# Exploitation d'une série de mesures

#### INTRODUCTION.

Un chercheur, dans son laboratoire, peut étalonner régulièrement ses appareils de mesures.

Les longueurs d'ondes de certaines raies, l'énergie de certaines particules, leur durée de vie, la périodicité de tel ou tel phénomène électronique, etc... sont connus avec assez de finesse pour qu'il puisse contrôler leur justesse, et, par des mesures répétées, apprécier leur précision et leur fidélité.

Dès lors que ces qualités sont bien connues, par des considérations qui sont au demeurant de type statistique, et qu'il a pu apprécier la répétabilité de sa méthode de mesure, le chercheur est à même, effectivement, d'estimer une grandeur inconnue à partir des indications de ses appareils, et de proposer un intervalle dans lequel la valeur exacte de cette grandeur a de fortes chances de se trouver. (On omet par contre trop souvent de préciser le niveau de confiance correspondant à cet intervalle, c'est-à-dire la probabilité pour que celui-ci contienne effectivement la valeur exacte.)

Dans nos classes, au contraire, c'est la même collection d'appareils qui sert à tous. Il n'est guère possible de les réétalonner et d'attacher à chacun d'eux sa courbe de correction, si bien que nous connaissons mal leurs erreurs.

Même pour les appareils de mesures électriques, dont on connaît la classe de précision permettant, théoriquement, de déduire une limite supérieure de leurs erreurs, il arrive fréquemment qu'on s'aperçoive que celle-ci, à la longue, est sous-estimée.

Aussi les calculs d'incertitudes classiques, individuels, basés sur la notion d'encadrements  $s\hat{u}rs$ , donnent-ils des incertitudes prohibitives quand on essaie d'être rigoureux, et des intervalles ne recelant que très peu de chances de contenir la valeur exacte, s'excluant parfois mutuellement, quand on réduit les incertitudes « primaires » de manière à ce que l'incertitude relative finale ne soit pas trop grande.

Or, on peut procéder différemment.

Chaque élève ne connaît que très imparfaitement les appareils qu'il emploie. Soit, on ne cherchera donc pas à fonder la recherche d'un intervalle de confiance sur l'étude de leurs performances. Par contre, si nous avons bien le sentiment que nos appareils sont tous un peu faux, nous savons aussi, par expérience, que leurs erreurs n'ont ni la même valeur, ni le même signe, pour les appareils d'une même série. Si nous érigeons cette constatation banale en hypothèse de travail, alors l'exploitation des résultats collectifs, à la fin d'une séance de travaux pratiques, permet de proposer, pour chaque grandeur mesurée simultanément ou à tour de rôle, un intervalle de confiance, en général très convenable, pour le niveau de confiance choisi.

En somme, nous pouvons compenser le manque de connaissance intime de nos appareils de mesures, par une analyse statistique *a posteriori*.

Nous nous proposons donc, dans les pages qui suivent, d'indiquer — ou plutôt de rappeler, car plusieurs articles du B.U.P. (\*) ont déjà traité ce sujet — des techniques permettant d'évaluer, en classe, des intervalles de confiance aux niveaux de confiance choisis. De plus, à l'usage certainement, des seuls professeurs, nous proposons quelques autres sujets de réflexion...

Nous diviserons notre étude en trois parties :

- A. Buts d'une analyse de mesures.
- B. Comparaison entre l'analyse statistique et les calculs d'incertitudes traditionnels.
- C. Propositions de méthodes pour atteindre quelques objectifs .

## A. BUTS D'UNE ANALYSE DE MESURES

#### I Intervalle de confiance et niveau de confiance.

Si nous cherchons à mesurer une grandeur dont la valeur exacte est X, nous pouvons, en considérant un certain nombre de mesures indépendantes de cette grandeur, non seulement estimer X par une valeur  $\overline{X}$  qui est aussi proche que possible de X, mais encore donner un encadrement ( $\overline{X} - \Delta X$ ;  $\overline{X} + \Delta X$ ) pour

<sup>(\*)</sup> N° 505: Mai 1968 (tout le B.U.P.).

Nº 517: Juillet-août-septembre 1969 (page 1231).

Nº 529: Novembre 1970 (page 125).

Nº 531: Janvier 1971 (page 365).

 $N^{\circ}$  537 : Juillet-août-septembre 1971 (page 1091).

Nº 547: Juillet-août-septembre 1972 (page 1221).

lequel nous sommes en mesure de préciser la probabilité correspondante de receler X.

L'intervalle de confiance ainsi déterminé doit être accompagné du niveau de confiance qu'on a choisi.

 $\Delta X$ , classiquement, est l'incertitude au niveau de confiance choisi. Ainsi, si l'on désire que l'intervalle de confiance ait 95 % de chances de contenir X, l'incertitude  $\Delta X_{95\%}$  correspondante sera inférieure à  $\Delta X_{99\%}$  telle que la probabilité pour que l'intervalle  $(\overline{X} - \Delta X_{99\%}; \overline{X} + \Delta X_{99\%})$  contienne X soit de 0,99.

(On dit dans ce cas que l'intervalle de confiance est déterminé au niveau de confiance 99 %).

## Intérêt d'une telle notion.

En physique, comme dans d'autres disciplines, on peut donner des intervalles de confiance encadrant strictement la vraie valeur que l'on cherche à estimer. Mais, ou bien ces intervalles sont très larges car on les majore pour avoir la certitude qu'ils contiennent la vraie valeur, ou bien, pour les réduire tout en conservant cette notion de certitude, il faut apporter aux mesures (et donc à l'appareillage, aux tolérances, etc.) un soin impliquant des investissements considérables en temps, en argent, etc.

En fait, la fierté du physicien, travaillant souvent à la limite des possibilités de ses appareils, le pousse à donner un intervalle de confiance aussi étroit que possible. S'il n'accompagne pas celui-ci de l'indication d'un niveau de confiance, cela conduit à des situations telles que celle qui a trait à la mesure de la vitesse de la lumière par des expérimentateurs (1) tels que Cornu, Foucault, Newcomb, ou Michelson (en 1902). Ces auteurs, parmi d'autres, donnèrent tous des intervalles de confiance, présentés comme des intervalles de certitude alors justement, qu'ils ne contiennent pas la valeur de c actuellement admise!

En fait, les intervalles qu'ils avaient annoncés ne recelaient qu'une certaine probabilité de présence de la vraie valeur de c.

De même, nous prenons toujours pour incertitude (supposée au niveau de confiance 100 %) sur les indications des appareils de mesures électriques, la valeur déduite de la classe de précision de l'appareil, et de la considération du calibre utilisé.

Or, il n'est pas rare, dans un établissement, de constater, en plaçant tous les ampèremètres de la collection (et de la même classe de précision) en série, que l'étendue de leurs indications n'est pas compatible (parce que trop large) avec cette notion

<sup>(1)</sup> Voir Bruhat, Optique 6me éd., page 29.

d'intervalle d'incertitude contenant à coup sûr la vraie valeur de I. (C'est en général moins flagrant pour les voltmètres qui vieillissent mieux car les shunts sont plus fragiles que les résistances additionnelles!)

Nous montrerons au § C. VIII qu'on peut, dans un lycée donné, retoucher légèrement cette notion d'incertitude sur une mesure de I, U, ou P, en considérant l'ensemble des indications des appareils mesurant une même grandeur. On sera alors capable de préciser le niveau de confiance qu'on attache à l'incertitude choisie (qui pourra être celle que donne la classe de l'appareil).

Insistons sur le fait que le traitement statistique des données, en classe, ne constitue qu'un *moyen* pour compenser notre connaissance insuffisante des appareils, particulièrement en ce qui concerne leur justesse.

La méthode statistique ne doit donc pas être présentée comme une fin. On ne peut d'ailleurs pas prétendre que ce mode de discussion des mesures soit toujours utilisé, ou même utilisable, dans les laboratoires de recherche ou de métrologie.

