présente la puissance des tests en fonction de la pente de la tendance pour α =10%.

	n=100 M=1000			n =500 M =200		
b	T1	T2	T3	T1	T2	T3
0	7.6	9.8	9.3	14.5	12.0	9.5
1	8.1	15.1	88.1	14.5	34.0	100
5	8.5	100	100	13.5	100	100

La première conclusion qui s'impose est que T1 est incapable de détecter une tendance indépendamment de la taille de l'échantillon, T3 pouvant détecter une pente bien inférieure à celle détectée par T2. Bien entendu les résultats s'améliorent avec la taille de l'échantillon.

.Changement de moyenne

Dans ce cas, on modifie la valeur moyenne de la moitié de l'échantillon:

$$xi = gi + m \qquad i = n/2,n$$

Qualitativement les résultats sont identiques au cas précédent.T3 peut détecter des variations de moyenne 5 fois plus faible que T2.

.Présence d'une fréquence pure

On opère la transformation suivante sur l'échantillon

$$xi = gi + \sqrt{Et} \cos (\pi f\, i/n) \qquad i=1,n$$

Nous nous sommes intéressés à deux cas extrêmes de fréquence pour diverses valeurs du rapport signal sur bruit (RSB = Et/σ_g^2).
Lorsque la période de la sinusoïde est de l'ordre de la taille de l'échantillon, on retrouve le même comportement que pour les cas précédents.
Si la fréquence du signal augmente,T1 est alors le plus sensible,T3 est alors incapable de mettre en évidence cette alternative (voir tableau).

n=500 M =200

	f =1			f= 150		
RSB	T1	T2	T3	T1	T2	T3
0.01	14.5	9.5	49.5	14.5	13.0	10.0
0.04	14.5	11.0	91.0	14.5	10.5	9.0
0.25	14.0	53.0	100	35.5	25.5	6.5

4.1.2) alternative d'échelle

Pour simuler une variation de la variance de l'échantillon on opére la transformation suivante:

$$xi = gi \qquad i=1,n/2 -1$$
$$xi = \sqrt{a}\, gi \qquad i=n/2 \ ,n$$

Dans cette situation les résultats obtenus sont trés mauvais quelque soit la taille de l'échantillon. Pour améliorer le pouvoir de détection de ces tests,on utilise non plus les données brutes mais la valeur absolue des données centrées ($yi = |xi - m|$).

Cette procédure préconisée dans [1] permet d'améliorer les performances des tests de façon significative (voir tableau).

	n=100 M=1000					
a	T1	T2	T3	T1y	T2y	T3y
1.5	7.8	10.3	12.1	10.7	12.1	14.2
2.0	7.6	9.7	9.2	11.0	11.9	22.0
0.1	7.6	9.5	9.4	9.6	80.7	100

4.1.3 Dépendance statistique

Tous les tests utilisés supposent que les données sont indépendantes entre elles.Or,cette hypothése n'est en général pas vérifiée dans la pratique. Pour aborder ce probléme nous avons étudié une dépendance autorégréssive du premier ordre en effectuant la transformation suivante sur l'échantillon.

$$xi = \sum_{j=0}^{n} c^j\, gi-j \qquad i=1,n$$

Les résultats regroupés dans le tableau suivant mettent en évidence la supériorité des tests de séquences.

	n=100 M=1000			n =500 M =200		
c	T1	T2	T3	T1	T2	T3
0.1	13.5	15.5	13.0	29.0	41.5	10.0
0.2	23.1	33.4	16.7	60.5	88.5	18.5
0.5	64.5	94.0	31.7	99.5	100	31.0

4.2 Puissance des tests de normalité

Vu l'importance de la loi gaussienne en statistique ce probléme a été considérablement étudié.La littérature fourmille de résultats de puissance pour de nombreuses procédures.De façon générale les statistiques basées sur la fonction de répartition empirique sont plus puissantes que le test du χ^2 lorsque les paramétres de la distribution sont connues [7].

Celui-ci reste trés utilisé car on dispose d'une bonne approximation de la loi de la statistique lorsque les paramètres sont estimés sur l'échantillon.Ce dernier avantage est rendu désuet par les résultats de Stéphens en ce qui concerne la loi gaussienne.Dans ce cas Stephens [9] donne les modifications à apporter au test de Kolmogorov,qui permettent de déterminer sa distribution asymptotique.
Pour de faibles valeurs de la taille de l'échantillon ($n<100$) on dispose de résultats numériques. La comparaison du test de D'Agostino avec le test de Kolmogorov tourne à l'avantage du premier si les paramètres sont supposés connus. Par contre lorsqu'il sont estimés sur l'échantillon les résultats ,s'inversent [9]. L'étude comparative des performances du test de D'Agostino avec le Kurtosis menée en [6] montre que ce dernier est généralement plus puissant lorsque la déviation de la loi gaussienne est due essentiellement au kurtosis;les différences observées sont néanmoins trés faibles.Ce qui peut militer en faveur du test de D'Agostino est qu'il est capable de détecter des déviations de normalité dues aussi bien au kurtosis qu'à la dissymétrie de la distribution.
Nous avons voulu comparer ces tests pour des tailles d'échantillons que nous manipulons habituellement en acoustique sous-marine.Pour ce faire nous avons effectué une étude de puissance en considérant comme alternative une loi mélange de gaussienne qui s'avère trés utile pour modéliser des bruits non-gaussiens[3]
La loi simulée est :

$$* \quad p(x) = 0.3\ f(x,0,2.7) + 0.7\ f(x,0,0.27)$$

$$\text{où } f(x,m,\sigma^2) \simeq N\ (m,\sigma^2)$$

Cette loi est symétrique et posséde un kurtosis de 6. Cette étude menée sur 200 échantillons de taille 500 fournit des résultats remarquables.En effet tous les tests réfutent l'hypothése gaussienne à 100%.

