REVUE DE STATISTIQUE APPLIQUÉE

J. ULMO
D. SCHWARTZ
A. VESSEREAU

**Problèmes relatifs aux échantillonnages à plusieurs niveaux**

*Revue de statistique appliquée*, tome 5, n° 1 (1957), p. 57-66
<http://www.numdam.org/item?id=RSA_1957__5_1_57_0>

# PROBLÈMES RELATIFS AUX ÉCHANTILLONNAGES A PLUSIEURS NIVEAUX

*par*

**J. ULMO**

*Statisticienne à l'I.R.S.I.D.*

**D. SCHWARTZ et A VESSEREAU**

*Ingénieurs des Manufactures de l'État*

*La présente étude se rapporte aux problèmes d'échantillonnage du type : échantillonnage à plusieurs niveaux avec nombre constant de prélèvements dans les diverses unités à un même niveau. Par exemple :*

*Dans la production d'un centre de fabrication d'ampoules de B.C.G., on choisit au hasard* l *lots de fabrication, on prélève au hasard dans chacun d'eux un même nombre* a *d'ampoules, chacune donnant naissance à* t *tubes de culture, c'est-à-dire à* t *mesures du nombre de particules vivantes.*

*Ou encore : dans une livraison de papier Kraft, on choisit au hasard* r *rames, dans chacune on prélève au hasard* f *feuilles, chacune donnant naissance à un même nombre* m *de mesures (par exemple de résistance à l'éclatement).*

*De même dans la production d'une cokerie, on choisit au hasard des lots, dans chaque lot on prélève au hasard des sous-lots, ou échantillons, chaque échantillon donnant lieu à un certain nombre de mesures (par exemple de résistance à la dégradation sous l'action de chocs).*

*Dans ces divers exemples, la suite des opérations permet de distinguer d'une part l'échantillonnage proprement dit, d'autre part la mesure d'un caractère donné : ces deux types d'opérations peuvent, dans la pratique, être d'essence et de signification très différentes. Mais dans le modèle mathématique, on peut ne pas tenir compte de cette différence, et considérer un groupe* m *mesures comme un échantillonnage parmi toutes les mesures possibles. Un cas fréquent et simple sera celui de l'échantillonnage à deux niveaux : un niveau d'échantillonnage proprement dit et un niveau de mesure. Par exemple : pour la détermination de la valeur en huile d'une variété d'arachide, on prélève un certain nombre d'échantillons de la variété, et chaque échantillon donne lieu, selon les normes de la méthode internationale, à cinq analyses de teneur en huile, etc.*

*Dans les divers exemples énumérés, la pratique d'un échantillonnage à plusieurs niveaux trouve sa raison d'être dans le fait couramment vérifié que chaque niveau apporte un élément propre de variabilité : par exemple, dans le cas des ampoules de B.C.G. la prise en considération des niveaux « lot » « ampoule », « tube de mesure », résulte du fait qu'il y a une plus grande ressemblance entre tubes provenant d'une même ampoule et entre ampoules provenant d'un même lot.*

*Quant à la constance du nombre des prélèvements dans les unités à un niveau donné, elle est motivée par des considérations telles que les suivantes : dans le cas par exemple des ampoules de B.C.G., on procède au même nombre de mesures pour les diverses ampoules par raison de symétrie, et si on prélève le même nombre d'ampoules dans les divers lots, c'est qu'on admet en première approximation, ou bien que le nombre d'ampoules varie peu d'un lot à l'autre, ou que cette variation est sans rapport avec celle du caractère étudié.*

*Les auteurs envisagent ici deux catégories de problèmes :*

*— En premier lieu, pour un plan d'échantillonnage donné (nombre donné de prélèvements à chaque niveau), estimer à partir des mesures obtenues certaines caractéristiques de tendance centrale ou de dispersion, portant sur la population échantillonnée : par exemple, valeur moyenne du nombre de particules dans les ampoules, variabilité due à l'effet lot, à l'effet ampoule, etc.*

*— En deuxième lieu, rechercher le plan d'échantillonnage optimum en fonction du but fixé qui sera, selon le cas, l'estimation la plus précise d'une caractéristique de tendance centrale ou d'une caractéristique de dispersion.*

