

## ENQUETE PLURIDISCIPLINAIRE AUPRES DES PECHEURS DU DELTA CENTRAL DU NIGER: PLAN DE SONDAGE ET ESTIMATEURS ASSOCIES.

---

MORAND P., QUENSIERE J., HERRY C.

### RESUME

La réalisation d'une enquête par échantillonnage dans une société de pêcheurs Sahéliens posait des problèmes méthodologiques liés d'une part à l'ampleur et à la complexité de l'univers étudié, d'autre part à l'absence de véritable base de sondage au départ, et enfin à la poursuite d'objectifs multiples, pluridisciplinaires.

C'est pourquoi un plan de sondage assez complexe dû être adopté, combinant diverses méthodes d'échantillonnage.

Conséquence de cette complexité des difficultés statistiques surgissent, notamment lorsqu'il s'agit d'évaluer des variances ou précisions. Pour y faire face, on a utilisé des approches classiques (estimateurs H.T.) ou "modernes"(Monte-Carlo). Quelques résultats numériques inédits sont présentés à cette occasion.

### INTRODUCTION

Le progrès dans la compréhension de la structure, du fonctionnement et, à terme, de l'évolution des systèmes pêches (cf. contribution de J. Quensière) passe nécessairement, à certains moments, par l'acquisition de nouvelles données, c'est-à-dire par des études descriptives. Dans ce cadre, un mot s'impose de prime abord dès lors qu'il s'agit du domaine humain: enquête. Mais il recouvre en fait des pratiques différentes, adaptées aux besoins et à la démarche de chaque discipline. La forme décrite ici est bien précise: il s'agit d'une enquête statistique par échantillonnage, faisant appel aux techniques de sondage **aléatoire** (et non pas "raisonné"), lesquelles ont pour principal mérite de permettre une inférence explicite des paramètres de la fraction observée (l'échantillon) à "l'univers" étudié. Après une présentation synoptique du plan de sondage, cet article présente quelques estimations possibles sur la base des données récoltées, ce qui autorise au passage une évaluation de la

qualité de l'enquête, tant du point de vue de sa tractabilité statistique que de sa précision.

## I L'ESTIMATION "INFÉRENTIELLE", UN OBJECTIF PARMIS D'AUTRES DANS UNE ENQUÊTE PLURIDISCIPLINAIRE

Au moment de l'élaboration du plan, les objectifs visés deviennent les premières "contraintes" à prendre en compte. Dans cette enquête préliminaire du programme Delta Central du Niger (Quensière, 1988), différentes disciplines étaient impliquées dès le départ et participaient à la définition des objectifs. De l'examen de ceux-ci ressortaient plusieurs axes déterminants pour la suite des opérations:

Tout d'abord, les questions posées concernaient non pas un seul mais plusieurs niveaux (plus ou moins "emboîtés") d'éléments de l'univers Delta: les individus, les familles ("marmites" ou concessions), les agglomérations, ces niveaux constituant autant de populations statistiques échantillonnables.

Concernant ces différents objets, on pouvait dégager deux directions majeures dans la recherche des connaissances:

(1) Principalement, l'enquête devait recueillir une grande variété de renseignements à-propos de chacun de ces types d'éléments. Car, dépassant une simple ambition descriptive, l'enquête se devait d'aboutir à des modèles explicatifs, lesquels ne peuvent naître que de multiples possibilités de croisements entre de nombreuses variables.

(2) Secondairement, il était souhaitable que l'enquête fournisse des estimations régionales de nombre, de quantités ou de proportion, notamment en ce qui concerne les agents de production (individus actifs, ménages de pêcheurs...), car de tels chiffres ne sont pas disponibles sous la forme souhaitable dans les recensements administratifs.

La poursuite conjointe de ces deux directions d'objectifs (relevant respectivement de l'optique "analyse des données" et de la statistique inférentielle) n'allait pas sans difficultés, l'une (1) impliquant une forte taille individuelle des questionnaires, l'autre (2) nécessitant non seulement le recours aux techniques de sondage aléatoire mais aussi une taille d'échantillon respectable pour pouvoir espérer une précision suffisante.

Si les résultats de l'enquête ont fait l'objet de nombreuses publications dans le cadre de la première direction, il n'y a eu, à ce jour, que très peu d'exploitation dans le sens de la seconde.

C'est donc à la préparation de cette tâche que nous nous employons ici, en complétant les outils nécessaires aux inférences (déjà

fournies pour partie dans Morand, 1988) par des évaluations de précision.