4.3) Puissance des tests d'homogénéité

Tout d'abord on a considéré les mêmes alternatives que pour les tests d'échantillon aléatoire.Les tailles des deux échantillons étant identiques la transformation est appliquée sur le second échantillon le premier restant gaussien normalisé.Les résultats complets sont donnés en [4] .On se contente ici de les commenter.On s'est par ailleurs intéressé au cas où les deux échantillons ont une valeur moyenne et une variance identique mais des distributions différentes.

4.3.1) Alternative de position

Lorsque l'on superpose sur le second échantillon une tendance affine, la non homogénéité est détectée de façon satisfaisante par les deux tests pour des valeurs trés faibles de la pente de la tendance.Le test de Wilcoxon est légérement plus efficace.Par exemple pour une pente de 0.08 le test de Kolmogorov détecte 75% des échantillons de taille 100 à comparer à 87% pour Wilcoxon.
Lorsque l'alternative est une simple différence de moyenne on retrouve le même comportement.La difference de puissance observée est de l'ordre de 10%.
Si l'on superpose une fréquence pure sur le second échantillon, on remarque tout d'abord que les résultats sont indépendants de la fréquence de la sinusoïde.Le test de Wilcoxon est alors totalement inefficace.Quant au test de Kolmogorov, pour atteindre une puissance de 90% , il faut attendre une valeur de rapport signal sur bruit de 4.Il faut noter qu'un test d'échantillon aléatoire aurait mis en évidence cette alternative bien plus efficacement.

4.3.2) Alternative d'échelle

Lorsque l'on teste l'homogénéité de deux échantillons de variances différentes en utilisant les données brutes, le test de Wilcoxon est insensible à cette différence, le test de Kolmogorov ne détecte à 100% que lorsque le rapport des variances est de 90%.
Si l'on utilise les tests sur la valeur absolue des échantillons, les tests deviennent alors efficaces. Ainsi un rapport de variance de 50% est détecté dans 51% des cas pour le test de Wilcoxon et 40% pour Kolmogorov sur des échantillons de taille 100.

4.3.3) Distributions différentes

La question posée est de savoir si les tests utilisés peuvent détecter une non-homogénéité due à des distributions différentes indépendamment des paramètres de position et d'échelle. Pour ce faire, on teste l'homogénéité entre deux échantillons de lois différentes préalablement centrées et réduites.
Le premier échantillon étant gaussien,on simule tour à tour quatre lois alternatives pour le second (χ^2_2, χ^2_{10}, uniforme, loi mélange de gaussienne). Les résultats mettent en valeur la totale inefficacité du test de Wilcoxon face à ces diverses alternatives. L'efficacité du test de Kolmogorov n'étant appréciable que lorsque la distribution est trés dyssimétrique. Les études théoriques prouvent que ce dernier est asymptotiquement efficace,les résultats obtenus pour des échantillons de taille 100 suggérent donc que la convergence vers la situation asymptotique est relativement lente.En fait un test d'adéquation appliqué sur l'échantillon aurait mis en évidence cette alternative.

	Uniforme	χ^2_2	χ^2_{10}	Mélange*
KS	0.0	52.0	23.0	7.0

5.) CONCLUSION

L'étude de puissance des tests du caractére aléatoire a mis en évidence la supériorité du test de Mann Kendall pour des déviations de position.Les tests de séquences sont à préférer lorsque l'on suspecte la présence d'une sinusoïde dans l'échantillon ou la dépendance statistique des données.Pour détecter une alternative d'échelle il est toujours préférable d'utiliser les tests sur les valeurs absolues des données.
Les tests de normalité ne nécessitent pas des tailles d'échantillon elevées pour donner des résultats satisfaisants.Il est préférable de les utiliser conjointement afin d'avoir une meilleure idée sur la distribution alternative.
En ce qui concerne les tests d'homogénéité ,il sont relativement efficaces pour détecter des déviations dues aux paramétres de la distribution, mais ils sont insensibles aux alternatives liées à la forme de la distribution.

6.) BIBLIOGRAPHIE

[1] C.R BAKER Statistical tests for the analysis of sonar data. U.S.N.J of underwater acoustic,1976

[2] J.V BRADLEY Distribution-free statistical tests. Prentice-Hall,1968

[3] H. CHERIFI Modélisation,Identification,détection de bruit non gaussiens.CEPHAG N°30/86

[4] H. CHERIFI Evaluation de tests d'hypothéses. CEPHAG N°13/87

[5] M. CHRAIBI Tests statistiques.CEPHAG N°42/86

[6] R.B D'AGOSTINO An omnibus test of normality for moderate and large size samples. Biometrika 58,1971.

[7] M.G KENDALL & A STUART The advanced theory of statistics griff.,1958

[8] J.L PIEDNOIR Statistiques non paramétriques. Revue du CETHEDEC 3ème trimestre 1977

[9] M.A STEPHENS EDF statistics for goodness of fit and some comparisons.J.A.Stat.Ass 69,1974