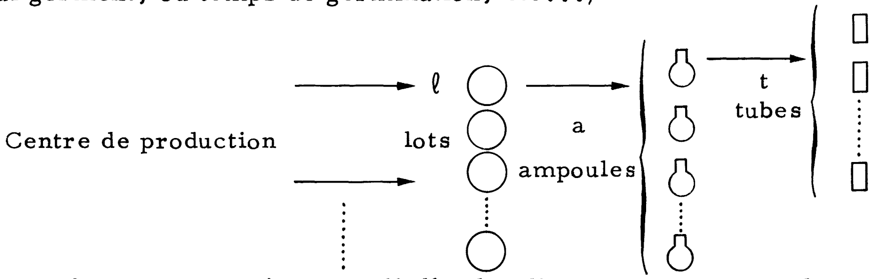
## ESTIMATION SUR ÉCHANTILLON POUR UN PLAN D'ECHANTILLONNAGE DONNÉ

Nous prendrons comme exemple concret celui des ampoules de B.C.G.

Le plan d'échantillonnage est donné :

- choix au hasard de $\ell$ lots de production,
- prélèvement au hasard de a ampoules dans chacun de ces lots
- constitution de t tubes de culture par ampoule.

On détermine sur chaque tube la valeur d'un caractère x (nombre de particules qui germent, ou temps de germination, etc...)




Nous ferons l'hypothèse que l'effet des divers niveaux se traduit par une addition de valeurs aléatoires de telle façon que si $x_{ijk}$ est la valeur déterminée sur le tube K de l'ampoule j du lot i :

(1)
$$x_{ijk} = x_0 + \underset{\text{(lot)}}{\lambda} + \underset{\text{(ampoule)}}{\alpha_{ij}} + \underset{\text{(tube)}}{\theta_{ijk}}$$

Les $\lambda$ ainsi que les $\alpha$ et les $\theta$ sont supposés indépendants en probabilité et distribués normalement autour d'une moyenne nulle, avec des variances :

$\sigma_L^2$ = variance due à l'effet lot

$\sigma_A^2$ = " " " " ampoule

$\sigma_T^2$ = " " " " tube (ou variance due à la mesure)

$x_0$ est donc la moyenne du caractère x pour l'ensemble des lots.

### I) ESTIMATION DE LA VALEUR MOYENNE DE x DANS L'ENSEMBLE DES LOTS

On estimera la moyenne $x_0$ par :

(2)
$$\overline{x} = \frac{\sum_{ijk} x_{ijk}}{\ell at}$$

Cette estimation est, en vertu des hypothèses faites, distribuée normalement autour de $x_0$ avec une variance qui se déduit de (1) et (2) :

$$\overline{x} = \frac{\sum_{ijk} x_{ijk}}{\ell at} = x_0 + \sum_i \frac{\lambda_i}{\ell} + \sum_{ij} \frac{\alpha_{ij}}{\ell a} + \sum_{ijk} \frac{\theta_{ijk}}{\ell at}$$

d'où :

(3)
$$\sigma_{\overline{x}}^2 = \frac{\sigma_L^2}{\ell} + \frac{\sigma_A^2}{\ell a} + \frac{\sigma_T^2}{\ell at}$$

On pourra estimer cette variance, et par suite un intervalle de confiance à un seuil quelconque pour $x_0$, si on peut estimer $\sigma_L^2$, $\sigma_A^2$, et $\sigma_T^2$. On trouvera ci-dessous (paragraphe 3) une méthode d'estimation de $\sigma_L^2$, $\sigma_A^2$, et $\sigma_T^2$, mais il peut être intéressant de tester auparavant la signification des effets lot, ampoule et tube.

2) TESTS DE SIGNIFICATION DES VARIABILITES DUES AUX DIFFERENTS NIVEAUX D'ECHANTILLONNAGE (analyse de la variance)

Il s'agit de voir si les hypothèses nulles $\sigma_L^2 = 0, \sigma_A^2 = 0, \sigma_T^2 = 0$, sont admissibles.