Cependant, quels que soient la méthode employée et les moyens dont on dispose, le but recherché est toujours le même : accompagner chaque mesure définitive d'une grandeur d'un intervalle de confiance.

Nous allons maintenant, à partir d'un exemple, essayer de montrer la diversité des problèmes qui peuvent se poser et à propos desquels nous pouvons utiliser une méthode d'analyse statistique. Il ne faut pas oublier, en effet, que si les opérations de type mathématique auxquelles on peut se livrer sur une série de mesures donnent toujours des résultats, il faut savoir si ces opérations sont fondées, et, si elles le sont, ce que signifient les résultats de l'analyse.

# II. Quels problèmes pouvons-nous poser?

Supposons qu'un établissement possède un lot de dix bobines à noyau de fer provenant du même constructeur, et possédant toutes, les mêmes caractéristiques nominales.

Un expérimentateur  $E_1$ , utilisant une méthode  $M_1$  (par exemple celle des trois voltmètres, avec U et R fixés, voir la figure 1) peut mesurer, dans ces conditions expérimentales, les résistances équivalentes de ces dix bobines. Notons  $S_1$  la série de nombres qu'il obtient.

Dix expérimentateurs, opérant à tour de rôle, peuvent, en utilisant la méthode M<sub>1</sub>, mais des appareils différents, bien que possédant les mêmes caractéristiques nominales, estimer chacun

à son tour la résistance équivalente d'une bobine particulière Bo.

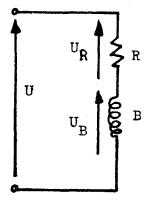


Fig. 1

Soit  $S_2$  la série des résultats des mesures obtenues dans ce cas là.

Les dix expérimentateurs peuvent encore mesurer, chacun pour son propre compte, la résistance de sa bobine, en suivant toujours la méthode  $M_1$ .

Soit S<sub>8</sub> cette troisième série de résultats de mesures.

Soit enfin  $S_4$  la série de dix mesures obtenues dans les mêmes conditions par les dix expérimentateurs utilisant cette fois une méthode  $M_2$  (fig. 2), et en particulier, une tension  $U_B$  notablement différente de la première.

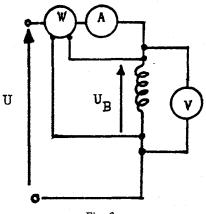


Fig. 2

Il est certain que les quatre séries de résultats n'autorisent pas les mêmes recherches.

Si l'expérimentateur  $E_1$  est assez adroit, si on n'a pas de raison particulière de suspecter la fidélité de son (ou de ses) voltmètre (s), on peut s'intéresser à la dispersion des résultats de la série  $S_1$ .

On aura ainsi une idée de la dispersion des valeurs des résistances des bobines au sein d'une même série de fabrication.

Par contre, la moyenne des mesures est sans objet, parce que le voltmètre utilisé peut très bien être faux et toutes les mesures entachées de ce fait de la même erreur.

La dispersion de la série  $S_2$  provient, elle, de la diversité des appareils utilisés et des expérimentateurs différents qui ont effectué les mesures et les calculs.

Cette série de résultats permet d'estimer la valeur de la résistance  $r_{\circ}$  de la bobine  $B_{\circ}$ , (dans les conditions de mesures fixées à l'avance) ainsi qu'un intervalle de confiance au niveau de confiance choisi, pour cette grandeur physique  $r_{\circ}$ .

Si une autre classe, se livrant à la même opération, (sur la bobine B<sub>o</sub> ou sur une autre) en utilisant la même collection d'appareils, obtient une dispersion nettement plus forte des résultats, alors on peut se poser la question de l'analyse comparée de ces dispersions en vue de savoir si elles révèlent que la deuxième classe, dans son ensemble, manipule moins bien que la première.

La série S<sub>3</sub> permet d'estimer, dans les conditions expérimentales précitées, la valeur moyenne des résistances des bobines de la série. Elle permet en outre de fixer un intervalle de confiance, à un niveau de confiance donné, pour cette valeur moyenne.

Bien entendu, si procédant par permutations circulaires, les dix expérimentateurs ont recommencé sur les bobines  $B_1$ ,  $B_2$ , ...  $B_6$  l'opération qui a donné la série  $S_2$ , on est en mesure d'estimer une nouvelle moyenne à partir des estimations collectives des résistances des bobines qui pourra être beaucoup plus précise.

Il peut sembler, en Physique, que l'estimation de la valeur moyenne des résistances des bobines à partir d'une série de résultats telle que  $S_4$  n'offre pas un grand intérêt.

En fait, en T.P.-cours, par exemple, elle permettra bien plus facilement que par des mesures individuelles, de dégager un phénomène physique.

Soit en effet  $\overline{r}_1$  la valeur moyenne obtenue pour les résistances des 10 bobines, toujours mesurées selon la méthode  $M_1$ , lorsque la tension U est égale à  $U_1$ .

Soit  $\vec{r}_2$  la nouvelle valeur moyenne lorsque  $U = U_2$ , etc.

En traçant la courbe  $\bar{r}=f(U)$ , on sera beaucoup plus à même de savoir si la résistance des bobines à noyau de fer ne dépend de la tension d'alimentation, que si chaque expérimentateur, agissant pour son propre compte, essaie de tracer la courbe r=f(U) à partir de ses mesures personnelles.

Certes, on raisonne sur une grandeur théorique, la résistance moyenne des bobines, mais chacun des points du graphe ainsi constitué est déterminé (parce que les erreurs individuelles se compensent en partie en prenant la moyenne) avec une bien meilleure précision que ceux des graphes obtenus à partir des mesures individuelles.

La série S<sub>4</sub>, enfin, a trait à la méthode M<sub>2</sub> et elle porte sur des conditions de mesures différentes (tension U<sub>B</sub> différente).

Elle peut conduire à une valeur moyenne  $\bar{r}_2$  différente, mais aussi à une dispersion des résultats très différente.

Si la différence des moyennes est très différente, on peut songer à interpréter cette variation comme provenant de la fonction r = f(U). Il existe des tests pour répondre à ce genre de questions (§ C.VI).

Si, enfin, les dispersions des mesures des séries  $S_a$  et  $S_4$  sont très différentes, leur comparaison peut aboutir à la conclusion que l'une des méthodes est supérieure à l'autre quant à la précision.

Bien des séances de T.P. peuvent donner lieu à l'établissement d'une série de mesures du type  $S_1$ ,  $S_2$ ,  $S_3$  ou  $S_4$ , ou à d'autres, car on peut encore combiner différemment les paramètres indépendants que sont les appareils, les objets servant de support aux mesures, et les expérimentateurs.

Cependant, l'estimation d'une grandeur physique proprement dite, ainsi que la détermination de son intervalle de confiance, s'opèreront à partir d'une série telle que  $S_2$ : mesures indépendantes, portant sur le même objet.

Nous venons de voir qu'à côté de ce premier problème, essentiel, il peut être intéressant, pour plusieurs raisons, de chercher à comparer des moyennes ou des dispersions.

Nous indiquerons, dans la partie C, des tests permettant de procéder à ces comparaisons, mais nous rencontrerons également d'autres problèmes comme celui de l'élimination des résultats aberrants, ou celui des ajustements linéaires, etc.

En fait, un problème important se posera souvent, si les mesures des élèves sont assez précises et si leur nombre est suffisant, c'est celui de la recherche des causes d'erreurs expliquant la dispersion observée, et dans certains cas, celui de la recherche des erreurs systématiques.

Le traitement statistique des mesures des élèves, en effet, est inopérant sur une erreur systématique due à la méthode de mesure.

Dans certains cas, toutefois, on pourra être amené à discuter ce type d'erreurs.

## Recherche d'erreurs systématiques.

Supposons que toutes les mesures d'une même série portent sur une même grandeur bien déterminée.

Si ces mesures ont le même poids statistique, c'est-à-dire si elles ont été réalisées par des expérimentateurs d'habileté comparable, avec des appareils de même classe de précision, c'est en prenant leur moyenne qu'on estime le mieux la grandeur mesurée.

