

A. KALLI

**Programme pour la discrimination barycentrique
entre deux populations et l'étude par permutation
aléatoire de la discriminabilité entre celles-ci**

Les cahiers de l'analyse des données, tome 12, n° 2 (1987),
p. 129-146

http://www.numdam.org/item?id=CAD_1987__12_2_129_0

© Les cahiers de l'analyse des données, Dunod, 1987, tous droits réservés.
L'accès aux archives de la revue « Les cahiers de l'analyse des données » implique l'accord avec les conditions générales d'utilisation (<http://www.numdam.org/conditions>). Toute utilisation commerciale ou impression systématique est constitutive d'une infraction pénale. Toute copie ou impression de ce fichier doit contenir la présente mention de copyright.

NUMDAM

Article numérisé dans le cadre du programme
Numérisation de documents anciens mathématiques

<http://www.numdam.org/>

PROGRAMME POUR LA DISCRIMINATION BARYCENTRIQUE ENTRE DEUX POPULATIONS ET L'ETUDE PAR PÉRMUTATION ALEATOIRE DE LA DISCRIMINABILITE ENTRE CELLES-CI

[PROG. DISC. ALEAT.]

A. Skalli ()*

1. Position et intérêt de l'étude

L'objet du présent article est l'étude de la validation des résultats d'une analyse discriminante barycentrique. La discrimination entre deux sous-ensembles en analyse des données s'impose fréquemment. C'est toujours le cas pour un ensemble d'individus composé de deux groupes ou deux catégories. En médecine, on étudie souvent un ensemble de sujets, les uns sont normaux les autres sont atteints d'une pathologie ou encore les uns sont décédés, les autres ont survécu après le même traitement. Dans les milieux bancaires on étudie des dossiers de crédits où les uns correspondent à des bons payeurs les autres à des mauvais. Pour les dépouillements d'enquête, on a souvent besoin de comparer ou d'opposer l'attitude des hommes à celles des femmes.

Ici l'on s'appuiera sur deux exemples pour illustrer nos propos: le premier concerne des mammifères vivant les uns en Guyane les autres au Gabon (C.A.D. Vol XI n° 4, 1986, pp. 403-440); le deuxième étudie des cellules normales et transformées par le virus polyome (C.A.D. VOL. XII, 1987, n° 1, pp.7 - 32).

(*) Laboratoire de statistique. Université Pierre et Marie Curie. 75252 PARIS CEDEX 05
Ecole Mohammadia d'ingénieurs. B.P. 765 Rabat. Maroc.

Notons que le problème auquel nous nous intéressons le plus particulièrement et qui est généralement plus difficile à discuter est celui où les deux groupes sont très semblables. En pareil cas, il n'est pas question de réaliser une discrimination avec un taux de bien classés suffisant pour être utile; mais se pose la question qu'en terme de statistique classique on appellerait: question de l'identité de la loi parente des deux populations. (Même si en fait les deux populations sont des ensembles clos qu'on ne peut considérer comme issus par échantillonnage d'une source aléatoire). Le premier exemple des formes des mammifères vivant dans deux pays différents (Chaque espèce étant décrite par son mode de vie) est caractéristique de cette situation puisqu'il y a recouvrement presque total des deux faunes. Les variables responsables de la petite différence existant entre celles-ci seront cependant mises en évidence.

En revanche, le deuxième exemple traite un cas où les deux groupes de cellules sont bien distincts: la discrimination est bonne et le fait que les deux populations ne se recouvrent pas est *a fortiori* hors de doute.

D'une manière générale on peut imaginer un intervalle allant de la discrimination parfaite au recouvrement total et le problème qui nous préoccupe consiste à situer le cas traité par discrimination par rapport au recouvrement total.

Cependant, dans la plupart des cas, au lieu d'appliquer directement la discrimination, on aura avantage à utiliser d'abord une méthode d'analyse factorielle des correspondances ou de classification permettant éventuellement de mettre en évidence des groupements privilégiés ou imprévus. Cela peut suggérer un changement de la démarche préétablie dans le but d'améliorer le pouvoir de discrimination (cf. [DISC. MORPH. CELL.] C.A.D.XII n° 1 pp. 7 - 32).

2. Rappel sur l'Analyse Discriminante Barycentrique (ADB)

2.1. La méthode : En analyse discriminante, on cherche à séparer deux sous-ensembles d'individus décrits par les mêmes variables. La méthode d'analyse discriminante que nous utilisons ici est l'analyse barycentrique (cf. ENS2 II pp. 102-104). On considère deux groupes G1 et G2 d'individus de cardinal n_1 et n_2 décrits par le même ensemble J de variables ou de modalités de variables logiques. On construit les 2 centres de gravité N et P des 2 sous-ensembles affectés d'un poids égal. Plus précisément on cumule toutes les lignes du même groupe et l'on divise le résultat par l'effectif ou plus généralement par le total de la ligne. Cette dernière opération donne aux deux classes le même poids. On représente ensuite chacune des modalités par un point j du segment NP tel que $|jN / jP| = \text{cum}(P,j) / \text{cum}(N,j)$ où $\text{cum}(N,j)$ désigne le nombre d'individus possédant la modalité j rapporté à l'effectif total du groupe ou plus généralement les rapports des valeurs qu'ont pour j les lignes P et N. Puis l'on place chaque individu i au barycentre des modalités j qu'il possède (ou plus généralement au barycentre des j affectés des masses $k(i,j)$). La représentation sur cet axe unique ainsi obtenue, au prix de calculs minimes, aboutit à une discrimination de l'ensemble des individus. Le point de discrimination sur cet axe est le milieu du segment NP du fait que les 2 classes ont même poids.