On a, en désignant par $\bar{x}$ la moyenne observée, $\bar{x}_{ij.}$ la moyenne observée pour les tubes de l'ampoule ij, et $\bar{x}_{i..}$ la moyenne observée pour les ampoules du lot i :

$$(4) \quad \sum_{ijk} (x_{ijk} - \bar{x})^2 = \sum_{ijk} (x_{ijk} - \bar{x}_{ij.})^2 + t \sum_{ij} (\bar{x}_{ij.} - \bar{x}_{i..})^2 + at \sum_{i} (\bar{x}_{i..} - \bar{x})^2$$

ou
$$Q = Q_T + Q_A + Q_L$$

avec les degrés de liberté

$$\ell at - 1 = (t-1)a\ell + (a-1)\ell + \ell - 1$$

Les différents termes de l'équation (4) se calculent comme d'ordinaire par :

$$Q = \sum_{ijk} x_{ijk}^2 - \ell at \bar{x}^2$$
$$Q_T = \sum_{ijk} x_{ijk}^2 - t \sum_{ij} \bar{x}_{ij.}^2$$
$$Q_A = t \sum_{ij} \bar{x}_{ij.}^2 - at \sum_{i} \bar{x}_{i..}^2$$
$$Q_L = at \sum_{i} \bar{x}_{i..}^2 - \ell at \bar{x}^2$$

On en déduit les estimations suivantes, indépendantes en probabilité.

TABLEAU I

| Estimations | | Valeurs moyennes | Degrés de liberté |
|---|---|---|---|
| Carré moyen "entre tubes" (d'une même ampoule) | $S_T^2 = \dfrac{Q_T}{a\ell(t-1)}$ | $\sigma_T^2$ | $a\ell(t-1)$ |
| Carré moyen "entre ampoules" (d'un même lot) | $s_{A,T}^2 = \dfrac{Q_A}{\ell(a-1)}$ | $t\sigma_A^2 + \sigma_T^2$ | $\ell(a-1)$ |
| Carré moyen "entre lots" | $s_{L,A,T}^2 = \dfrac{Q_L}{\ell - 1}$ | $at\sigma_L^2 + t\sigma_A^2 + \sigma_T^2$ | $\ell - 1$ |

---

(1) Ce résultat se déduit aisément du calcul de la variance entre moyennes d'ampoules d'un même lot i $\bar{x}_{ij.}$, dont $\dfrac{Q_A}{\ell t(a-1)}$ est une estimation sans biais :

On a en effet $\qquad \bar{x}_{ij.} = x_0 + \lambda_i + \alpha_{ij.} + \dfrac{\sum_k \theta_{ijk}}{t}$

et pour le lot i : $\qquad E_i(\bar{x}_{ij.}) = x_0 + \lambda_i$

soit $\qquad \bar{x}_{ij.} - E_i(\bar{x}_{ij.}) = \alpha_{ij} + \dfrac{\sum_k \theta_{ijk}}{t}$

dont la variance est $\qquad \sigma_A^2 + \dfrac{\sigma_T^2}{t}$

Les expressions des valeurs moyennes des estimations montrent que l'on doit procéder de la manière suivante :

On teste la signification de l'effet "ampoule" en comparant le carré moyen "entre ampoules" $\frac{Q}{(a-I)\ell}$ avec $(a-I)\ell$ degrés de liberté au carré moyen "entre tubes" $\frac{Q_T}{(t-I)a\ell}$ avec $(t-I)\,a\ell$ degrés de liberté par le test F.

On teste la signification de l'effet "lot" en comparant le carré moyen "entre lots" $\frac{Q_L}{\ell-I}$ avec $(\ell-I)$ degrés de liberté au carré moyen "entre ampoules", $\frac{Q_A}{(a-I)\ell}$ avec $(a-I)\,\ell$ degrés de liberté.

D'une façon générale, dans un échantillonnage à plusieurs niveaux, l'effet de chaque niveau est testé par comparaison du carré moyen correspondant à ce niveau à celui qui correspond au niveau immédiatement inférieur.

## 3) ESTIMATION DES VARIANCES AUX DIFFERENTS NIVEAUX D'ECHANTILLONNAGE

a) Estimations.

On a intérêt à connaître la variabilité de la production du centre à chaque niveau. On devra chercher à connaître :

outre $\sigma_T^2$, variance entre mesures finales

les variances $\sigma_A^2$ (due à l'effet "ampoule"),

et $\sigma_L^2$ (due à l'effet "lot")

Le tableau I (Paragraphe 2) montre que la variance $\sigma_T^2$ est estimée directement par :

$$s_T^2 = \frac{Q_T}{(t-I)a\ell} \qquad (5)$$

et que l'on peut estimer sans biais $\sigma_A^2$ et $\sigma_L^2$ respectivement par :

$$s_A^2 = \frac{s_{A,T}^2 - s_T^2}{t} = \frac{Q_A}{\ell t(a-I)} - \frac{Q_T}{\ell at(t-I)} \qquad (6)$$

$$s_L^2 = \frac{s_{L,A,T}^2 - s_{A,T}^2}{at} = \frac{Q_L}{at(\ell-I)} - \frac{Q_A}{\ell at(a-I)}$$

b) Intervalles de confiance.