## II LE PLAN DE SONDAGE: DES CONTRAINTES AUX GRAND~ CHOIX

Mis à part les objectifs de l'enquête, évoqués plus haut, quatre grands types de contraintes ont conditionné la conception du plan:

(1) L'absence de véritable base de sondage au départ, c'est-à-dire de listes exhaustives des éléments les plus fins sur lesquelles on aurait pu effectuer, avant toute phase de terrain, des procédures de tirages aléatoires.

(2) Une structure hiérarchique assez évidente des différents types d'éléments au sein du système étudié, depuis l'individu jusqu'à la concession et à l'agglomération, au-dessus de laquelle on s'accordera à définir la "zone géographique", puis la "sousrégion" (termes provisoires).

(3) Toutefois, une grande hétérogénéité au sein des populations correspondant à un même niveau d'objets: il y a une grande différence entre une agglomération constituée de trois cases de paille et une véritable ville comme Djenné, entre une sous-région dunaire traversée de quelques marigots et une autre s'étendant le long du fleuve Niger, englobant ses bras et ses plaines inondables.

(4) Enfin, des contraintes de terrain sévères: dispersion des éléments, distances à parcourir, absence de véritable réseau de communication.

Le premier point (1) rendait nécessaire la construction d'un plan d'échantillonnage à plusieurs degrés.

Le second (2) fournissait un support naturel pour cette construction.

Le troisième (3) a suggéré l'adoption, à divers niveaux, de l'échantillonnage par strates, permettant une modulation de l'effort ou des méthodes d'enquête selon les portions du système.

Enfin, le quatrième point a limité la taille de l'échantillon, compte-tenu d'un budget donné.

Finalement, le plan adopté reposait en premier lieu sur un découpage du Delta en deux sous-ensembles: d'une part la macrostrate "villes", regroupant les 34 agglomérations de plus de 2000 habitants (recensement 76, réactualisé 86) à l'exception de la ville même de Mopti (où n'habitent que très peu de pêcheurs), et d'autre part la macrostrate "rurale", comprenant le reste du Delta.

Il était prévu dès le départ d'enquêter deux fois (à l'étiage et en hautes eaux 1987) cette macrostrate rurale, de façon à disposer d'observations factuelles, et non pas seulement déclaratives, sur les changements saisonniers. Quant au sous-ensemble "villes", il ne devait être visité qu'une seule fois, à l'étiage, l'année de réalisation finalement retenue (pour des raisons contingentes) étant 1989.

Pour les buts visés ici, qui sont d'estimer des nombres ou des quantités à l'échelle régionale, il est préférable de se limiter aux données des deux phases réalisées à la même saison, même s'il s'agit d'années différentes: enquête rurale étiage 87 et enquête "villes" étiage 89. En effet le cycle saisonnier des déplacements de pêcheurs est un phénomène de telle ampleur que l'on ne peut pas réunir l'échantillon de la phase rurale "hautes eaux 87" avec celui observé en étiage dans la même macrostrate. Par contre, on considérera que ce cycle est suffisamment reproductible d'une part, et que la dérive générale interannuelle de 87 à 89 est suffisamment faible d'autre part, pour pouvoir traiter les deux phases d'étiage (rurale 87 et villes 89) comme si elles avaient été réalisées simultanément, c'est-à-dire en additionnant simplement leurs résultats.

### III DE LA STRUCTURE DU PLAN SONDAGE AUX ESTIMATEURS STATISTIQUES

Dans la macrostrate "villes", on a effectué un tirage aléatoire simple de  $a (=12)$  agglomérations dans la liste exhaustive qui en totalise  $A (=34)$ . A l'intérieur de chacune de ces 12 agglomérations, une liste exhaustive des  $C$  concessions reconnues comme ayant des membres pratiquant peu ou prou la pêche a été établie auprès des "vieux" du village, et un tirage aléatoire systématique (sans remise) de  $c$  d'entre-elles a été effectué. Le taux d'échantillonnage ( $c/C$ ), non constant, varie de 20% à 85%; il est d'autant moins élevé que  $C$  est grand. Toutes les concessions tirées ont été visitées et ont été décrites de façon exhaustive, tant au niveau de leurs ménages que de leurs individus.

Il s'agit donc d'une forme banale de plan à 2 degrés, la seule sophistication résidant dans l'emploi du tirage systématique (et non pas aléatoire simple) au niveau ultime.

Cependant, nous considérerons par la suite, comme beaucoup d'auteurs, que l'échantillonnage systématique est simplement une forme de tirage aléatoire sans remise, dans la mesure où la liste des concessions ne se présente pas sous un quelconque ordre corrélé aux variables étudiées - mise à-part la première concession citée qui est toujours plutôt grande, puisqu'il s'agit très généralement de celle du chef de village. Au pire, les variances d'estimation seront un peu surévaluées, d'où une légère sous-estimation de la précision de l'enquête (Anonyme, 1974).