On peut également avons-nous dit, centrer sur cette moyenne des intervalles de confiance correspondant à des niveaux de confiance choisis. Or, il peut arriver que l'on se propose de mesurer une grandeur connue (par exemple l'intensité du champ de la pesanteur en un lieu donné, ou la constante d'augmentation de pression d'un gaz à volume constant, dont on sait qu'elle est proche de 1/273).

Dans ce cas, si cette vraie valeur, connue, est extérieure à l'intervalle de confiance correspondant par exemple au niveau de confiance 99 %, ce sera peut-être le signe que certaines erreurs systématiques ont été commises. Il faudra donc les rechercher, et cette recherche aura été motivée par ce premier travail d'encadrement.

En fait, à l'aide d'un nombre relativement petit de n mesures indépendantes, nous chercherons presque toujours à estimer la moyenne  $\mu$  à laquelle aurait permis d'accéder un nombre très important, N, de mesures réalisées dans les mêmes conditions.

Or, si une méthode comporte une erreur systématique, la moyenne  $\mu$  serait elle-même différente de la vraie valeur X de la grandeur mesurée (fig. 3).

## III. De quels moyens disposons-nous?

Dans une salle de Travaux Pratiques, on dispose facilement de six à huit groupes d'élèves qui, en opérant au besoin par permutation circulaire en ce qui concerne la détermination de grandeurs caractéristiques d'objets techniques, peuvent fournir 6 à 8 mesures indépendantes d'une même grandeur.

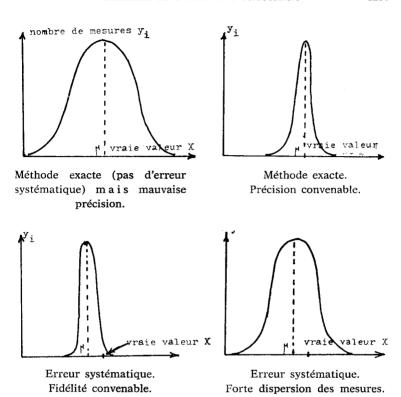


Fig. 3

(Nous considèrerons que deux mesures sont indépendantes lorsqu'elles sont réalisées par des expérimentateurs différents, munis d'appareils différents; cette dernière condition provenant de ce qu'en général nous ne connaissons pas, en classe, les petites erreurs systématiques liées aux appareils de mesure, et variant d'un appareil à l'autre).

Dans ces conditions, une série de six résultats permet généralement, comme nous le verrons, de donner un intervalle de confiance au niveau 95 % ou même 99 % qui n'est pas du tout ridicule.

Supposons que les mesures individuelles des élèves, portant sur une même grandeur X, soient effectuées à la précision « de 10 % », ce qui signifie que la plupart de ces mesures (par exemple 95 %), sont situées dans un intervalle  $(X - \widehat{\Delta}X : X + \widehat{\Delta}X)$ .

avec, justement 
$$\frac{\widehat{\Delta}X}{X} = 0.1 = 10 \%$$
.

Les statistiques nous apprennent alors qu'à partir de chaque échantillon de six de ces mesures, on peut déterminer un intervalle de confiance centré sur la moyenne de ces six mesures,  $\overline{X}$  —  $\Delta X_{95}$  %;  $\overline{X}$  +  $\Delta X_{95}$  %) tel que, en moyenne,

$$\frac{\Delta X_{95\%}}{\overline{X}} = \frac{1}{20} = 5\%$$
, et dans plus de 90 % des cas,  $\frac{\Delta X_{95\%}}{\overline{X}} \le 7\%$ .

En regroupant les résultats de deux demi-groupes de T.P., on pourra d'ailleurs affiner cet intervalle.

Il faudra, par contre, montrer les limites d'une telle précision, obtenue en considérant un nombre croissant de mesures indépendantes : elles proviennent en particulier des erreurs systématiques mises en jeu dans les méthodes utilisées, erreurs que l'on ne peut souvent pas éliminer (surtout en classe).

Nous verrons d'autre part (Annexe I) que si les mesures de départ sont deux fois moins précises, il faut quatre fois plus de mesures pour atteindre la même précision finale. C'est dire que l'amélioration de la précision par la multiplication du nombre des mesures coûte cher!

### B. COMPARAISON AVEC LES CALCULS D'INCERTITUDES CLASSIQUES

Le « calcul d'incertitude » traditionnel se fixait généralement pour objet, connaissant les incertitudes  $\Delta X$ ,  $\Delta Y$ ,  $\Delta Z$ , sur les mesures primaires X, Y, Z, de calculer l'incertitude  $\Delta A$  sur la grandeur A telle que :

$$A = f(X, Y, Z).$$

En fait, on ne connaît jamais X, Y, et Z mais seulement des estimations  $\overline{X}$ ,  $\overline{Y}$ ,  $\overline{Z}$  de ces grandeurs.

En posant  $\overline{X}=X+\delta X$  où  $\delta X$  est l'erreur sur X, et de la même manière  $\overline{Y}=Y+\delta Y$ ,  $\overline{Z}=Z+\delta Z$ , on supposait que :

$$|\delta X| \leqslant \Delta X$$
;  $|\delta Y| \leqslant \Delta Y$ ;  $|\delta Z| \leqslant \Delta Z$ . (1)

La grandeur A étant estimée par  $\overline{A} = f(\overline{X}, \overline{Y}, \overline{Z})$ , il y a évidemment une erreur  $\delta A$  commise sur la détermination de A.

En posant  $\overline{A} = A + \delta A$ , soit  $\delta A = f(\overline{X}, \overline{Y}, \overline{Z}) - f(X, Y, Z)$  on trouve, au premier ordre:

$$\delta \mathbf{A} \simeq f'_{\mathbf{X}}(\overline{\mathbf{X}}, \overline{\mathbf{Y}}, \overline{\mathbf{Z}}) \cdot \delta \mathbf{X} + f'_{\mathbf{Y}}(\overline{\mathbf{X}}, \overline{\mathbf{Y}}, \overline{\mathbf{Z}}) \cdot \delta \mathbf{Y} + f'_{\mathbf{Z}}(\overline{\mathbf{X}}, \overline{\mathbf{Y}}, \overline{\mathbf{Z}}) \cdot \delta \mathbf{Z}$$
(2)

Si l'on suppose que les inégalités (1) sont vérifiées, alors

l'équation (2) entraîne :

$$|\delta A| \leq \Delta A = |f'_x| \cdot \Delta X + |f'_x| \cdot \Delta Y + |f'_z| \cdot \Delta Z.$$

La quantité  $\Delta A$  était alors baptisée : incertitude sur la mesure de A.

## Première critique.

Une telle discussion ne fait pas apparaître toutes les causes d'erreur.

En effet, si la grandeur A est reliée aux variables X, Y, et Z par la relation A = f(X, Y, Z), c'est en général dans certaines conditions. Par exemple, c'est au minimum de déviation que

$$n = \sin \left(\frac{A+D}{2}\right)/\sin (A/2)$$
 et une erreur peut justement pro-

venir de ce qu'on applique cette formule sans que la déviation mesurée soit exactement la déviation minimale.

De même, dans les dosages, le point d'équivalence, décrit par les schémas de réaction, est apprécié par des changements de couleur plus ou moins nets, plus ou moins rapides. Les incertitudes sur les volumes de produits utilisés et le titre de la solution de référence ne sont pas seuls à intervenir dans l'incertitude sur le titre cherché.

#### Deuxième critique.

Les incertitudes  $\Delta X$ ,  $\Delta Y$ ,  $\Delta Z$  étaient généralement estimées de manière empirique.

Ainsi l'incertitude  $\Delta\theta$  sur la mesure d'une température à l'aide d'un thermomètre gradué au cinquième de degré était prise par exemple égale à 0,2 degré (ou à 0,4 degré si les résultats étaient trop éloignés de ceux qu'on attendait). Le moins que l'on puisse dire, c'est que c'était faire beaucoup d'honneur au fabricant.