($\alpha$). - On pourra obtenir une solution approchée à un niveau donné si le nombre de degrés de liberté est élevé à ce niveau; on voit en effet, dans les équations 6 que les variances estimées sont des fonctions linéaires des carrés moyens $\frac{Q_T}{(t-I)a\ell}$, $\frac{Q_A}{(a-I)\ell}$ et $\frac{Q_L}{\ell-I}$ ; or chacun de ces carrés moyens, pour un nombre élevé de degrés de liberté, est distribué normalement avec une variance connue :

- Si $ta\ell$ est élevé, $s_T^2$ est distribué normalement autour de $\sigma_T^2$ avec la variance

$$V\,(s_T)^2 = \frac{2\sigma_T^4}{(t-I)a\ell} \text{ que l'on peut estimer par } \frac{2\,s_T^4}{(t-I)a\ell} \qquad (7)$$

ce qui fixe les intervalles de confiance pour $\sigma_T^2$.

- Si $a\ell$ est élevé (et par conséquent $ta\ell$ a fortiori), $s_A^2$ est une combinaison linéaire de deux variables normales indépendantes, donc distribuée normalement autour de $\sigma_A^2$ avec une variance que l'on peut calculer; on a d'après (6) :

(8) $$\mathrm{Var}\,(s_A^2) = \frac{2}{t^2}\left[\frac{(t\sigma_A^2 + \sigma_T^2)^2}{(a-I)\ell} + \frac{\sigma_T^4}{(t-I)a\ell}\right]$$

dont l'estimation est :

(9) $$\text{estimation de Var}\,(s_A^2) = \frac{2}{t^2\ell}\left[\frac{(s_T^2 + ts_A^2)^2}{a-I} + \frac{s_T^4}{(t-I)a}\right]$$

expression qui sera utilisée plus loin à propos des plans d'échantillonnage optima.

- Enfin, si $\ell$ est élevé, on peut, par un calcul analogue au précédent, estimer la variance de $s_L^2$ en plus des deux autres; on a ainsi les intervalles de confiance pour les variances aux trois niveaux.

($\beta$). - Les solutions approchées ci-dessus, si elles ont des chances de convenir au niveau le plus bas ($\ell$ at grand) ont de moins en moins de chances d'être applicables au fur et à mesure que le niveau monte, le nombre de degrés de liberté diminuant. D'autres solutions doivent alors être recherchées. La solution exacte pour le niveau le plus bas est immédiate. C'est la solution classique reposant sur le fait que :

$$\frac{Q_T}{\sigma_T^2} = \frac{\ell\, a(t-I)\, s_T^2}{\sigma_T^2}$$ est distribué comme $\chi^2$ avec $\ell\, a(t-I)$ degrés de liberté.

On trouvera une méthode de la détermination des limites de confiance pour les variances dues aux autres niveaux d'échantillonnage dans BROSS (Biometrics, 6, 1950, 136) et TUCKEY (Biometrics, 7, 1951, 33). Nous la reproduisons ci-dessous en l'adaptant aux notations utilisées dans notre exemple.

On peut, d'après BROSS, estimer les limites de confiance de $\sigma_A^2$ au risque $\alpha$ par :

- limite inférieure $$L_1 = \frac{F - F_{\alpha/2}}{F'_{\alpha/2}\,F - F_{\alpha/2}}\, s_A^2$$

- limite supérieure $$L_2 = \frac{F - F_{(1-\alpha/2)}}{F'_{(1-\alpha/2)}\,F - F_{(1-\alpha/2)}}\, s_A^2$$

ou $$F = \frac{\text{carré moyen "entre ampoules"}}{\text{carré moyen "entre tubes"}} = \frac{s_T^2 + t\, s_A^2}{s_T^2}$$

$F_{\alpha/2}$ limite fiducielle au risque $\alpha/2$ pour F avec $\ell$ (a-I) et (t-I) $a\ell$ degrés de liberté

$F'_{\alpha/2}$ limite fiducielle au risque $\alpha/2$ pour F avec $\ell$ (a-I) et une infinité de degrés de liberté.