Le caractère très vaste de la macrostrate rurale du Delta a conduit à l'utilisation d'un plan plus complexe que celui adopté pour la macrostrate "villes". Il consiste tout d'abord dans la division du Delta en 9 strates géographiques, de surfaces variées, correspondant à des sous-ensembles régionaux reconnus par les études sociologiques préliminaires et présentant une certaine identité en termes de géographie physique.

Chacune de ces strates a été divisée en un certain nombre  $Z$  de petites zones géographiques (8 à 67), d'importances relativement homogènes (du moins à l'intérieur d'une même strate) tant du point de vue du nombre de villages englobés que de la surface. Un tirage aléatoire simple sans remise a été effectué dans chaque strate pour définir l'échantillon des  $z$  zones à visiter, avec un taux ( $z/Z$ ) variant de 7 à 47% selon l'hétérogénéité et l'intérêt supposés de la strate. Dans chaque zone ainsi désignée, on a visité toutes les  $A$  agglomérations, en établissant dans chacune la liste des  $C$  concessions reconnues comme ayant des membres pratiquant peu ou prou la pêche. Tout comme dans l'enquête "villes", un tirage systématique dans cette liste a fourni l'échantillon des  $c$  concessions qui allaient être visitées - en employant un taux ( $c/C$ ) décroissant en fonction de  $C$ , de telle façon que l'on puisse compter sur une précision par agglomération approximativement constante (Herman, 1986, p. 82). Ainsi, pour avoir un intervalle de confiance (risque admis: 0,05) de 15% autour d'une proportion type de 50%, il faut échantillonner 10 concessions lorsqu'il en y en a 177 ou plus, 9 quand il y en a 57 à 76, 8 quand il y en a 30 à 56 (etc . . . ) et évidemment une quand il n'y en a qu'une !

Il s'agit donc d'un cas particulier de plan à 4 degrés, comportant 2 recensements (pour le 1er et le 3e degré) et 2 tirages aléatoires, respectivement simple et systématique, pour les 2e et 4e degrés.

A un plan de sondage donné correspond un ensemble de calculs qui permettent de dégager les statistiques souhaitées pour décrire les populations d'éléments. Ces statistiques estimées (ou "estimateurs") peuvent être de plusieurs nature: dans le cas le plus fréquent, la variable est un descripteur quantitatif des éléments ultimes échantillonnés (ici, ce sont les concessions), ce qui conduit à rechercher un estimateur "quantité totale" et un estimateur "moyenne par élément" aux niveaux supérieurs (ici: strate, macrostrate ou Delta). Outre la valeur même de ces estimateurs, on est intéressé par leur précision, ce qui nécessite l'estimation de leur variance. On fournira ici les estimateurs "nombre total", "quantité totale", "ratio (rapport)" et "moyenne", ainsi que l'estimateur de variance du second et du troisième, pour chacun des deux plans. Les estimateurs proposés étant sans biais ou presque (ce sont ceux de Horvitz-Thompson, 1952), les variances fournies conduisent directement à la définition d'intervalles de confiance correspondant à l'erreur d'échantillonnage. A noter que l'erreur d'observation, très difficilement évaluable, n'est pas prise en compte ici.

### l) Cas de la macrostrate "villes": plan à 2 degrés

A : nombre total d' agglomérations (= 34)

a : nombre d' agglomérations échantillonnées (=12)

$C_i$  : nombre total de concessions de pêcheurs de l'agglomération i.

$c_i$  : nombre de concessions échantillonnées dans l'agglomération i.

$Y_{ii,j}$  : quantité y pour la concession j de l'agglomération i.

$n_i$  : nombre d'éléments (c.à.d. de concessions) d'une certaine catégorie observés dans l'échantillon de l'agglomération i.

la - Estimation d'un nombre total

(nombre de concessions d'une certaine catégorie)

$$\hat{N} = \frac{A}{a} \sum_{i=1}^a \frac{c_i}{C_i} n_i$$

lb - Estimation d'une "quantité totale"

(ex.: nombre d'individus, nombre de ménages)