## Troisième critique.

Le niveau de confiance qu'on pouvait accorder aux diverses incertitudes n'était jamais précisé.

En fait, on supposait qu'il était de 100 %... jusqu'à ce qu'on s'aperçoive que la vraie valeur, lorsqu'elle était connue, était extérieure à l'intervalle trouvé. Il ne restait plus alors qu'à augmenter les incertitudes  $\Delta X$ ,  $\Delta Y$ ,  $\Delta Z$  sur les mesures primaires.

En fait, ce problème est difficile à traiter rigoureusement, même par les méthodes statistiques que nous proposerons, sauf si l'on dispose d'un très grand nombre de mesures; mais même dans les cas où l'on ne dispose pas d'assez de renseignements pour fixer avec précision le niveau de confiance que l'on peut attacher à tel ou tel intervalle d'incertitude, il n'est pas équivalent de dire que celui-ci est de l'ordre de 90 % ou de 99 %, ou de laisser croire qu'il est de 100 %, autrement dit qu'il s'agit d'un intervalle de certitude de présence de la vraie valeur cherchée.

On voit que le calcul d'incertitude classique, dont le but, louable, était de déterminer un intervalle encadrant à coup sûr la vraie valeur cherchée, manquait parfois son objectif parce qu'il partait de bases trop fragiles, et parce qu'il ne prenait en compte que les erreurs « calculables ».

Enfin, s'il donnait de bons ordres de grandeur, son principe était incompatible avec une conception probabiliste des erreurs de mesure. Nous nous proposons de montrer sur un exemple qu'il est vain d'espérer marier les deux méthodes.

Pour fixer les idées, nous allons discuter un exemple classique de calcul d'incertitude : calcul de l'incertitude sur la valeur R d'une résistance mesurée avec un voltmètre et un ampèremètre de classes connues, par exemple 1,5 (voir Annexe V).

Nous notons  $\overline{U}$ ,  $\overline{I}$  les valeurs lues sur les appareils, estimant les valeurs vraies U et I qui demeurent inconnues, comme la vraie valeur R=U/I de la résistance.

On suppose qu'on peut parfaitement négliger le courant passant dans le voltmètre (méthode dite de la courte dérivation).

Supposons qu'on puisse affecter à l'intervalle d'incertitude déduit de la classe des appareils, le taux de confiance 99,8 %.

Soient  $\overline{U}=24\,V$  et  $\overline{I}=80\,mA$ , les indications des deux appareils de calibres respectifs  $30\,V$  et  $100\,mA$ . La valeur  $\overline{R}$  estimant R vaut donc  $300\,\Omega$ .

$$\Delta U = (1.5 \cdot 30)/100 = 0.45 V$$
  $\Delta I = (1.5 \cdot 100)/100 = 1.5 mA.$ 

En fait, il y a une chance sur 1 000 pour que la vraie valeur de la tension, U, soit supérieure à  $U_{\rm M}=24,45\,\rm V$ , et une chance sur 1 000 pour qu'elle soit inférieure à  $U_{\rm m}=23,55\,\rm V$ .

De même, la probabilité pour avoir I  $\geqslant$  I<sub>M</sub> = 81,5 mA est de 0,1 % tandis que la probabilité pour avoir I  $\leqslant$  I<sub>m</sub> = 78,5 mA est également de 0,1 %.

Par la formule 
$$\frac{\Delta R'}{\overline{R}} = \frac{\Delta U}{\overline{U}} + \frac{\Delta I}{\overline{I}}$$
, on trouverait  $\Delta R' = 11.5~\Omega$ .

Pour évaluer le niveau de confiance lié à cette incertitude, cherchons quelle est la probabilité pour que R soit par exemple supérieure à  $R_M = \overline{R} + \Delta R' = 311,25 \Omega$ .

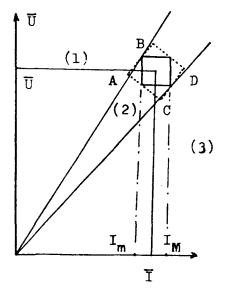
Remarquons que  $R_M$  correspond sensiblement au quotient  $U_M$  24,45

$$\frac{\Omega_m}{\Gamma_m} = \frac{2.5,10}{78,5} \cdot 10^3 \text{ soit } 311,5 \Omega.$$

Pour avoir  $R \geqslant R_M$ , il faut que le point représentatif du couple (U,I) soit dans la région (1) (figure ci-après). De même, pour avoir  $R \leqslant R_m$ , le point doit se trouver dans la région (3).

Or, en supposant que les erreurs  $\delta U$  et  $\delta I$  ont des probabilités de présence gaussiennes (autour de  $\overline{U}$  et de  $\overline{I}$ ), on peut montrer dans le cas particulier envisagé (valeurs respectives de  $\Delta U$ ,  $\overline{U}$ ,  $\Delta I$  et  $\overline{I}$ , ainsi que le niveau de confiance initial de 99,8 % correspondant à 3,1  $\sigma$ ) que la probabilité pour que le point représentatif du couple (U, I) se trouve dans le quadrilatère » ABCD est déjà supérieure à 0,99996.

La probabilité pour qu'il se trouve dans une des régions (1) ou (3) est donc inférieure à 4 • 10<sup>-5</sup>.



On a donc changé de niveau de confiance, puisque l'intervalle classique ( $R_m$ ;  $R_M$ ) dépasse pour R le niveau de confiance 99,996 % alors qu'on était parti, pour U et I du niveau 99,8 %.

Quel est alors l'intervalle de confiance que l'on peut attribuer à R, au même niveau de confiance (ici 99,8 %) que les intervalles  $\overline{U} \pm \Delta U$  et  $\overline{I} \pm \Delta I$ ?

Pour répondre à cette question, il faut avoir des renseignements sur la manière dont sont étalonnés les appareils.

En effet, alors que la vraie valeur de la tension est U, le voltmètre indique  $\overline{U}=U+\delta U$  où  $\delta U$  est une variable aléatoire continue pour laquelle nous avons supposé connue la seule propriété :

Prob. 
$$(| \delta U | > \Delta U = 0.45 \ V) = 0.2 \%$$
.

De même, alors que la vraie valeur de l'intensité du courant vaut I. l'ampèremètre indique  $\overline{I} = I + \delta I$  avec :

Prob. (
$$|\delta I| \geqslant \Delta I = 15 \text{ mA}$$
) = 0,2 %.

On a ensuite

$$\overline{R} = \frac{\overline{U}}{\overline{I}} = \frac{U + \delta U}{I + \delta I} = \frac{U + \delta U}{I \left(1 + \frac{\delta I}{I}\right)} \simeq \frac{U + \delta U}{I} \cdot \left(1 - \frac{\delta I}{I}\right)$$

Soit 
$$\overline{R} = \frac{U}{I} + \frac{\delta U}{I} - \frac{U}{I} \cdot \frac{\delta I}{I}$$
, et en remplaçant U/I par R :

$$\frac{\overline{R} - R}{R} = \frac{\delta R}{R} = \frac{\delta U}{U} - \frac{\delta I}{I}.$$

La variable aléatoire R apparaît donc au premier ordre près, comme une combinaison linéaire des variables aléatoires  $\delta U$  et  $\delta I$ .

Si ces deux dernières sont des variables gaussiennes,  $\delta R$  est également une variable gaussienne (la loi de Gauss est stable pour l'addition mais cette propriété est spécifique de la loi de Gauss.