On rappelle que $$F_{n_1,n_2,1-\frac{\alpha}{2}} = \frac{1}{F_{n_2,n_1,\alpha/2}}$$

TUCKEY montre que, dans le cas $F < F_{\alpha/2}$ (variance due au facteur ampoule non significative au risque $\alpha/2$), on doit préférer pour la limite supérieure

$$L'_2 = \frac{F - F_{1-\alpha/2}}{t\, F'_{1-\alpha/2}}\, s_A^2$$

la limite inférieure étant $L_I = 0$.

## PLANS D'ÉCHANTILLONNAGE OPTIMA

On peut chercher les plans qui donnent la meilleure précision possible, soit pour la moyenne, soit pour les variances aux divers niveaux.

### I - ON S'INTÉRESSE A LA MOYENNE

On peut prendre comme optimum un plan qui minimise la variance de la moyenne, soit d'après (3)

$$\sigma_{\bar{x}}^2 = \frac{\sigma_L^2}{\ell} + \frac{\sigma_A^2}{\ell a} + \frac{\sigma_T^2}{\ell at}$$

dont on a une estimation en remplaçant les trois variances par leurs estimations tirées de (5) et (6) . Ce problème, qui présente a priori une infinité de solutions, devra être traité en tenant compte des possibilités offertes pour t, a, et $\ell$ , et des prix des prélèvements à chaque niveau. Il est clair qu'on ne pourra le résoudre qu'en connaissant les estimations de $\sigma_L^2$ , $\sigma_A^2$ , $\sigma_T^2$ , ce qui nécessite une expérience préliminaire.

A côté de ce problème général, qui ne peut être résolu que d'une façon particulière à chaque cas, il est intéressant d'étudier le cas fréquent où <u>on fixe la limite supérieure du nombre total des mesures, aucune autre restriction</u> (prix, durée, etc...) <u>n'étant imposée</u>.

Dans le cas présent, si on suppose donné le nombre de mesures

$$\ell at = n$$

la condition $\sigma_{\bar{x}}^2$ minimum s'écrit :

$$\frac{at\,\sigma_L^2 + t\,\sigma_A^2 + \sigma_T^2}{n} \quad \text{minimum}$$

$$\text{ou} \quad at\,\sigma_L^2 + t\,\sigma_A^2 + \sigma_T^2 \quad \text{minimum}$$

Comme $\sigma_L^2$ , $\sigma_A^2$ , $\sigma_T^2$ , sont fixes et positifs, les entiers a et t vérifiant au mieux les conditions sont :

$$a = t = I$$

avec, par suite, $\ell = n$

L'optimum consiste donc à <u>échantillonner au maximum au niveau le plus élevé, et au minimum aux niveaux inférieurs</u>.

Cette règle est vraie, non seulement quand on impose $\ell at = n$ mais naturellement aussi quand on impose $\ell at \leqslant n$ : en effet si on prend $\ell$ , a, t, tels que $\ell at = n_1 < n$, ce plan est inférieur ou égal au plan $(n_1, I, I)$, lui-même moins bon que le plan $(n, I, I)$.

Elle peut servir en outre de guide si on ajoute quelques restrictions supplémentaires à certains niveaux. Par exemple, si à la condition $\ell at \leqslant n$ on ajoute la condition $\ell \leqslant \ell o$ (avec évidemment $\ell o < n$), on a le plus souvent intérêt à échantillonner au maximum au niveau le plus élevé en prenant $\ell = \ell o$, puis au maximum restant possible au niveau ampoule (c'est-à-dire au plus grand entier contenu dans $\frac{n}{\ell o}$), en prenant toujours le minimum au niveau tube. On ne peut toutefois donner ici à la règle un caractère absolu, car on est amené, dans une telle série de divisions, à perdre des mesures sur le nombre total n permis, ce qui peut faire perdre à la solution son caractère optimum; on peut alors, en utilisant les divisibilités, être amené, soit à prendre une solution légèrement différente, soit à consentir un léger dépassement des limites imposées.