Quantité estimée dans une agglomération:

$$\hat{Y}_i = \frac{C_i}{c_i} \left( \sum_{j=1}^{c_i} Y_{ij} \right)$$

Quantité estimée dans la strate ville:

$$\hat{Y} = \frac{A}{a} \left( \sum_{i=1}^a \hat{Y}_i \right)$$

lb' - Estimation de la variance  $\sigma^2 \hat{Y}$  de l' estimateur quantité totale  
(Gourieroux, 1981)

$$\hat{\sigma}^2_{\hat{Y}} = \frac{A(A-a)}{a} \frac{1}{a-1} \sum_{i=1}^a (\hat{Y}_i - \frac{\hat{Y}}{A})^2 + \frac{A}{a} \sum_{i=1}^a \frac{C_i(C_i-c_i)}{c_i} \frac{1}{c_i-1} \sum_{j=1}^{c_i} (Y_{ij} - \frac{\hat{Y}_i}{c_i})^2$$

**Remarque:** A chaque ligne apparait un facteur de type  $\frac{(N(N-n))}{N}$ , où N est le nombre total d'éléments et n le nombre d'éléments échantillonnés. Dans la littérature, on trouvera souvent ce terme présenté d'une façon un peu différente [  $\frac{N^2(1 - \frac{n}{N})}{n}$  ] (par ex. dans Scherrer - 1983), ce qui a l'avantage de mieux mettre en évidence la présence de la "correction pour population finie":  $(1 - \frac{n}{N})$ .

lc - Estimation d'une variable ratio (ou rapport, ou quotient).(ex.: nombre d'individus par ménage, sex-ratio, effort de pêche par homme actif)

Lorsque l'on veut estimer le rapport moyen entre deux variables quantitatives y et y' observées sur chaque élément, on doit employer l'estimateur ratio. [A noter que la plupart des proportions, ainsi que certaines moyennes, se ramènent à des ratios.]

$$\hat{R} = \frac{\hat{Y}}{\hat{Y}'}$$

où  $\hat{Y}$  et  $\hat{Y}'$  sont les deux estimateurs de quantité. (cf. 1b)

Cette estimateur est légèrement biaisé, mais le biais est négligeable si l'effectif de l'échantillon est grand (cas présent).

lc'- Estimation de la variance  $\hat{\sigma}^2_{\hat{R}}$  de l'estimateur ratio

Les formules d'estimation sont plus complexes que dans le cas d'une quantité simple. Cochran (1977) décrit un estimateur approché, que nous ré-écrivons ici en adoptant le type de présentation de Gourieroux (1981):

$$\hat{\sigma}_{R}^{2\wedge} = \frac{1}{\hat{Y}^2} \frac{A(A-a)}{a} \frac{1}{a-1} \sum_{i=1}^a (\hat{Y}_i - \hat{R} \hat{Y}_i)^2$$

$$+ \frac{A}{a} \frac{1}{\hat{Y}^2} \sum_{i=1}^a \frac{C_i (C_i - c_i)}{c_i} \frac{1}{c_i - 1}$$

$$\sum_{j=1}^{c_i} [(\hat{Y}_{ij} - \hat{R} \hat{Y}_{ij}) - (\hat{Y}_i - \hat{R} \hat{Y}_i)]^2$$

Id - Estimation d'une quantité moyenne par élément (ex.: nombre d'individus par concession)

Si l'on disposait d'une base de sondage complète, c'est-à-dire de listes exhaustives des populations d'éléments (et ce jusqu'au niveau le plus fin: la concession), on pourrait écrire de véritables estimateurs de moyenne en divisant simplement les estimateurs de quantité par les nombres totaux N d'éléments correspondant:

$$\hat{\bar{Y}} = \frac{\hat{Y}}{N}$$

Et l'on aurait comme variance pour cet estimateur

$$\hat{\sigma}_{\bar{Y}}^2 = \frac{\hat{\sigma}_Y^2}{N^2}$$

Malheureusement, et c'est d'ailleurs là une raison majeure du recours à l'échantillonnage à multiples degrés, on ne dispose ni d'une base de sondage ni d'une connaissance exacte de ce nombre N, mais seulement d'une estimation N issue de l'enquête.

On estimera donc la moyenne comme s'il s'agissait d'un rapport, avec un estimateur de quantité au numérateur et un estimateur de nombre total au dénominateur:

$$\hat{\bar{Y}} = \frac{\hat{Y}}{\hat{N}}$$



**2) Cas de la macrostrate "rurale" : plan à 4 degrés**

On conviendra tout d'abord des notations suivantes:

indices: i: n° de strate;

j: n° de zones;

k: n° d'agglomération;

m: n° de concession;

S : nombre total de strates (=9)

$Z_i$  : nombre total de zones de la strate i

$z_i$  : nombre de zones échantillonnées dans la strate i

$A_{i,j}$  : nombre total d'agglomérations de la zone j de la strate i

$a_{i,j}$  : nombre d'agglomérations échantillonnées de la zone j de la strate i (ici, on aura toujours:  $a_{ij} = A_{ij}$ )

$C_{i,j,k}$  : nombre total de concessions de l' agglomérations k de la zone j de la strate i.