Dans ce dernier cas, toutefois on montre facilement le résultat suivant : si  $\Delta U$  et  $\Delta I$  sont les intervalles de confiance correspondant au niveau de confiance  $\tau$ , alors l'intervalle de confiance  $\Delta R$ , sur R, au même niveau de confiance  $\tau$  est donné par l'expression :

$$\Delta R = R \sqrt{\left(\frac{\Delta U}{U}\right)^2 + \left(\frac{\Delta I}{I}\right)^2}$$

Dans la pratique, R est estimée à partir des valeurs  $\overline{\mathbf{U}}$  et  $\overline{\mathbf{I}}$ , si bien qu'on prend :

$$\frac{\Delta R}{R} = \sqrt{\left(\frac{\Delta U}{\overline{U}}\right)^2 + \left(\frac{\Delta I}{\overline{I}}\right)^2}$$

Dans le cas où  $\delta U$  et  $\delta I$  ne sont pas des variables gaussiennes, l'incertitude  $\Delta R$  au niveau de confiance  $\tau$  est bien entendu inférieure à la quantité  $\Delta R' = \overline{R} \cdot (\Delta U/\overline{U} + \Delta I/\overline{I})$ , mais son expression exacte dépend de la forme des lois auxquelles obéissent les variables  $\delta U$  et  $\delta I$ .

Nous dirons un mot, en Annexe V, de la forme de ces lois.

## Que peut-on conserver des calculs d'incertitudes classiques?

Le « calcul d'incertitude », mené à son terme, est à proscrire, mais le calcul des petites variations, sur lequel on s'appuyait pour établir sa validité — alors que les prémisses étaient fausses — demeure lui, parfaitement à sa place dans une réflexion sur les incertitudes.

Certes, la formule (2) du début du § B est pratiquement inexploitable dans la plupart des cas, d'une part parce qu'on ne connaît pas les signes des erreurs  $\delta X$ ,  $\delta Y$ , et  $\delta Z$ , et d'autre part, parce qu'on ignore même, à franchement parler, dans la plupart des cas, les limites supérieures des quantités  $|\delta X|$ ,  $|\delta Y|$ , et  $|\delta Z|$ .

Cependant, une formule du type :

$$\delta A = f'_{x} \cdot \delta X + f'_{y} \cdot \delta Y + f'_{z} \cdot \delta Z,$$

ou, si f(X, Y, Z) s'écrit sous forme de monôme  $X^a \cdot Y^b \cdot Z^r$ :

$$\frac{\delta A}{A} = \alpha \frac{\delta X}{X} + \beta \frac{\delta Y}{Y} + \gamma \frac{\delta Z}{Z}$$

a le mérite de montrer sur quelle mesure directe il faut particulièrement soigner l'acquisition du résultat.

(Si  $\alpha = 1$ ;  $\beta = 2$ ;  $\gamma = 1$ ; il est visible qu'une erreur sur Y aura sur  $\delta A$ , une répercussion plus importante qu'une erreur sur X ou Z se traduisant par la même erreur relative.)

Nous avons vu au § précédent que si  $\Delta X$ ,  $\Delta Y$ , et  $\Delta Z$  sont des intervalles de confiance, correctement évalués, à un niveau de confiance  $\tau$  donné, le calcul classique conduisait à un intervalle de confiance trop élevé pour  $\Delta A$ . Il donne toutefois un ordre de grandeur correct.

Cet ordre de grandeur peut être utilisé pour rechercher, le moment venu, la cause de certaines erreurs.

Reprenons l'exemple d'une résistance R que dix élèves, opérant sur des appareils différents, mais dans des conditions semblabes, (tensions voisines), mesurent à tour de rôle.

Soient  $\Delta U$  et  $\Delta I$ , calculés à partir de la classe des appareils,

et tels par exemple, que 
$$\frac{\Delta U}{\overline{U}} = 5\%$$
 et  $\frac{\Delta I}{\overline{I}} = 3\%$ .

Supposons que, grâce aux dix mesures différentes de R, on puisse déterminer, grâce aux techniques que nous verrons bientôt, un intervalle de confiance à 95 % de R; de largeur  $2\Delta R$ .

Il est bien évident, d'après ce que nous avons déjà dit, qu'on

devrait avoir 
$$\frac{\Delta R}{\overline{R}} < 8 \%$$
.

S'il en était autrement, il faudrait en rechercher la cause soit dans une détérioration très sensible de la collection des appareils, soit dans un manque de soin des opérateurs, soit, encore dans la présence de toute autre cause de variabilité négligée a priori.

Ce qui précède montre que pour être utilisables, les calculs d'incertitudes classiques doivent porter sur des quantités  $\Delta X$ ,  $\Delta Y$ ,  $\Delta Z$  telles que les intervalles  $(\overline{X} - \Delta X \; ; \; \overline{X} + \Delta X)$  etc. contiennent à coup sûr la vraie valeur X (resp. Y, Z, etc.).

Dans ces conditions, l'intervalle  $(\overline{A}-\Delta A\;;\;\overline{A}+\Delta A)$  serait lui aussi un intervalle de certitude de présence de la vraie valeur A.

Il faut, de plus, que les conditions d'application de la formule liant A aux quantités X, Y, et Z soient parfaitement vérifiées.

C'est l'ensemble de ces contraintes qui est rarement réalisé, ou, s'il l'est, qui conduit à des intervalles trop larges et donc sans intérêt.

D'ailleurs, lorsque un tel calcul est justifié, le traitement statistique des mesures n'apportera rien de plus.

Si la longueur d'une plaque rectangulaire de faibles dimensions est mesurée au centimètre près, on peut évidemment acquérir la certitude que celle-ci est comprise par exemple, entre 24 et 25 cm; et si dix élèves, procèdent à tour de rôle à sa mesure, en étant munis d'une règle murale graduée au centimètre, ils trouveront le même résultat!

## C. METHODES PROPOSEES POUR ATTEINDRE LES OBJECTIFS DU § A

Nous nous proposons maintenant d'indiquer des méthodes pour répondre de manière plus ou moins précise aux différents problèmes que nous avons soulevés.

Devant les séries statistiques des résultats de nos élèves, selon le problème posé, nous indiquons en général deux méthodes, l'une étant plus efficace, l'autre plus rapide.

Ce choix ne se posera peut-être plus dans quelques années, lorsque tous les laboratoires des lycées disposeront de machines à calculer munies des fonctions statistiques. On pourra alors n'utiliser que la méthode basée chaque fois sur l'estimation de l'écart-type.

Pour l'instant, il me semble que si l'on ne veut pas transformer les séances de T.P. en heures de mathématiques appliquées, il faut proposer des méthodes rapides, faciles à mettre en œuvre, et rappelant de manière claire les buts poursuivis, même si, du point de vue purement scientifique, ces méthodes ne sont pas tout à fait les meilleures. C'est ce que nous ferons en proposant les méthodes basées sur l'étendue.

## I. Hypothèses de travail.

Qu'on le veuille ou non, nous ne disposerons, pour chaque série de résultats que d'un petit nombre de mesures (de 5 à 30).

Pour qu'il en soit autrement, il faut se livrer à une gymnastique peu compatible avec la progression d'une classe : (mobiliser plusieurs classes sur un même sujet, rassembler tous les résultats, attendre une séance ultérieure pour exploiter ces résultats collectifs, etc. constitue un processus que l'on peut bien mettre en œuvre une fois dans l'année, mais non systématiser, et il faut que la discussion des résultats des mesures soit, effectivement, systématique.)

Or, tous les problèmes que nous nous poserons seraient résolus si, disposant d'élèves d'une patience à toute épreuve, d'heures de T.P. à ne plus savoir qu'en faire, d'appareils de qualités semblables en nombre illimité, etc., nous pouvions avoir pour chaque série de mesures deux ou trois mille résultats!

Nous ferons l'hypothèse que si nous disposions effectivement de ces 3000 résultats, que nous les groupions en classes en calculant l'effectif de chaque classe, nous trouverions, en portant tous ces résultats sur un graphique, une courbe de Gauss.

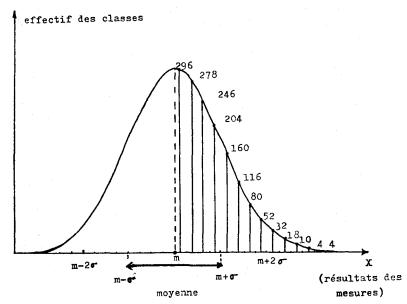
On sait qu'une telle courbe est souvent appelée courbe « normale », justement parce que, très souvent, les conditions sont réunies pour qu'il en soit ainsi (nombreuses causes d'erreurs, d'importances semblables).