La règle de l'échantillonnage maximum au niveau le plus élevé présente plusieurs avantages importants :

a) Etant indépendante des valeurs de $\sigma_L^2$, $\sigma_A^2$, $\sigma_T^2$, la règle ne nécessite pas leurs estimations. Elle ne nécessite donc aucune expérience préliminaire pour calculer des estimations de $\sigma_L^2$, $\sigma_A^2$, $\sigma_T^2$, et n'est pas influencée par l'imprécision relativement grande de ces estimations. Par contre elle ne permet pas d'estimer $\sigma_L^2$, $\sigma_A^2$ et $\sigma_T^2$. Elle permet toutefois d'estimer $\sigma_{\bar{x}}^2$ donc de déterminer des intervalles de confiance pour la moyenne $x_0$.

b) Etant indépendante des valeurs $\sigma_L^2$, $\sigma_A^2$, $\sigma_T^2$, la règle est par suite indépendante du caractère mesuré. Par exemple, si on s'intéresse à deux caractères mesurés dans les tubes (nombre de colonies, temps de germination), le plan (n,I,I) est optimum pour chacun des caractères (ce qui n'est pas nécessairement vrai dans le cas général).

c) Enfin, il faut signaler qu'on peut avoir, dans un plan d'échantillonnage ($\ell$, a, t) un raté ou un manquant, ce qui se produit généralement au dernier niveau (mesure); par exemple, on a un tube cassé. Si cet accident se produit dans le plan (n, I, I), on a seulement une mesure perdue, et il reste un plan correct (n-I, I, I). Au contraire dans le plan général ($\ell$, a, t) la perte de I tube conduirait en toute rigueur, si l'on voulait conserver un plan d'échantillonnage à t constant, à supprimer toutes les mesures du lot correspondant. On peut éviter cette perte d'information en gardant le plan initial, avec estimation de la donnée manquante; mais le calcul est plus compliqué, moins exact, et ne peut convenir s'il y a trop de manquants.

La règle de l'échantillonnage maximum au niveau le plus élevé s'applique en particulier à toute une catégorie de problèmes à deux niveaux : échantillonnage proprement dit et mesure.

On a alors e échantillons, donnant lieu chacun à m mesures. Il arrive souvent que la seule limitation imposée est le nombre total me = n de mesures. Il faut alors prélever n échantillons, en procédant à une mesure seulement sur chacun. (En langage ordinaire, il vaut mieux mal connaître un grand nombre d'échantillons qu'en bien connaître un petit nombre).

Si pour des raisons de sécurité, on impose la condition de la mesure répétée en double exemplaire au moins, on prendra comme optimum $\left(\frac{n}{2}, 2\right)$. Un tel plan a par ailleurs l'intérêt de donner une idée de l'erreur de mesure.

## 2) ON S'INTERESSE AUX VARIANCES

Il peut arriver qu'on s'intéresse, non à la moyenne, mais aux variances. C'est ce qui arrivera, en particulier, dans le cas où, s'intéressant à la moyenne, on doit, pour dresser le plan optimum, faire un essai préliminaire permettant d'estimer les variances. Mais il se présente ici une difficulté: il y a une variance par niveau et le plan optimum pour la variance due à un niveau ne sera pas nécessairement optimum pour les autres. Des indications générales seront toutefois données plus loin. Nous étudierons d'abord le cas de deux niveaux seulement.

### A) Echantillonnage à 2 niveaux

Nous continuons à utiliser l'exemple des ampoules, et nous supposons qu'on étudie un lot seulement. Le plan d'échantillonnage est caractérisé par :

a = nombre des ampoules (tirées au hasard dans le lot)
t = nombre de tubes de mesure par ampoule

Deux variances peuvent nous intéresser, $\sigma_T^2$ et $\sigma_A^2$

a) <u>Plan optimum pour $\sigma_T^2$</u>

On peut prendre comme plan optimum celui qui minimise la variance de $s_T^2$, dont la valeur est donnée par (7) (en faisant $\ell$ = I)

$$V(s_T^2) = 2\frac{\sigma_T^4}{(t-I)a}$$

Le plan optimum devra rendre maximum le dénominateur $(t-I)a = ta - a$.

Si nous admettons ici encore que la seule restriction imposée soit le nombre des mesures $ta \leqslant n$, il faut rendre minimum a, ce qui donne a = I.

Le plan optimum est donc, ce qui n'est guère étonnant :

une seule ampoule, avec n mesures.

Ce plan ne permet naturellement pas d'estimer $\sigma_A^2$.