Cette analyse barycentrique équivaut à analyser un tableau de correspondance $\{N,P\} \times J$ dont les deux lignes principales sont les descriptions moyennes des individus qui composent respectivement chaque groupe G_1 et G_2 . Les n_1+n_2 individus sont adjoints en lignes supplémentaires. L'axe unique issu de cette analyse est celui construit par les calculs barycentriques.

En termes géométriques, et en dehors du cadre même de l'analyse des correspondances on considérera, si deux populations G_1 et G_2 sont représentées par des nuages N_1 et N_2 de points dans un espace euclidien E , (N_1 et N_2 étant amenés à avoir même masse totale), que tout individu i_1 de G_1 , ou i_2 de G_2 peut être projeté orthogonalement sur la droite g_1g_2 joignant les centres de gravité des deux nuages: et la discrimination proposée consiste à affecter cet individu (ou un individu nouveau is) à celui des groupes du centre auquel sa projection est le plus proche (i.e. du côté auquel elle se trouve sur g_1g_2 relativement au milieu). Sous cette forme générale, nous croyons pouvoir dire que la méthode de discrimination remonte à Sir R. Fisher.

2.2 Mode d'interprétation des résultats de l'ADB : Deux résultats importants doivent attirer l'attention: le premier apparaît du fait de la projection simultanée sur le même axe des individus et des modalités de variables. Il importe alors de noter celles qui caractérisent le mieux chacun des deux groupes d'individus si ces derniers sont bien discriminés. Le deuxième résulte du calcul des pourcentages partiels des individus de chacun des groupes qui ont des coordonnées barycentriques du côté opposé à celui du centre de leur classe relativement à l'origine de l'axe. Ces taux d'erreur nous permettent d'apprécier la qualité de la discrimination.

Lorsque le taux d'erreur est très faible de l'ordre de 10% la discrimination est nette. Quant ce taux augmente, la description spécifique de chaque groupe devient de plus en plus floue. Dans les cas limites on peut se demander si la répartition des individus n'est pas proche d'une répartition aléatoire. Pour cela on se place d'un point de vue inverse et l'on compare la répartition réelle des individus sur les 2 groupes à celle où les deux groupes seraient composés d'individus tirés au hasard. Cette répartition aléatoire sera simulée un grand nombre de fois (une centaine ou plus), et à chaque fois une ADB est réalisée (la simplicité des calculs le permet cf. § 2.1), pour avoir une tendance générale des cas aléatoires.

3. Simulations

3.1 Simulations aléatoires: On considère un ensemble I d'individus caractérisés par un ensemble J de variables ou de modalités de variables. I est la réunion de I_1 et I_2 connus *a priori*. La comparaison de ces deux ensembles, par une analyse discriminante barycentrique, permet de mettre en évidence leurs différences et ressemblances. Pour comparer cette répartition réelle à une répartition où les individus sont affectés aléatoirement à l'un ou l'autre groupe on procède par des simulations. Dans chacune on construit des ensembles I_1 et I_2 remplis aléatoirement par des individus de I en respectant uniquement les cardinaux de I_1 et I_2 du cas réel.

3.2 Description d'une simulation: On pose au départ $r_1=n_1$ et $r_2=n_2$ les cardinaux de I_1 et I_2 dans le cas réel. On lit le premier individu de I , on tire au hasard un nombre z compris entre 0 et 1 ; si ce nombre est inférieur ou égal à r_1/r_1+r_2 on affecte alors le premier individu à la classe 1 sinon on l'affecte à la classe 2 et on déduit de r_1 sinon de r_2 l'unité. On lit ensuite le deuxième individu et l'on recommence jusqu'à la lecture complète des individus de I . On obtient ainsi une partition de l'ensemble I en deux classes de mêmes effectifs que I_1 et I_2 réels. On réalise l'analyse discriminante barycentrique sur le tableau obtenu. On recommence cette simulation un grand nombre de fois.

3.3 Discussion d'une simulation : Notons que la lecture des individus se fait séquentiellement du premier au dernier en affectant chacun à la 1-ère ou à la 2-ème classe selon que la proposition ($z \leq r_1/r_1+r_2$) est vraie ou fausse.

Il est possible de procéder autrement en tirant au hasard des nombres entiers compris entre 1 et n_1+n_2 et l'on affecte les n_1 premiers nombres tirés à la première classe et le restant à la deuxième. Ce procédé est plus coûteux en temps machine à cause des répétitions possibles des nombres à tirer. La probabilité de tirer un nombre au hasard parmi n_1+n_2 est de $1/(n_1+n_2)$ mais en pratique cela est très différent et les mêmes nombres peuvent sortir plusieurs fois, il faudra alors retenir le premier et éliminer les autres pour éviter la répétition d'individus. Cet algorithme doit continuer jusqu'à l'obtention de n_1 individus. Le temps d'exécution serait grand et nous avons alors préféré un algorithme plus simple et de temps d'exécution prévisible.