L'écart-type σ est l'écart quadratique moyen défini par :

$$\sigma = \sqrt{\frac{\Sigma (X - m)^2}{N}}$$

expression dans laquelle m est la moyenne des mesures et N le nombre total de mesures.

Rappelons que 95,4 % des mesures sont contenues dans l'intervalle  $(m-2\,\sigma\,;\,m+2\,\sigma)$ , soit 95 % dans l'intervalle  $(m-1,96\,\sigma\,;\,m+1.96\,\sigma)$ .



99 % des mesures appartiennent à  $(m-2.57 \sigma; m+2.57 \sigma)$ . 99.73 % appartiennent à  $(m-3 \sigma; m+3 \sigma)$ .

Inversement, et ceci est important, si on fait une seule mesure supplémentaire  $X_i$ , ou si l'on fait une mesure isolée en connaissant par ailleurs  $\sigma$ , il y a 95 % de chances pour que l'intervalle ( $X_i - 1.96 \sigma$ ;  $X_i + 1.96 \sigma$ ) contienne la valeur moyenne m.

Comme nous avons dit qu'en l'absence d'erreur systématique, la moyenne m représente la vrate valeur de la grandeur mesurée, nous en déduisons que l'incertitude, au niveau de confiance 95 % sur la mesure de la grandeur X, estimée par la mesure unique X, vaut 1,96  $\sigma$ .

Toujours pour une loi normale, d'écart-type  $\sigma$  connu, on a donc la correspondance :

NIVEAU DE CONFIANCE	INCERTITUDE
	INCERTITUDE
68 %	1,0 σ
95 %	1,96 σ
95,4 %	2,0 σ
99 %	2,57 σ
99.7 %	3.0 σ

On voit donc que la connaissance de l'écart-type lié à la mesure d'une grandeur dans certaines conditions, permettrait

d'avancer considérablement le problème des incertitudes. Malheureusement, le plus souvent, en classe, on ne connaît pas  $\sigma$ , et on ne peut que chercher à l'estimer.

Nous verrons au paragraphe C. V comment nous pourrons, pour des petits échantillons, vérifier que la composition de ceuxci ne viole pas de manière flagrante l'hypothèse de la normalité des mesures autour de la moyenne m.

En fait, cette hypothèse de normalité n'est nécessaire que pour fixer rigoureusement les niveaux de confiance attachés aux incertitudes ou les seuils de confiance auxquels on accepte ou on rejette telle hypothèse (dire qu'une hypothèse est testée au seuil 5 %, signifie que la probabilité de la rejeter à tort est de 0.05).

Si la loi de distribution des erreurs n'est pas normale, les méthodes exposées resteront valables en principe, mais l'interprétation probabiliste des niveaux de confiance ou des seuils de confiance définis plus haut est plus ou moins approchée.

## II. Estimation d'une moyenne et d'un intervalle de confiance.

Nous continuerons à noter X la valeur exacte inconnue,  $X_1, X_2, ... X_4, X_n$  les n mesures de X entachées d'erreurs, et  $\overline{X}$  la valeur par laquelle nous estimerons X.

## a) Estimation de X.

Si, après élimination éventuelle des résultats aberrants, les n mesures restantes ont a priori le même « poids statistique », le meilleur estimateur de X est la moyenne arithmétique des n mesures.

$$\overline{\mathbf{X}} = \frac{\sum_{i=1}^{n} \mathbf{X}_{i}}{n}$$

Il n'en serait évidemment pas de même si l'on collectait des résultats obtenus dans des conditions très différentes : mesures de g par le pendule simple, les unes en mesurant T sur deux ou trois périodes, les autres sur 30 ou 40; mesures d'une résistance les unes avec des appareils de classe 0,5, les autres avec du matériel courant de classe 1.5.

Dans ce dernier cas, toutefois, on pourrait pondérer les résultats, toutes choses étant égales par ailleurs par l'inverse du carré des classes des appareils (voir Annexe IV).

Nous supposerons toujours que l'estimation de X par la moyenne arithmétique des  $X_i$  sera légitime.

b) Estimation de l'intervalle de confiance de X.

Méthode basée sur l'écart-type et la loi de Student.

On estime d'abord l'écart-type théorique défini au  $\S$  C. I par le coefficient s défini ci-dessous :

$$s = \sqrt{\frac{\sum_{i=1}^{n} (X_{i} - \overline{X})^{2}}{n-1}}$$

(Pour simplifier les calculs, signalons que le numérateur de l'expression qui se trouve sous le radical et qui est la somme des carrés des écarts à la moyenne de l'échantillon des n mesures, s'écrit encore :

$$[(\Sigma X_i^2) - n \overline{X}^2]$$

L'intervalle de confiance est alors donné par l'expression cidessous où le coefficient t, qui suit la loi de Student, dépend, comme le montre le tableau correspondant, du nombre n des mesures effectuées, et du niveau de confiance choisi.

$$\overline{\mathbf{X}} - t \cdot \frac{s}{\sqrt{n}} \leqslant \mathbf{X} \leqslant \overline{\mathbf{X}} + t \cdot \frac{s}{\sqrt{n}}$$

n	2	3	4	5	6	7	8	9	10
t 95 %	12,7	4,30	3,18	2,78	2,57	2,45	2,37	2,31	2,26 3,25
t 99 %	63,7	9,93	5,84	4,60	4,03	3,71	3,50	3,36	3,25

n	12	14	16	18	20	30	50	100	∞ 1,96 2,57
t 95 %	2,20	2,16	2,13	2,11	2,09	2,04	2,01	1,98	1,96
t 99 %	3,11	3,01	2,95	2,90	2,86	2,76	2,68	2,63	2,57

Exemple. Mesure de g. (Pendules simples de longueur  $l=1\,\mathrm{m}$ ; mesure de T sur 50 périodes.)

On prend 
$$g = \frac{4 \pi^2 l}{T^2}$$
. On trouve:  
9.68 9.85 9.85 9.77 9.87 9.79

Soit n = 6.

 $\bar{g} = 9.80 \text{ m/s}^2$ . écart-type estimé :  $s = 7.1 \cdot 10^{-2}$ .  $t_{95\%} = 2.57$ .

$$t \cdot \frac{s}{\sqrt{n}} = \frac{2,57 \cdot 7,1 \cdot 10^{-2}}{2,45} = 7,5 \cdot 10^{-2}.$$

Au taux de confiance 95 %, on a : 9,73  $\leq \frac{4 \pi^2 l}{T^2} \leq 9,87$ .

## Remarque.

Les coefficients ci-dessus sont tirés des tables dites de Student. Lorsque n variables aléatoires  $\mathbf{X}_i$  telles que  $\overline{\mathbf{X}} = \frac{\sum \mathbf{X}_i}{n}$ , etc. suivent la même loi « normale » de moyenne m et d'écart-type  $\sigma$ , on montre en effet que la quantité  $t = \frac{n - \overline{\mathbf{X}}}{s / \sqrt{n}}$  suit une loi, étudiée par Student, à  $\mathbf{v} = n - 1$  degrés de libertés.

Des tables donnant la fonction de répartition de cette loi existent dans tous les ouvrages traitant de statistiques.