Le plan optimum permettant en outre d'estimer $\sigma_A^2$ est : $a = 2$, $t = \frac{n}{2}$.

b) <u>Plan optimum pour $\sigma_A^2$</u>

On peut prendre comme plan optimum celui qui minimise la variance de $s_A^2$, dont la valeur est donnée d'après (8), en faisant $\ell$ = I :

$$V(s_A^2) = \frac{2}{t^2}\left[\frac{(\sigma_T^2 + t\,\sigma_A^2)^2}{a-1} + \frac{\sigma_T^4}{a\,(t-I)}\right]$$

En développant et en posant

$$ta = n \text{ (nombre total de mesures)}$$

$$\frac{\sigma_A^2}{\sigma_T^2} = \rho$$

on a ·

(10)
$$\frac{V(S_A^2)}{2\sigma_T^4} = \frac{t}{n-t}\rho^2 + \frac{2}{n-t}\rho + \frac{n-1}{n\,(n-t)(t-1)}$$

Dans cette expression, $\rho$ ne dépend pas du plan choisi. Si l'on suppose encore que la seule condition est n donné, c'est une fonction de t, et il s'agit de trouver la valeur de t qui la minimise, ceci dans l'intervalle de variation possible pour t, c'est-à-dire :

valeur maximum $t = \frac{n}{2}$ (il faut au moins a = 2)
valeur minimum $t = 2$ (avec t = I, il n'est pas possible de calculer $s_T^2$, et par conséquent $s_A^2$, comme le montrent les équations (6) )

La fonction de t représentée en (10) doit alors être étudiée dans cet intervalle, pour déterminer la valeur qui la rend minimum. On est amené à calculer sa dérivée et à étudier son signe. L'étude détaillée avec la discussion des divers cas possibles, amène finalement aux conclusions suivantes, qui sont fonction du rapport : $\rho = \frac{\sigma_A^2}{\sigma_T^2}$

1° Si $\sigma_A^2 \geqslant \sigma_T^2$ ($\rho \geqslant I$), l'optimum a lieu pour $t = 2$, $a = \frac{n}{2}$. Autrement dit, <u>lorsque la variance due à l'effet "ampoule" égale ou dépasse la variance due aux "tubes" (ou encore dans cet exemple, lorsque la variance due à l'échantillonnage</u>

égale ou dépasse la variance due à la mesure), il faut échantillonner au maximum au niveau le plus élevé, et prendre le minimum t = 2 au niveau le plus bas.

2° Si $\sigma_A^2 < \sigma_T^2$ $(\rho < I)$

soit $\rho_1$ la racine positive de l'équation.

(11) $$\rho^2 n^2 \left(\frac{n}{2} - 1\right)^2 + 2\, n\rho \left(\frac{n}{2} - 1\right)^2 - (n - 1) = 0$$

On doit distinguer deux cas :

Si $\dfrac{\sigma_A^2}{\sigma_T^2} > \rho_1$, l'optimum a lieu pour t égal à la plus grande racine de l'équation

(12) $$t^2 n\rho\,(n\rho + 2) + 2\,t \left[ n - 1 - n\rho\,(n\rho + 2) \right] + n\rho\,(n\rho + 2) - n^2 + 1 = 0$$

racine qui tend vers $I + \dfrac{I}{\rho}$ quand $n \longrightarrow \infty$ si $\rho$ n'est pas trop petit (1).

Si $\dfrac{\sigma_A^2}{\sigma_T^2} \leqslant \rho_1$, l'optimum a lieu pour $t = \dfrac{n}{2}$, $a = 2$.

Cette discussion envisage l'ensemble des cas théoriquement possibles, mais tous ne sont pas également plausibles dans la pratique. C'est ce qu'on peut voir, en particulier, en examinant les diverses conditions énumérées quand n est grand.

Cas de n grand :

1er cas : $\sigma_A^2 \geqslant \sigma_T^2$, l'optimum reste $t = 2$, $a = \dfrac{n}{2}$

2ème cas : $\sigma_A^2 < \sigma_T^2$

La racine $\rho_1$ de l'équation (11) qui sépare les deux éventualités possibles devient $\rho_1 \sim \dfrac{2}{n^2}$

Cette valeur devient rapidement très petite quand n augmente, et du point de vue pratique l'éventualité $\dfrac{\sigma_A^2}{\sigma_T^2} \leqslant \rho_1$ sera très rare, elle correspond à un effet ampoule négligeable.