3.4 Mode d'interprétation des résultats des simulations: Pour apprécier l'éloignement du cas réel aux cas simulés aléatoirement on étudie principalement l'évolution de la valeur propre du cas réel et de chaque simulation, les nombres d'individus bien placés dans le groupe 1, dans le groupe 2 et leur total.

Rappelons que la valeur propre caractérise la dispersion du nuage. Celle-ci augmente lorsque la distance séparant les deux centres de gravité des 2 classes croît, ce qui correspond à une discrimination plus nette. La valeur propre et la somme des nombres d'individus bien placés dans le groupe 1 et dans le groupe 2 donnent sous des formes voisines, mais non identiques, une indication sur la discrimination.

Pour l'interprétation et en utilisant un langage de la théorie des tests, nous mettons à l'épreuve l'hypothèse nulle H_0 que la répartition des individus sur les deux groupes du cas réel est elle-même aléatoire et l'on observe la position des paramètres du cas réel par rapport à ceux des cas simulés aléatoirement. On peut alors calculer le risque d'erreur maxima commis en annulant H_0 . Le centile du paramètre du cas réel nous permet de calculer ce risque. Pour une meilleure visualisation de la variation de ces paramètres, un histogramme de l'ensemble des valeurs de chaque paramètre sera tracé.

3.5 Propriété géométrique de l'A.D.B. et intérêt des simulations Notons une propriété avantageuse de l'A.D.B. qui découle de la projection en éléments supplémentaires de tous les individus sur la droite des barycentres. Celle-ci, mieux

que toute autre, permettra, par la détermination d'un point , une bonne séparation des deux groupes. Considérons un cas limite où les deux groupes seraient très ressemblants; c'est le cas précisément où les individus sont tirés aléatoirement. Le taux de succès est certainement supérieur à 50%. Cela provient de cette propriété géométrique de l'A.D.B.. En effet tout se passe comme si l'on cherchait à se positionner dans l'espace pour avoir la meilleure séparation entre les deux groupes définis *a priori* . Il est naturel de penser que la projection sur la droite des deux centres de gravité des deux groupes donnera cette meilleure séparation entre ceux-ci. C'est la raison pour laquelle même dans le cas où les groupes sont composés aléatoirement, le taux de succès est nettement supérieur à 50%. Pour permettre dans ces cas limites une meilleure appréciation de la discrimination nous proposons cette comparaison entre les résultats de la discrimination du cas réel et celles des cas simulés aléatoirement. Si les paramètres obtenus dans le cas réel sont respectivement voisins de ceux obtenus dans les cas simulés on peut en déduire une forte ressemblance entre les deux groupes.

Signalons que l'intérêt de cette méthode augmente lorsque les deux groupes donnés se rapprochent.

4. Description de l'algorithme

4.1 Listage de l'algorithme

Pour S:=1 pas 1 jusqu'à NS faire début

R[1]:=N[1]; R[2]:=N[2];

pour C:=1 pas 1 jusqu'à 2 faire pour J:=1 pas 1 jusqu'à CJ faire
CUM[C,J]:=0;

CI:=N[1]+N[2];

pour I:=1 pas 1 jusqu'à CI faire début

CC:= si $Z \leq R[1]/(R[1]+R[2])$ alors 1 sinon 2;

CLAS[I]:=CC; R[CC]:=R[CC]-1;

Pour J:=1 pas 1 jusqu'à CJ faire

CUM[CC,J]:=CUM[CC,J]+K[I,J] fin;

SOM[1]:=0

SOM[2]:=0

Pour C:=1 pas 1 jusqu'à 2 faire

Pour J:=1 pas 1 jusqu'à CJ faire SOM[C]:=SOM[C]+CUM[CJ];

pour C:=1 pas 1 jusqu'à 2 faire

pour J:=1 pas 1 jusqu'à CJ faire CUM[C,J]:=CUM[C,J]/SOM[C];

pour J:=1 pas 1 jusqu'à CJ faire

GJ[J]:=(CUM[1,J]-CUM[2,J])/(CUM[1,J]+CUM[2,J]);

pour I:= 1 pas 1 jusqu'à CI faire début

FI[I]:=0;

KI:=0;

pour J:=1 jusqu'à CJ faire

FI[I]:=FI[I]+(K[I,J]*GJ[J]);

KI:=KI + K[I,J];

FI[I]:=FI[I]/KI fin;

LAM[S]:=0;

```

pour J:=1 pas 1 jusqu'à CJ faire LAM[S]:=LAM[S]+| CUM[1,J]/NV*GJ[J]| ;
FC[1]:=1; FC[2]:=-1; JUS[1]:=0; JUS[2]:=0; FAU[1]:=0; FAU[2]:=0;
pour I:=1 pas 1 jusqu'à CI faire
si FI[I]*FC[CLAS[I]]>0 alors JUS[CLAS[I],S]=JUS[CLAS[I],S]+1
                               sinon FAU[CLAS[I],S]=FAU[CLAS[I],S]+1
fin

```

4.2 Commentaire de l'algorithme

L'algorithme présenté ici représente le noyau central de l'ensemble des simulations et discriminations barycentriques. Pour la description des entrées et surtout des sorties il suffit de se reporter au listage complet du programme en Fortran au § 5.