$c_{i,j,k}$  : nombre de concessions échantillonnées dans l'agglomération k de la zone j de la strate i.

$Y_{i,j,k,m}$  : quantité y pour la concession m de l'agglomération k de la zone j de la strate i.

$n_{i,j,k}$  : nombre d'éléments (c.à.d. de concessions) d'une certaine catégorie, observés dans l'agglom. k de la zone j de la strate i.

2a - Estimation d'un nombre total

$$\hat{N} = \sum_{i=1}^S \frac{Z_i}{z_i} \sum_{j=1}^{z_j} \frac{A_j}{a_j} \sum_{k=1}^{a_j} \frac{C_k}{c_k} n_{i,j,k}$$

2b - Estimation d'une quantité totale Y.

Quantité estimée dans une agglomération :

$$\hat{Y}_{ijk} = \frac{C_{ijk}}{c_{ijk}} \left( \sum_{m=1}^c Y_{ijkm} \right)$$

Quantité estimée dans une zone j :

$$\hat{Y}_{ij} = \sum_{k=1}^a Y_{ijk}$$

Quantité estimée dans une strate i :

$$\hat{Y}_i = \frac{Z_i}{z_i} \left( \sum_{j=1}^z \hat{Y}_{ij} \right)$$

Quantité estimée dans la macrostrate:

$$\hat{Y} = \sum_{i=1}^S \hat{Y}_i$$

2b' - Estimation de la variance de l'estimateur quantité totale

(Gourieroux, 1983)

$$\begin{aligned} \sigma^2 \hat{Y} &= \sum_{i=1}^S \frac{Z_i(Z_i - z_i)}{z_i} \frac{1}{z_i - 1} \sum_{j=1}^z (\hat{Y}_{ij} - \frac{\hat{Y}_i}{Z_i})^2 \\ &+ \sum_{i=1}^S \sum_{j=1}^z \frac{z_i}{z_i} \frac{A_{ij}(A_{ij} - a_{ij})}{a_{ij}} \frac{1}{a_{ij} - 1} \sum_{k=1}^a (\hat{Y}_{ijk} - \frac{\hat{Y}_{ij}}{Z_{ij}})^2 \\ &+ \sum_{i=1}^S \sum_{j=1}^z \sum_{k=1}^a \frac{z_i A_{ij}}{z_i a_{ij}} \frac{C_{ijk}(C_{ijk} - c_{ijk})}{c_{ijk}} \frac{1}{c_{ijk} - 1} \sum_{m=1}^c (Y_{ijkm} - \frac{\hat{Y}_{ijk}}{C_{ijk}})^2 \end{aligned}$$

La seconde ligne étant forcément nulle (puisque  $a_{i,j} = A_{i,j}$ ), le calcul de cette variance se limite à l'addition de deux termes: l'un (première ligne ci-dessus) résulte de l'échantillonnage et de la variabilité des zones dans chaque strate, le second (troisième ligne) provient essentiellement de l'échantillonnage et de la variabilité des concessions dans les agglomérations visitées (bien qu'il contienne aussi un facteur de dilatation égal à l'inverse du taux d'échantillonnage des zones correspondantes).

2c - Estimation d'un quotient ou ratio (cf. lc)

2c' - Estimation de la variance de l'estimateur ratio.

En partant de la variance de l'estimateur quantité, on obtient la variance de l'estimateur ratio de la même façon que pour la macrostrate "villes" (cf. & lc'), c'est-à-dire en appliquant un facteur général  $\frac{1}{Y^2}$  d'une part et d'autre part en remplaçant, lors des calculs des moments à chacun des niveaux, les valeurs

$$Y_{\text{indices}} \text{ par } (Y_{\text{indices}} - \hat{R} Y_{\text{indices}})$$

2d - Estimation d'une moyenne par élément

Concernant cet estimateur, les mêmes considérations que celles faites à propos de la macrostrate "villes" s'impose (cf. ld).

### 3) Combinaison des estimateurs de quantité des deux macrostrates

Pour estimer une quantité totale sur l'ensemble du Delta, on se contente de sommer les deux estimateurs de quantité obtenus sur chacune des macrostrates. De même, les variances issues des deux macrostrates seront simplement additionnées puisqu'elles sont indépendantes.