L'éventualité rencontrée en pratique sera $\dfrac{\sigma_A^2}{\sigma_T^2} > \rho_1$, et le plan optimum défini en conséquence par t = racine de l'équation (12) $\sim 1 + \dfrac{1}{\rho} \sim 1 + \dfrac{\sigma_T^2}{\sigma_A^2}$

En définitive, on prendra comme règle générale le plan optimum défini par

$$t = I + \frac{\sigma_T^2}{\sigma_A^2}$$

avec un minimum de $t = 2$ quand $\sigma_T^2 < \sigma_A^2$

REMARQUE - C'est également le plan optimum pour l'estimation de la moyenne quand on désire avoir une estimation de $\sigma_T^2$ et $\sigma_A^2$.

c) Plan optimum pour $\sigma_T^2$ et $\sigma_A^2$

Le plan optimum pour $\sigma_T^2$ qui permette en même temps l'estimation de $\sigma_A^2$ est, comme on l'a vu :

$$a = 2 \qquad t = \frac{n}{2}$$

(1) Plus précisément, il faut que $\dfrac{I}{\sqrt{n}}$ soit négligeable devant $\rho$

Le plan optimum pour $\sigma_A^2$ varie, selon la valeur de $\frac{\sigma_A^2}{\sigma_T^2}$ du plan :

du plan : $a = \frac{n}{2} \quad t = 2$

au plan : $a = 2 \quad t = \frac{n}{2}$

Ces deux optima sont donc loin d'être toujours conciliables et peuvent même s'opposer.

Il y a un cas où le même plan est optimum pour les deux, c'est le cas où $\frac{\sigma_A^2}{\sigma_T^2}$ est très petit, le plan $\left(a = 2,\ t = \frac{n}{2}\right)$ étant optimum : mais ce cas, qui correspond à une variance négligeable de l'effet ampoule, et où l'échantillonnage à deux niveaux perd à la limite sa signification, n'a guère d'intérêt pratique.

Le cas le plus fréquent en pratique correspondra à une variance d'échantillonnage égale ou supérieure à la variance de mesure. On aura alors un conflit entre :

le plan (a = 2), $t = \frac{n}{2}$ optimum pour $\sigma_T^2$

et le plan $t \# I + \frac{\sigma_T^2}{\sigma_A^2}$ (et souvent t = 2) optimum pour $\sigma_A^2$

Nous signalerons que la solution

$$a = \frac{n}{2} \quad t = 2$$

qui a le double avantage d'être optimum pour $\sigma_A^2$ et pour la moyenne (dès qu'on renonce à la solution a = n, t = I, qui ne donne pas les variances), n'est en même temps pas trop mauvaise dans l'absolu pour $\sigma_T^2$ (quoique étant la moins bonne), parce que le nombre de degrés de liberté à ce niveau reste de toute façon élevé (égal à $\frac{n}{2}$). Si au contraire, on optait pour une solution du type $\left(a = 2,\ t = \frac{n}{2}\right)$ on aurait une solution optimum pour $\sigma_T^2$, mais très mauvaise pour $\sigma_A^2$, le nombre des degrés de liberté à ce niveau tombant à I, - et en outre mauvaise pour la moyenne.

Si on veut trouver un plan qui soit réellement optimum "à la fois pour $\sigma_T^2$ et $\sigma_A^2$", il faut préciser ce qu'on entend par là. On peut (par exemple) rechercher un plan tel que les estimations de $\sigma_T^2$ et $\sigma_A^2$ soient obtenues avec la même précision, ou encore avec la même précision relative : il faudrait dans le premier cas un plan tel que

$$V(s_A^2) = V(s_T^2)$$

et dans le deuxième cas, un plan tel que :

$$\frac{V(s_A^2)}{(s_A^2)} = \frac{V(s_T^2)}{(s_T^2)}$$

relations qui, dans chacun des deux cas, en remplaçant $s_A^2$ et $s_T^2$ par leurs valeurs (8) et (9) donneront t et a.

B) Echantillonnage à plus de 2 niveaux

Les calculs deviennent complexes; en outre, il devient difficile de trouver le plan optimum à la fois pour les divers niveaux.

Des considérations analogues à celles qui viennent d'être exposées pour 2 niveaux conduisent à préconiser, lorsque la seule restriction est le nombre n des mesures, un plan du type :

$$\frac{n}{2^k},\ 2,\ 2,\ 2,\ \ldots,\ 2, \text{ s'il y a } k + I \text{ niveaux.}$$