On note que la définition de t conduit bien à :

$$m = \overline{X} + t \cdot s / \sqrt{n}.$$

#### Méthode basée sur l'étendue.

En fait, c'est la même méthode que la précédente, mais l'estimation de l'écart-type, au lieu de se faire à partir du coefficient s, est effectuée à partir de l'étendue r des mesures (différence entre la valeur maximale  $X_{\rm M}$  et la valeur minimale  $X_{\rm m}$ ) Annexe II.

$$r = X_{M} - X_{m}$$

q étant un coefficient qui dépend du nombre de mesures conservées, et du niveau de confiance désiré, on a :

$$\overline{X} - q \cdot r \leq X \leq \overline{X} + q \cdot r$$

Nombre de mesures	q 95 %	q 99 %			
2	6.35	31.8			
$\bar{3}$	1.30	3.01			
4	0.72	1.32			
5	0.51	0.84			
6	0.40	0.63			
7	0.33	0.51			
8	0.29	0.43			
9	0.26	0.37			
10	0.23	0.33			

12	0.19	0.28
14	0.17	0.24
16	0.15	0.21
18	0.14	0.19
20	0.13	0.18

Bien que moins précise (pour un même niveau de confiance, elle conduit généralement à des intervalles de confiance légèrement plus larges que la précédente), cette méthode est néanmoins acceptée par l'A.F.N.O.R. (N.F. 06-042) jusqu'à n=12.

Exemple: mesure de g précédente.

$$\overline{g} = 9.80$$
  $r = 9.87 - 9.68 = 0.19$   $q_{95\%} = 0.40$ 

Intervalle de confiance au niveau de confiance 95 % :

en parfait accord avec le résultat précédent.

## Le problème des chiffres significatifs à conserver.

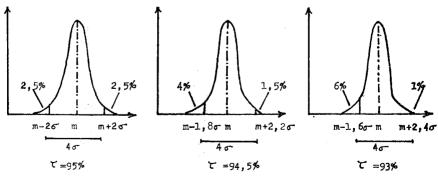
On lie souvent cette question, relative à l'énoncé définitif de la mesure considérée, à la considération de la largeur de l'intervalle de confiance associé.

Notons à ce sujet que la valeur annoncée sert également à centrer l'intervalle de confiance, et qu'un décentrage de celui-ci, à largeur égale, modifie le niveau de confiance associé.

Pour illustrer cette assertion, nous raisonnerons sur une population normale dont on connaîtrait la moyenne et l'écart-type (figures ci-dessous).

Supposons que le niveau de confiance choisi soit de 95 %, correspondant à une « incertitude »  $\Delta X = 2 \sigma$ .

Le fait de décaler le centre de l'intervalle de 0,2  $\Delta X$ , soit 0,4  $\sigma$ , modifie le niveau de confiance de 95 %, sa valeur initiale, à 93 %.



(Les résultats de cette étude ne seraient pas sensiblement modifiés si, au lieu d'une loi normale, on avait raisonné sur la loi de Student à laquelle satisfait la quantité  $(\overline{X} - m)/(s/\sqrt{n})$ .

Le cas que nous avons envisagé correspond très exactement au changement que l'on provoquerait dans le niveau de confiance accompagnant le résultat final, si l'on modifiait le résultat d'une mesure primitivement annoncé sous la forme :

$$X \in 17.7 \pm 1.5$$

et que l'on donnerait finalement sous la forme :

$$X \in 18 \pm 1.5$$
.

Il ne faut donc pas que la suppression des chiffres significatifs entraı̂ne un décentrage de l'intervalle initial supérieur à  $0.1~\Delta X$ , si l'on a choisi le niveau de confiance 95 %. (Un décentrage égal à  $0.1~\Delta X$ ' de l'intervalle de confiance à 99 %, fait passer ce niveau de confiance à 98.7~%, ce qui est également acceptable.)

#### III. Résultats aberrants.

C'est un problème auquel nous sommes fréquemment confrontés. Certaines de nos mesures ne sont-elles pas « douteuses », autrement dit ne s'écartent-elles pas trop des autres ? Ne convientil donc pas de les éliminer ?

Nous allons voir qu'en dehors des raisons expérimentales, techniques, ou évidentes pour éliminer ou corriger un résultat, il existe des tests statistiques des valeurs aberrantes.

Malheureusement, aucun d'eux ne permet de s'assurer avec une efficacité suffisante du caractère aberrant d'une valeur observée. La présence de telles valeurs peut, en effet, être due à des causes très diverses :

- variations aléatoires dues à tel groupe d'élèves au lieu de tel autre;
- erreur de lecture d'un appareil de mesure ou erreur de notation de cette mesure;
- erreur de calcul;
- non normalité de la distribution des résultats : un grand nombre de mesures aurait donné une population très largement étalée vers les petites ou les grandes valeurs, ou vers les deux.

Les tests que nous indiquons ci-après sont sensibles à la non normalité de la loi, pour autant que l'on puisse en juger sur un petit nombre de mesures. Il faudra donc ne les utiliser que lorsque les autres moyens de discussion seront impossibles ou épuisés.

Il peut arriver qu'une mesure soit matériellement incompatible soit avec la nature du phénomène physique étudié, soit avec l'emploi de l'appareil de mesure utilisé, soit pour toute autre raison résultant du caractère particulier du problème étudié : elle ne peut provenir que d'une erreur et il convient évidemment de l'éliminer. Pour cela, on voit qu'il faut demander aux élèves de noter les conditions de leurs observations : calibres utilisés, nombre de divisions, etc.

En ce qui concerne les indications de tel ou tel appareil qui s'écarterait sensiblement des autres, il faudra, là encore, être prudent dans l'élimination des mesures qui en découlent. Nous avons, en effet, admis comme hypothèse, que les erreurs de justesse des divers appareils se compensaient au moins partiellement. Or, c'est peut-être le thermomètre donnant les indications les plus faibles qui permet de compenser celles du reste du lot...

Avant de passer aux tests proprement dits, nous voudrions discuter d'une méthode courante, y compris dans l'industrie, pour éliminer les valeurs que l'on est tenté de considérer comme trop petites ou trop grandes pour qu'elles appartiennent au même ensemble que les autres, il s'agit du test des  $3\sigma$ .

Dans un échantillon de mesures tirées d'une population-mère dont l'écart-type est connu, on considère comme aberrantes les observations s'écartant de la moyenne d'une quantité supérieure à  $3\,\sigma$ .

Or, si la probabilité pour qu'une mesure isolée, tirée d'une population normale, soit supérieure à  $m+3\sigma$  n'est que de  $\varepsilon=1,35\cdot 10^{-3}$ , soit 0,135 %, la probabilité pour que la plus grande des valeurs d'un échantillon de n mesures soit supérieure à  $m+3\sigma$  varie avec n et augmente avec lui.

Le raisonnement qui le prouve est simple : l'événement contraire est celui où toutes les valeurs de l'échantillon sont inférieures à  $m+3\sigma$ .

Cette probabilité vaut  $(1-\epsilon)^n$ .

La probabilité pour que la plus grande valeur (au moins), d'un lot de n, tirées d'une population normale de moyenne m et d'écart-type  $\sigma$  soit supérieure à  $m+3\sigma$  vaut donc  $1-(1-\epsilon)^{n}$ .

Pour 6 mesures, on trouve que cette probabilité vaut 0,8 %.

Pour 20 mesures : 2,7 %.

Pour 75: 10%, etc.

La relation précédente permet en outre de montrer qu'il y a par exemple cinq chances sur cent pour qu'un échantillon de 50 mesures tirées d'une population normale de moyenne m et d'écart-type  $\sigma$  contienne au moins une valeur supérieure à  $m+3.1 \sigma$ .

Si l'on ramène la taille de l'échantillon à 10 mesures, il faut réduire cette valeur à  $m+2.57\,\sigma$  pour qu'il reste toujours cinq chances sur cent d'obtenir un échantillon possédant au moins une mesure dépassant cette limite. Enfin, pour n=5, on trouve que cinq pour cent des échantillons possèdent une mesure supérieure à  $m+2.32\,\sigma$ .

Si on considère comme aberrantes les situations qu'on rencontre normalement moins de cinq fois sur cent, on voit qu'il faudrait prendre comme écart maximal à la moyenne, si elle était connue, de même que  $\sigma$ : 3,1  $\sigma$  pour n=50; 2,57  $\sigma$  pour n=10 et 2,32  $\sigma$  pour n=5.