## IV QUELQUES ESTIMATIONS NUMERIQUES ET LEUR PRECISION

### 1) Macrostrate "villes"

Estimations de quantités (Y), avec leur variance d'estimation

( $\hat{\sigma}^2$ ) et leur précision ( $1,96\frac{\hat{\sigma}}{\hat{Y}}$ ), cette dernière étant définies sur la base d'un intervalle de confiance représentant 95% de la distribution (définition conservée tout au long de ce travail):

Tableau I

variable	estimat $\hat{Y}$	terme 2 (eq.lb')	terme 1 (eq.lb')	$\hat{\sigma}^2$	$\hat{\sigma}$ $(1,96\frac{\hat{\sigma}}{\hat{Y}})$
Population pech. (nb total indiv.)	33663	434868	48,8 10 <sup>6</sup>	49,25 10 <sup>6</sup>	7017 (40%)
Hommes 'actifs' (12 < âge < 65)	8829	39138	3,47 10 <sup>6</sup>	3,51 10 <sup>6</sup>	1874 (41%)
Population Bozo	18918	305669	39,3 10 <sup>6</sup>	39,6 10 <sup>6</sup>	6291 (65%)
nb de ménages	3612	1293	510694	511987	715 (39%)

Estimations de ratios:

ex.: le nombre d'individus par ménage.

Soit  $\hat{Y}$  l'estimation du nombre d'individus et  $\hat{Y}'$  celle du nombre de ménages:

$$\text{On a: } \hat{Y} = \frac{\hat{Y}}{\hat{Y}'} = \frac{33663}{3612} = 9,32$$

En appliquant l'estimateur de variance décrit plus haut (lc'), on obtient:

$$\hat{\sigma}^2 R = 0,1405 \quad , \text{ d'où une précision de } (1,96\frac{\hat{\sigma}}{\hat{Y}}) = 7,8\%$$

Il apparaît que l'estimation du ratio, variable de structure, est beaucoup plus précise que celles des deux quantités qui le constituent. On retrouve ici une constatation générale en matière de résultats de sondage.

## 2) Macrostrate rurale

Estimations de quantités (Y), avec leur variance d'estimation ( $\hat{\sigma}^2$ ) et leur précision ( $1,96 \frac{\hat{\sigma}}{\hat{Y}}$ ):

**Tableau II**

variable	estimat $\hat{Y}$	terme 2 (eq.2b) :	terme 1 (eq.2b)	$\hat{\sigma}^2$	$\hat{\sigma}$ $(1,966 \frac{\hat{\sigma}}{\hat{Y}})$
Population pech.	191271	$4,2 \cdot 10^6$	$273 \cdot 10^6$	$277,2 \cdot 10^6$	16650 (17%)
Hommes 'actifs' (12 < âge < 65)	52905	468046	$8,5 \cdot 10^6$	$18,9 \cdot 10^6$	4355 (16%)
Population Bozo	104640	$2,2 \cdot 10^6$	$131 \cdot 10^6$	$133 \cdot 10^6$	11531 (21%)
nb de ménages	26392	41446	$6,74 \cdot 10^6$	$6,78 \cdot 10^6$	604 (19%)

## 3) Delta global

A partir des résultats sur les deux macrostrates (Tab.I et II), on obtient, par sommation, des chiffres pour le Delta global (Mopti exclu). La population totale de pêcheurs est ainsi estimée à  $(191271 + 363603 =) 224874$  individus, avec une variance de  $(277,2 \cdot 10^6 + 49,2 \cdot 10^6 =) 326,4 \cdot 10^6$ , ce qui définit une précision de 15,7%.

De même, on peut ainsi estimer:

- le nombre de Bozos à 123558 (précision: 20,8%)
- le nombre de ménages (en étiage) à 30004 (précision: 17,6%)
- le nombre d'hommes actifs à 61734 (précision: 15,1 %)

A noter, pour comparaison avec ce dernier résultat, le chiffre de 54112 pêcheurs fourni par Konare (1977), sur la base d'une opération de recensement de la pêche menée en 1973-74.

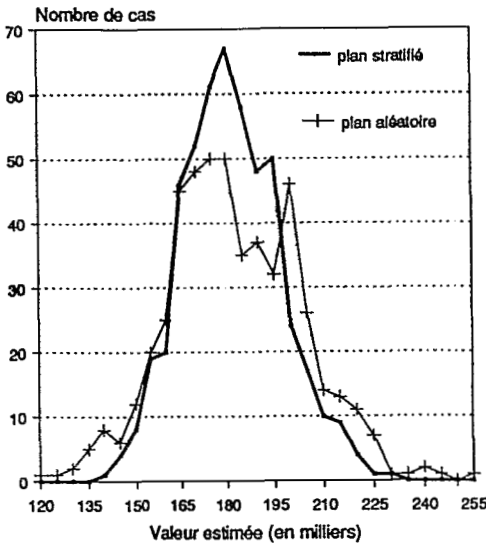
Des chiffres plus nombreux seront fournis dans une publication ultérieure (Herry et Morand, en préparation).