Soit S le rang de la simulation variant de 1 à NS . EI et EJ sont les ensembles des individus et des modalités de variables, CI et CJ représente respectivement leurs cardinaux. Posons $N[1]$ cardinal de la première classe, $N[2]$ cardinal de la deuxième classe d'où $CI=N[1]+N[2]$. $R[1]$ et $R[2]$ sont initialisés respectivement à $N[1]$ et $N[2]$ et déduits d'une unité chaque fois qu'un individu est affecté à l'une des deux classes.

Z est un nombre tiré au hasard compris entre 0 et 1. La procédure générant ces nombres aléatoires n'est pas représentée dans cet algorithme mais figure dans le programme Fortran. On note CC le numéro de la classe qui prend la valeur 1 ou 2 selon que Z est inférieur ou supérieur à $R[1] / (R[1] + R[2])$.

On note $CUM[CC,J]$ la somme des paramètres J sur tous les individus de la classe CC .

$K[I,J]$ est le tableau initial.

$CUM[C,J]$ est ensuite divisé par $SOM[C]$ la somme totale des valeurs de la classe C .

$GJ[J]$ et $FI[I]$ sont respectivement les coordonnées barycentriques des variables J et des individus I .

Les paramètres caractéristiques de chaque simulation sont :

$LAM[S]$ valeur propre

$JUS[CLAS[I],S]$ nombre de bien classés de $CLAS[I]$ ($CLAS[I]=1$ ou 2)

$FAU[CLAS[I],S]$ nombre de mal classés de $CLAS[I]$

$JUST[S]=JUS[1,S]+JUS[2,S]$


```

dimension cum(2,250),g(250),fi(800),txf(2),fc(2)
integer r(2),c,c1,cj,cc,s,n(2),clas(800),i1f(800),ig(250)
integer jus(2,0:999),fau(2),it(2,250),it1(250),justt(0:999)
real lam(0:999),k(800,250),nv
character nom(800)*4,idf(2,250)*4,id*1,ic(2)*1,ir(2)*1
character fice*8,fics*8,fmte*80,idj(250)*4,leg(2)*20
character a(120)*1,d(120,40)*4
call lecpa(c1,cj,ns,lh,ic,ir,fice,fics,fmte,idj,leg,nv)
call lecdon(c1,cj,ic,fice,fmte,k,cum,n,clas,id,nom)
call descri(c1,cj,ns,lh,ic,ir,fics,idj,cum,k,fi,gj,n,nom,
1 lam,fc,clas,a,d,ig,i1f,it,it1,jus,fau,txf,justt,leg,nv)
end
subroutine lecpa(c1,cj,ns,lh,ic,ir,fice,fics,fmte,idj,leg,nv)
integer ci,cj
real nv
character ic(2)*1,ir(2)*1,fice*8,fics*8,fmte*80,idj(250)*4
character leg(2)*20
c Ce programme realise la discrimination barrycentrique de deux classes
c selon le premier caractere de chaque individu
write(*,100)
100 format('Entrez dans l ordre et sans separateur en format '
1 'i3,4a1' /
2 'le nombre total de lignes '/'de variables '/'de modalites ' /
3 'de simulations desirees (0 sinon) '/' de classes desirees dans '
4 ' l histogramme '/'les 2 caracteres distinctifs des 2 classes' /
5 '2 caracteres choisis pour la representation sur l histogramme')
read(*,101)ci,nv,cj,ns,lh,ic(1),ic(2),ir(1),ir(2)
101 format(i3,f3.0,3i3,4a1)
write(*,*)ci,nv,cj,ns,lh,ic(1),ic(2),ir(1),ir(2)
write(*,106)ir(1),ir(2)
106 format('Entrez pour la legende de l'histo. en 2a20 ' /
1 'la definition en clair de ',a1,' et de ',a1)
read(*,109)leg(1),leg(2)
109 format(2a20)
write(*,102)
102 format('Entrez en 2a8 les noms des fichiers de donnees '
1 ' et des resultats')
read(*,103)fice,fics
103 format(2a8)
open(1,file=fice,status='old')
rewind 1
write(*,104)cj
104 format('Entrez le format de lecture des donnees'/' commençant
1 par (a1,a4, suivi en format fn.d des ',i3,' variables')
read(*,105)fmte
105 format(a80)
go to 60
write(*,107)cj,cj
107 format('Entrez les identificateurs des ',i3,
1 variables en'i3,'a4')