## V SIMULATION DE LA PRINCIPALE SOURCE DE VARIANCE. EVALUATION A *POSTERIORI* DE L'APPORT DE LA STRATIFICATION

En observant les résultats intermédiaires des calculs de précision, on peut voir apparaître la principale source de variance pour l'estimation des quantités totales sur le Delta: elle se situe sur le premier terme de la variance d'estimation de la macrostrate rurale (cf. 2b'), lorsque, à partir des estimations des différentes quantités par zone (lesquelles sont généralement précises), on passe à l'estimation des quantités par strate. [Remarque: le passage au niveau macrostrate rurale n'est ensuite qu'une simple somme et ne "produit" donc pas d'imprécision spécifique.] Deux facteurs peuvent jouer ici: un nombre trop faible de zones échantillonnées par strate et/ou une variabilité inter-zones intrastrates élevée.

Ce dernier point pouvait laisser croire que l'utilisation d'une stratification géographique n'a pas eu les effets bénéfiques escomptés. Nous avons voulu approfondir cette question, c'est-à-dire évaluer l'apport de la stratification à la précision en ce qui concerne la macrostrate rurale - qui, par son poids dominant, conditionne largement la précision des résultats globaux. On a procédé par simulation. Le principe de cette méthode est de disposer d'un univers fictif qui, tout en présentant des propriétés d'hétérogénéité semblables à celles de l'univers réel, puisse subir un nombre illimité d'échantillonnages, en supportant l'utilisation de différentes sortes de plan. Comme il s'agit ici d'étudier le seul effet de la stratification pour ce qui concerne l'échantillonnage des zones, on se contentera de re-crée une population de zones et non pas un univers complet jusqu'au niveau individu. La façon la plus simple de construire cette population est de répliquer l'échantillon des zones, strate par strate, selon un nombre de fois inverse du taux d'échantillonnage, de la même façon que dans une procédure de "Bootstrap" (Deville, 1987). On obtient ainsi une population de zones ayant des variances interstrate et intra-strate quasi conformes à celles de l'échantillon observé par l'enquête. [La seule inexactitude est liée au fait que l'inverse du taux pratiqué dans l'enquête n'est pas toujours un entier, ce qui oblige à un arrondi pour obtenir le nombre de réplifications à effectuer. C'est aussi ce qui explique que la macrostrate rurale ainsi reconstruite ne comporte que 242 zones au lieu de 247 - et ne totalise que 183712 habitants concernés par la pêche - et non pas 191271 comme l'a estimé l'enquête.]

Nous avons donc effectué deux séries de 500 simulations d'échantillonnages, l'une selon la méthode du tirage aléatoire simple sans remise, l'autre selon le plan stratifié analogue à celui employé lors de l'enquête. Le résultat calculé est l'estimation du nombre total d'individus. Voici la distribution de ce résultat dans chacune des deux séries:



plan aléatoire simple sans remise:

$$\hat{Y} = 184232$$

$$\hat{\sigma}^2 Y = 436,6 \cdot 10^6$$

(précis: 22,2%)

plan stratifié:

$$\hat{Y} = 183446$$

$$\hat{\sigma}^2 Y = 233,4 \cdot 10^6$$

(16,3 %)

(on notera que cette variance empirique est bien du même ordre que celle estimée analytiquement à partir de l'enquête: cf. terme 1 Tab II)

On conclura sans hésiter à l'utilité de la stratification adoptée lors de l'enquête de la macrostrate rurale.

## CONCLUSIONS

Tout d'abord, on a montré que cette enquête pluridisciplinaire autorise des estimations statistiques de quantités, de ratios et de nombres totaux. Quant aux précisions obtenues, s'il faut reconnaître qu'elles sont tout juste convenables pour les quantités (de l'ordre de 15% dans les meilleurs cas), elles semblent nettement meilleures si l'on s'intéresse aux structures, notamment aux ratios. Cette constatation doit à l'avenir inciter à une utilisation combinée des résultats de l'enquête (fournissant de bons ratios intercatégoriels) et des statistiques administratives (plus précises, car il s'agit de recensements, mais très globales), ceci dans le but d'obtenir de bons chiffres concernant les "nombres de ménages par catégorie". Rappelons que de telles valeurs constituent la base indispensable

pour l'extrapolation des résultats de l'enquête halieutique en cours (Laé et Bousquet, 1990) -