```

```

60   read(1,108)(idj(j),j=1,cj)
108  format(20a4)
      write(*,108)(idj(j),j=1,cj)
      end
      subroutine lecdon(c1,cj,ic,f1ce,fmte,k,cum,n,clas,id,nom)
      integer ci,cj,n(2),clas(800)
      real k(800,250),cum(2,250)
      character ic(2)*1,f1ce*8,fmte*80,id*1,nom(800)*4
      do 50 i=1,ci
      read(1,fmte)id,nom(i),(k(i,j),j=1,cj)
      if(id.eq.ic(1)) then
      clas(i)=1
      else
      clas(i)=2
      endif
50   continue
      end
      subroutine cumul(ci,cj,k,cum,clas,n)
      integer ci,cj,clas(800),n(2),c
      real k(800,250),cum(2,cj),som(2)
      do 51 c=1,2
      do 51 j=1,cj
51   cum(c,j)=0.
      n(1)=0
      n(2)=0
      do 52 i=1,ci
      c=clas(i)
      n(c)=n(c)+1
      do 52 j=1,cj
52   cum(c,j)=cum(c,j)+k(i,j)
      som(1)=0.
      som(2)=0.
      do 55 j=1,cj
55   som(1)=som(1)+cum(1,j)
      som(2)=som(2)+cum(2,j)
      do 54 c=1,2
      do 53 j=1,cj
53   cum(c,j)=cum(c,j)/(som(c))
54   continue
      end
      subroutine discr1(ci,cj,ns,lh,ic,ir,fics,idj,cum,k,f1,gj),
1   n,nom,lam,fc,clas,a,d,ig,ilf,it,lt1,jus,fau,txf,justt,leg,nv)
      integer ci,cj,n(2),clas(800),ilf(800),ig(250),it(2,120)
      integer lt1(120),cc
      integer r(2),s,jus(2,0:999),fau(2),justt(0:999)
      real cum(2,250),k(800,250),lam(0:999),f1(800),gj(250),fc(2),nv
      real g(0:999),h(0:999)
      character ic(2)*1,ir(2)*1,fics*8,idj(250)*4,nom(800)*4
      character idf(2,300)*4,a(120),d(120,120)*4,v1*4,v2*4,ldh(0:999)*4
      character titre*60,vr*10,leg(2)*20

```

```

s=0
call cumul(c1,cj,k,cum,clas,n)
call factor(c1,cj,k,cum,f1,g1,lam,fc,clas,n,s,nv)
open(2,file=fics,status='new')
write(2,118)
118 format(////,10x,'LABORATOIRE DE STATISTIQUE UPMC'//
1 10x,'PROFESSEUR J.P. BENZECRI'//
2 10x,'AUTEUR : A. SKALLI'////)
write(2,119)
119 format(20x,'RESULTAT DE L'ANALYSE DISCRIMINANTE '
1 'BARYCENTRIQUE'/20x,49('_')////)
write(2,121)
121 format(///'+',7('-----'),'---+ '/
1 '!Sim. n0 * val.propre*nb. gr.1*b.cl.gr1*'
2 '% erreur1*b.cl.gr2* % erreur 2*jus.tot!%err.tot!'/
3 '+',7('-----'),'---+ ')
call disbar(c1,cj,k,cum,f1,gj,fc,jus,fau,s,n,nom,clas,fics,lam
1 ,txf,justt)
write(2,122)
122 format('+',7('-----'),'---+ ')
call impres(f1,gj,ci,cj,lh,ic,ir,clas,idj,fics,nom
1 ,a,d,ig,1f,it,1t1,idf,fc,leg)
if(ns.eq.0)return
ix=99
write(2,112)ns
112 format(///20x,'RESULTATS DES ',i3,' SIMULATIONS'//
1 20x, ' _____',////)
write(2,111)
111 format(///'+',7('-----'),'---+ '/
1 '!Sim. n0 *val.propre * nb.gr.1*b.pl.gr1*'
2 '% erreur 1*b.pl.gr2* % erreur 2*jus.tot!%err.tot!'/
3 '+',7('-----'),'---+ ')
do 66 s=1,ns
r(1)=n(1)
r(2)=n(2)
do 52 i=1,c1
t=real(r(1))/(real(r(1))+real(r(2)))
call randu(ix,iy,hazard)
ix=iy
if(hazard.le.t)then
cc=1
else
cc=2
endif
clas(i)=cc
r(cc)=r(cc)-1
52 continue
call cumul(c1,cj,k,cum,clas,n)
call factor(c1,cj,k,cum,f1,gj,lam,fc,clas,n,s,nv)
call disbar(c1,cj,k,cum,f1,gj,fc,jus,fau,s,n,nom,clas,fics,lam