Du point de vue des méthodes, il faut reconnaître que ces résultats n'ont pu être obtenus que par le recours à un plan de sondage sophistiqué, lequel a conduit à des formules d'estimateurs parfois très lourdes, prohibant dans certains cas l'expression des variances. Certes, la complexité du plan était justifiée: la stratification s'avère efficace, l'échantillonnage par zones géographiques, c.à.d. par grappes d'agglomérations, était incontournable (eu égard aux contraintes logistiques), l'échantillonnage à taux variable des concessions dans les agglomérations de même. Mais il nous semble cependant que l'on a approché les limites de ce qu'il est raisonnable de concevoir en matière de plan de sondage. Ceci est d'autant plus vrai que les conséquences de cette complexité ne se limitent pas aux difficultés d'inférence: les développements selon l'approche dite de l'analyse des données (analyses factorielles, classifications) doivent aussi être menées avec une attention très particulière pour ne pas conduire à des résultats biaisés.

On a donc quelques raisons de conclure sur la difficulté qu'il y a à concilier des objectifs différents dans une même enquête. On peut se demander si, à l'avenir, il ne serait pas préférable de séparer l'effort d'obtention des connaissances quantitatives "absolues" (nombre totaux, quantités totales) de celui visant la description fine des structures ou la compréhension des phénomènes. Une stratégie judicieuse pourrait ainsi consister, après une phase exploratoire d'identification des objets et catégories pertinentes, à procéder en deux temps:

-Effectuer tout d'abord un sondage aléatoire à fort taux d'échantillonnage (mais petit questionnaire), conçu pour une tractabilité statistique inférentielle parfaite, et visant uniquement des estimations précises de nombre ou quantités par catégorie ou type d'objet.

-Puis développer, au fur et à mesure des besoins, des enquêtes "sur mesure" répondant aux différentes problématiques (pluri)disciplinaires. En précisant que, dans ce cas, l'échantillonnage représentatif raisonné (*par quotas* par exemple, en se basant sur les estimations issues du sondage aléatoire) sera souvent beaucoup plus pratique, tout en satisfaisant aux exigences d'une "analyse des données" multivariée.

## REMERCIEMENTS

Il nous est agréable de remercier ici Melle Habibatou NDiaye, qui nous a été d'un grand secours pour le traitement informatique des données.



## REFERENCES BIBLIOGRAPHIQUES

- ANONYME (1974): 'Manuel des enquêtes démographiques par sondage en Afrique'. Nations Unies. 263 pp.
- COCHRAN W.G. (1977): 'Sampling techniques'. John Wiley ~ Sons . 428 PP -
- DEVILLE J.C. (1987): 'Replications d'échantillons: demiéchantillons, jackknife et bootstrap'. in 'Les sondages', Dreesbeke, Fichet et Tassi eds . Economica . pp. 147-171.
- GOURIEROUX C . (1981): ' Théorie des sondages' . coll . "Economie et Statistiques Avancées". Economica. 270 pp.
- HERMAN J. (1986): 'Analyse de données qualitatives. 1. Traitement d'enquêtes, échantillon, répartitions, associations.' Masson. coll. 'Méthodes + Programmes'. 183 pp.
- KONARE A. (1977): 'Collecte, traitement et commercialisation du poisson en plaines inondables', in CIFA Working Party on River and Floodplain Fisheries, contributions by members of the Working Party, pp. 32-45
- LAE R. et BOUSQUET F. (1990): 'Définition d'une procédure d'échantillonnage des pêches artisanales dans le Delta Central du Niger'. Séminaire et Rapport d'étape INRZFH-ORSTOM: Nov. 90. A paraître.
- MORAND P. (1988): 'Méthodologie de traitement des données d'enquêtes'. in 'INRZFH-ORSTOM: Etudes halieutiques du Delta Central du Niger: Enquête Statistique auprès des Pêcheurs. Premiers résultats . 'pp. 3-1 à 3-10 .
- QUENSIERE J. (1988) : 'Etudes halieutiques du Delta Central du Niger. Enquête statistique auprès des pêcheurs. Premiers résultats: Introduction'. doc. INRZFH-ORSTOM. pp. I-I à I-V
- QUENSIERE J. (1990): 'Systémique et pluridisciplinarité: l'exemple du programme d'étude de la pêche dans le Delta Intérieur du Niger.' Seminf 1990.
- SHERRER B. (1983): 'Techniques de sondage en écologie', in Stratégies d'échantillonnage en Ecologie. S. Frontier ed. Masson. 494 pp.