```

```

1 ,txf,justt)
66 continue
   titre='HISTOGRAMME DES VALEURS PROPRES DE TOUTES LES SIMULATIONS '
   call histog(lam,ns,titre,lh)
   do 50 i=0,ns
50 lam(i)=real(jus(1,i))
   titre='HISTOGRAMME DES NBS D INDIVIDUS BIEN PLACES DU GROUPE 1'
   call histog(lam,ns,titre,lh)
   do 51 i=0,ns
51 lam(i)=real(jus(2,i))
   titre='HISTOGRAMMES DES NBS D INDIVIDUS BIEN PLACES DU GROUPE 2'
   call histog(lam,ns,titre,lh)
   do 62 i=0,ns
62 lam(i)=real(justt(i))
   titre='HISTOGRAMME DES TOTAUX BIEN PLACES (JUSTE GR1+GR2) '
   call histog(lam,ns,titre,lh)
   v1='N1 '
   v2='N2 '
   do 60 i=0,ns
   g(i)=real(jus(1,i))
   h(i)=real(jus(2,i))
   write(idh(i),100)
60 continue
100 format('s',i3)
   call graph(g,h,idh,v1,v2,ns)
   call ro(g,h,ns,n)
   end
   subroutine ro(g,h,ns,n)
   character tex1b*35,tex2b*35,texcr*45
   real g(0:999),h(0:999)
   integer n(2)
   gb=0.
   hb=0.
   do 50 i=0,ns
   gb=gb+g(i)
50 hb=hb+h(i)
   cr=0.
   cd1=0.
   cd2=0.
   gb=gb/real(ns+1)
   hb=hb/real(ns+1)
   do 40 i=0,ns
   cr=cr+(g(i)-gb)*(h(i)-hb)
   cd1=cd1+(g(i)-gb)*(g(i)-gb)
   cd2=cd2+(h(i)-hb)*(h(i)-hb)
40 continue
   tx1=100.*gb/real(n(1))
   tx2=100*hb/real(n(2))
   cd1=sqrt(cd1)
   cd2=sqrt(cd2)

```

```

cr=cr/(cd1*cd2)
tex1b=' nb. moyen de bien places du gr. 1 : '
tex2b=' nb moyen de bien places du gr.2 : '
texcr='coefficient de correlation de ces 2 series : '
write(2,100)tex1b,gb
write(2,101)tx1
101 format(//,15x,'Leur taux moyen :',f6.2)
write(2,100)tex2b,hb
write(2,101)tx2
write(2,110)texcr,cr
100 format(///,10x,a35,f6.2)
110 format(///,10x,a45,f8.5)
end
subroutine factor(c1,c1,k,cum,fi,gj,lam,fc,clas,n,s,nv)
real lam(0:999),cum(2,250),k(800,250),f1(800),gj(250),fc(2),nv
real k1
integer c1,cj,clas(800),n(2),s
do 60 j=1,cj
60 gj(j)=(cum(1,j)-cum(2,j))/(cum(1,j)+cum(2,j))
lam(s)=0.
do 63 j=1,cj
63 lam(s)=lam(s)+cum(1,j)*gj(j)
do 61 i=1,c1
f1(i)=0.
k1=0.
do 62 j=1,cj
k1=k1+real(k(1,j))
62 f1(i)=f1(i)+k(1,j)*gj(j)
61 fi(i)=f1(i)/k1
fc(1)=1.
fc(2)=-1.
end
subroutine sbar(c1,cj,k,cum,fi,gj,fc,jus,fau,s,n,nom,clas,
1 fics,lam,txf,justt)
integer ci,cj,s,fau(2),fc(2),clas(800),n(2),jus(2,0:999),s,c
integer justt(0:999)
character nom(800)*4,fics*8
real lam(0:999),k(800,250),cum(2,250),f1(800),gj(250),txf(2)
fau(1)=0
fau(2)=0
jus(1,s)=0
jus(2,s)=0
do 65 i=1,c1
if(fi(i)*fc(clas(i)).gt.0.) then
jus(clas(1),s)=jus(clas(1),s)+1
else
fau(clas(1))=fau(clas(1))+1
endif
65 continue
do 66 c=1,2

```

```

66   txf(c)=(real(fau(c))*100.)/real(fau(c)+jus(c,s))
      txt=real((fau(1)+fau(2))*100.)/real(c1)
      justt(s)=jus(1,s)+jus(2,s)
      write(2,112)s, lam(s),n(1),jus(1,s),txf(1),jus(2,s),txf(2),justt(s)
1   ,txt
112  format('!',i6,' ! ',f8.6,2x,'!',i6,' !',i6,' ! ',f6.2,'% !',
1   i6,1x,' ! ',f6.2,'% !',i6,' !',f7.2,' !')
      end
      subroutine histog(x,ns,titre,lh)
      real x(0:999)
      integer ix(0:999),ns,lh,it1(120)
      character titre*60,d(120)*1,b1*1,ast*1,vr*11
      character texte*20,texte1*55
      b1=' '
      ast='*'
      vr='Simulations'
      ncs=0
      xmn=x(0)
      xmx=x(n)
      do 47 i=1,ns
      if(x(i).gt.x(0)) ncs=ncs+1
      if(x(i).gt.xmx)then
      xmx=x(i)
      else if(x(i).lt.xmn)then
      xmn=x(i)
      else
      endif
47   continue
      write(2,100)titre
100  format(///20x,a60,/20x,60('-')//)
      write(2,99)ns,vr
99   format(30x,i3,1x,a11,/ ,30x,15('-')////)
      do 48 j=1,lh
48   it1(j)=0
      do 49 i=0,ns
49   ix(i)=nint((x(i)-xmn)*(real(lh-1))/(xmx-xmn))+1
      do 50 i=0,ns
50   it1(ix(i))=it1(ix(i))+1
      imx=it1(1)
      do 53 j=2,lh
53   if(it1(j).gt.imx)imx=it1(j)
      iq=0
      do 56 i=1,ix(0)
56   iq=iq+it1(i)
      texte=' Centile du cas reel'
      ic1=(iq*100)/(ns+1)
      do 54 i=imx,1,-1
      do 55 j=1,lh
      d(j)=b1
55   if(it1(j).ge.i)d(j)=ast

```

```

54  write(2,110)(d(j),j=1,1h)
110  format(130a1)
      write(2,110)('- ',j=1,1h)
      write(2,111)(b1,j=1,ix(0)-1),'^'
      write(2,111)(b1,j=1,ix(0)-1),'|'
      write(2,111)(b1,j=1,ix(0)-1),'|'
111  format(130a1)
      write(2,111)(b1,j=1,ix(0)-4),'c','a','s',' ','r','e','e','l'
      write(2,112)texte,ic1
112  format(///,a20,' : ',i3//)
      textel='Nombre precis de cas simules meilleur que le cas reel : '
      write(2,113)textel,ncs
113  format(//,a55,i3)
      titre=' '
      end
      subroutine impres(f1,g1,c1,c1,1h,ic,ir,clas,1dj,fics,nom
1  ,a,d,ig,11f,it,it1,idf,fc,leg)
      integer clas(800),11f(800),c1,c1,ig(250),it(2,1h),it1(1h)
      integer fau(2)
      real f1(800),g1(250),fc(2)
      character ic(2)*1,ir(2)*1,1dj(250)*4,a(1h)*1,b1*1,leg(2)*20
      character nom(800)*4,b14*4,d(1h,1h)*4,tr*4,idf(2,250)*4,tr2*4
      character t1x1*28,t1x2*28
      b1=' '
      t1r='- '
      b14=' '
      tr='+---'
      if(1h.le.33) then
        1h1=1h
      elseif(1h.le.66) then
        1h1=int(1h/2)
          elseif(1h.le.99) then
            1h1=int(1h/3)
              else
                1h1=int(1h/4)
            endif
      endif
c Recherche du min et du max de f1
      amin=fi(1)
      amax=fi(1)
      do 50 i=2,c1
        if(fi(i).lt.amin) then
          amin=fi(i)
        elseif(fi(i).gt.amax) then
          amax=fi(i)
        else
          endif
50  continue
      if(amax.gt.abs(amin)) then
        b1=amax
      else

```

```

      b1=abs(am1n)
      endif
      do 51 i=1,c1
51      ilf(i)=nint(((f1(i)+b1)*(lh-1)/(2*b1))
      do 53 i=1,lh
          it(1,i)=0
          it(2,i)=0
53      a(i)=b1
          do 52 i=1.c1
              j1=clas(i)
              if(ic(j1).eq.ic(1)) then
                  it(1,ilf(i)+1)=it(1,ilf(i)+1)+1
              else
                  it(2,ilf(i)+1)=it(2,ilf(i)+1)+1
              endif
52      continue
          imx1=it(1,1)
          imx2=it(2,1)
          do 58 i=2,lh
              if(it(1,i).gt.imx1)imx1=it(1,i)
58      if(it(2,i).gt.imx2)imx2=it(2,i)
          write(2,90)
90      format(//20x,'Analyse discriminante barycentrique '
1 20x,'-----' ///
210x,'Coordonnees barycentriques et entieres des individus :'/
310x,'-----',///
4 ,!' 6('-----+)',/'!'
56('Ident*coord.bar,his*')/'!'
6 6('-----+'))
          i6=c1/6
          jr=c1-6*i6
          jr1=c1-jr
          jr2=jr1-i6
          do 57 i=1,i6
57      write(2,91)(ic(clas(i+j)),nom(i+j),f1(i+j),ilf(i+j)+1,j1=0,jr2,i6)
          write(2,91)(ic(clas(jr1+j)),nom(jr1+j),f1(jr1+j),ilf(jr1+j),j1=1,jr)
91      format('!6(a1,a4,'!',f8.6,1x,'!',i3,'!'))
          fau(1)=0
          fau(2)=0
          do 40 i=1,c1
              if(fi(i)*fc(clas(i)).lt.0.) then
                  fau(clas(i))=fau(clas(i))+1
                  idf(clas(i),fau(clas(i)))=nom(i)
              else
                  endif
40      continue
          write(2,101)(idf(1,j),j=1,fau(1))
101      format(///20x,'Individus du 1er groupe mal classes'/20x,35('-')
1  ,///(4x,20(a4,''),/)
          write(2,102)(idf(2,j),j=1,fau(2))

```



```

write(2,95)lh,amin,amax
102 format(///20x,'Individus du 2eme groupe mal classes'/20x,36('-'
1 ),///(4x,20(a4,'')/))
write(2,110)(ir(1),leg(1),i=1,2)
110 format(2(5x,a1,' : ',a20/))
do 77 i=imx1,1,-1
do 78 j=1,lh
a(j)=bl
78 if(it(1,j).ge.i)a(j)=ir(1)
write(2,100)(a(j),j=1,lh)
77 continue
lhm=lh/2
write(2,100)(tir,j=1,lhm-1,'0',(tir,j=lhm+1,lh)
do 56 i=1,imx2
do 76 j=1,lh
a(j)=bl
76 if(it(2,j).ge.i)a(j)=ir(2)
56 write(2,100)(a(j),j=1,lh)
100 format(132a1)
c Calcul de l'histogramme des variables
c Recherche du min et du max de gj
aminj=gj(1)
amaxj=gj(1)
do 60 j=2,cj
if(gj(j).lt.aminj) then
aminj=gj(j)
else if(gj(j).gt.amaxj) then
amaxj=gj(j)
else
endif
60 continue
if(amaxj.gt.abs(aminj)) then
bj=amaxj
else
bj=abs(aminj)
endif
tix1=' Plus grand batonnet du gr.1'
tix2=' Plus grand batonnet du gr.2'
write(2,119)tix1,imx1,tix2,imx2
119 format(///(3x,a28,' : ',i2)/)
do 61 j=1,cj
61 ig(j)=nint((gj(j)+bj)*(lh1-1)/(2*bj))
do 62 i=1,lh
it1(i)=0
do 62 j=1,lh1
62 d(i,j)=bl4
do 63 j=1,cj
i1=1
64 if(d(i1,ig(j)+1).eq.bl4) then
it1(ig(j)+1)=it1(ig(j)+1)+1

```

```

        d(i1,ig(i)+1)=idj(i)
        else
            i1=i1+1
            go to 64
        endif
63    continue
        write(2,93)
        imxj=i1(1)
        do 68 i=2,lh
68    if(i1(1).gt.imxj)imxj=i1(1)
93    format(///10x,'Coordonnees barycentriques et entieres ',
1    ' des variables'
2    / 10x,'-----',///
3    ///,'!'6('-----+')/
4    '!'6('ident! cord.bar!his.!' )/
5    '!'6('-----+')))
        i6=cj/6
        jr=cj-6*i6
        jr1=cj-jr
        jr2=jr1-i6
        do 65 i=1,i6
65    write(2,94)(idj(i+j),gj(i+j),ig(i+j)+1,i=0,jr2,i6)
        write(2,94)(idj(jr1+j),gj(jr1+j),ig(jr1+j),j=1,jr)
94    format('!'6(a4,'!',f8.3,'!',i3,'!'))
        write(2,95)lh1,aminj,amaxj
95    format(///' L'histogramme suivant a ',i3,' classes min=',
1    f10.6,10x,' max=',f10.6///)
        do 66 i=1:imxj,1,-1
66    write(2,96)(d(i1,i),j=1,lh1)
96    format(100a4)
        lhm=lh1/2
        tr2='0---'
        write(2,97)(tr,j=1,lhm),tr2,(tr,j=lhm+1,lh1-1)
97    format(50a4)
        end
        subroutine randu(ix,iy,yfl)
            iy=ix*65539
            if(iy)5,6,6
5            iy=iy+2147483647+1
6            yfl=iy
            yfl=yfl*.4656613 e-9
        end
        subroutine graph(a,b,ld,v1,v2,ns)
            dimension a(0:999),b(0:999)
            character id(0:999)*4,ir(0:60,0:120)*1,nm(0:60)*4,v1*4,v2*4,tr*1
            real mina,minb,maxa,maxb
            character nj(0:120)*4,tx1*28,tx2*32
            tr='_'
            mina=a(0)
            minb=b(0)

```

```

maxa=a(0)
maxb=b(0)
do 50 i=1,ns
  if(mina.gt.a(i))  mina=a(i)
  if(minb.gt.b(i))  minb=b(i)
  if(maxa.lt.a(i))  maxa=a(i)
  if(maxb.lt.b(i))  maxb=b(i)
50  continue
do 51 i=0,60
  nm(i)=' '
do 51 j=0,120
51  ir(i,j)=' '
do 61 i=0,120
  nj(i)=' '
61  continue
  xa=maxa-mina
  xb=maxb-minb
do 52 i=1,ns
  ip=nint(60*(a(i)-mina)/xa)+1
  jp=nint(120*(b(i)-minb)/xb)+1
  if(ir(ip,jp).eq.' ') then
    ir(ip,jp)='*'
    nj(jp)=id(i)
  else if(ir(ip,jp).eq.**') then
    ir(ip,jp)='1'
    go to 34
  else if(ir(ip,jp).eq.'1') then
    ir(ip,jp)='2'
    go to 34
    else if(ir(ip,jp).eq.'2') then
      ir(ip,jp)='3'
      go to 34
      else if(ir(ip,jp).eq.'3') then
        ir(ip,jp)='4'
        go to 34
      else
        ir(ip,jp)='5'
      endif
34  nm(ip)=id(i)
52  continue
  tex1='Representation graphique du'
  tex2='Taux de bien classes 1=f(taux2)'
  write(2,123)tex1,tex2
123  format(20x,a28/20x,a32////)
  write(2,140)v1
140  format(1h,6x,a4)
do 53 i=60,0,-1
80  write(2,150)nm(i),(ir(i,j),j=0,120)
150  format(2x,a4,'|',121a1)
  53  continue
  write(2,145)(tr,k=0,110),v2
145  format(6x,111a1,a4,'_____')
  write(2,133)(nj(i)(1:1),i=0,120)
  write(2,133)(nj(i)(2:2),i=0,120)
  write(2,133)(nj(i)(3:3),i=0,120)
  write(2,133)(nj(i)(4:4),i=0,120)
133  format(7x,121a1)
  return
end

```