Cette modulation de l'écart maximal fixant le seuil des valeurs aberrantes en fonction du nombre des mesures est a fortiori nécessaire lorsqu'on ne connaît pas les caractéristiques m et  $\sigma$  de l'ensemble de la distribution et qu'on doit les estimer grâce à la moyenne  $\overline{X}$  de l'échantillon d'une part, et grâce à l'estimateur s précédemment défini ou à l'étendue  $r = X_M - X_m$  d'autre part.

Le test suivant précise d'ailleurs ce point.

#### Les tests.

1° TESTS DE L'ÉCART A LA MOYENNE.

Ordonnons les résultats de mesures dont nous disposons,  $X_1, X_2, ..., X_n$  dans l'ordre de croissance :

$$X_m \dots X_j \dots X_M$$

et supposons que nous doutions de la valeur maximale  $X_m$ . (Dans les expressions ci-dessous,  $X_m$  sera remplacé par  $X_m$  si la mesure douteuse est la valeur minimale  $X_m$ .)

La décision de conserver ou de rejeter la valeur douteuse, d'après S. Aïvazian (Etude statistique des dépendances; Edi-

tions de Moscou) dépend de la valeur du rapport 
$$v = \frac{X_M - \overline{X}}{2}$$

et du seuil de signification  $\alpha$  que l'on a choisi (le seuil de signification est la probabilité, ou encore le risque, évalué en % d'éliminer à tort une mesure douteuse).

On peut remplacer le calcul de v, faisant intervenir l'estimation de l'écart-type  $\sigma$  par la quantité s déjà citée, par celui de w qui estime  $\sigma$  à partir de l'étendue  $r = X_M - X_m$  (Annexe II).

$$w = \frac{X_{M} - \overline{X}}{r}.$$

On se fixe donc un seuil de signification (nous n'avons indiqué que les coefficients relatifs aux seuils 5% et 1%) et l'on cherche dans l'un des tableaux ci-dessous l'un des coefficients  $v_{\max}$   $(n, \alpha)$  ou  $w_{\max}$   $(n, \alpha)$  correspondant au seuil considéré et au nombre n de mesures.

On cherche alors si l'inégalité  $v \leqslant v_{\max}$ , (ou  $w \leqslant w_{\max}$ ) est vérifiée ou non.

Si elle ne l'est pas, on élimine  $X_M$  (ou  $X_m$ ), de toute considération ultérieure.

n	v <sub>max</sub> (5 %)	v <sub>max</sub> (1 %)	w <sub>max</sub> (5 %)	Wmax (1 %)
5	1,87	1,96	0,81	0,84
6	2,00	2,13	0,79	0,84
7	2,09	2,27	0,77	0,84
8	2,17	2,37	0,76	0,83
9	2,24	2,46	0,75	0,83
10	2,29	2,54	0,74	0,82
12	2,39	2,66	0,73	0,82
14	2,46	2,76	0,72	0,81
16	2,52	2,84	0,71	0,80
18	2,58	2,90	0,71	0,80
20	2,62	2,96	0,70	0,79
30	2,79	3,16	0,68	0,77
40	2,90	3,22	0,67	0,76
50	2,99	3,37	0,66	0,75

## Remarque.

Le test ci-dessus est optimiste en ce sens qu'il conduit à conserver, surtout lorsque n est faible, des valeurs assez étonnantes. Il est en fait assez sensible à la non normalité de la population. (Si  $X_M$  est très supérieure à  $\overline{X}$ , cette moyenne est affectée par la présence de  $X_M$  si bien que la différence  $X_M - \overline{X}$  est minimisée, tandis qu'au dénominateur, s ou r sont fortement augmentés. Le rapport v (ou w) atténue donc beaucoup le caractère insolite d'une mesure.

Cet effet va cependant dans le sens souhaité au début de cette discussion.

Nous donnerons deux exemples d'application de cette méthode.

# Exemple.

On a lancé un pendule (dans une classe de Terminale C) et on a demandé aux élèves, sans préparation, en vue de leur faire prendre conscience de la nécessité de procéder plus méthodiquement, de mesurer la durée correspondant à 20 périodes de ce pendule.

Voici la série des nombres obtenus :

40,3 40,4 40,5 40,5 40,5 40,6 40,6 40,8 40,8 40,8 40,9, 41,0 et 41.2 secondes.

On se demande si la dernière mesure est aberrante.

On a: 
$$\overline{X} = 40,68 \text{ s}$$
;  $s = 0,258$ ;  $r = 0.9 \text{ et } n = 13.$ 

Choisissons le seuil standard 5 %. Par interpolation, les tableaux ci-dessous donnent  $v_{max} = 2,43$  ou  $w_{max} = 0,73$ ; or :

$$v = \frac{41,2-40,68}{0,258} = 2,02$$
  $w = \frac{41,2-40,68}{0,9} = 0,58.$ 

Au simple vu des données statistiques, on ne peut donc pas éliminer cette dernière mesure. (En fait, l'appareil donnait, comme on a pu s'en rendre compte par la suite, des résultats toujours supérieurs aux autres et on l'a donc éliminé quand même!)

Autre exemple.

Mesure de g (Terminale D; pendule de 1 m; chronométrage manuel sur 50 périodes):

$$g_s(m/s^2) = 10,49 9,68 9,85 9,85 9,16 9,77 9,87 9,79.$$

$$\overline{g} = 9.81$$
;  $s = 0.36$ ;  $r = 1.33$ ;  $g_M = 10.49$  est « douteuse ».

$$v = (10.49 - 9.81)/0.36 = 1.89 < 2.17$$
; de même  $w = 0.51 < 0.76$ .

On conserve donc toutes les mesures (en l'absence d'autres renseignements).

Nous allons maintenant proposer un autre test, élaboré par Dixon, dans « Annals of Mathematical Statistics, 22 (1951) », et partiellement repris dans la norme NF X 06-041. (Le test réellement proposé par Dixon est plus complet que ce que nous en rapportons.)

2° Test de Dixon.

On range les mesures par valeurs croissantes.

Soit  $X_m$ ,  $X'_m$ , ...  $X'_M$ ,  $X_M$  la suite obtenue. On note  $r = X_M - X_m$  l'étendue.

On calcule alors le rapport

$$Q_m = \frac{X'_m - X_m}{r} \quad \text{ou} \quad Q_M = \frac{X_M - X'_M}{r}$$

selon que c'est la valeur extrême X<sub>m</sub> ou X<sub>M</sub> qui semble douteuse.

Si le rapport correspondant  $(Q_m \text{ ou } Q_M)$  dépasse une valeur  $Q_{max}$  dépendant du seuil de confiance choisi (nous nous sommes limités aux seuils 10 %, ce qui est faible, et 1 %) et du nombre n de mesures, alors on rejette la valeur douteuse et on continue à nouveau le test

seuil 10 %	n	3	4	5	6	7	8	9	10	15	20
Scull 10 70	Qmax	0,94	0,76	0,64	0,56	0,51	0,47	0,44	0,41	0,35	0,32
seuil 1%	n		4	5	6	7	8	9	10	15	20
Scut 1 %	Qmax		0,93	0,82	0,74	0,68	0,63	0,60	0,57	0,475	0,425

## Application aux exemples précédents.

1° Mesures de 20 périodes.

 $Q_M = (41,2-41,0)/0.9 = 0.22$ ; n = 13. Par interpolation, au seuil 10 %, nous trouvons pour un nombre de mesures égal à 13 :  $Q_{max} = 0.38$ .

Considérant le seul point de vue statistique, on ne peut donc rejeter cette valeur.

2° MESURES DE g PAR LE PENDULE SIMPLE.

 $Q_M = (10.49 - 9.87)/1.33 = 0.47$ ; n = 8;  $Q_{max}(8.1\%) = 0.63$ . Là encore, on ne rejettera pas cette valeur douteuse (à moins qu'on ait d'autres raisons de le faire), bien qu'au seuil 10 %, on soit juste à la limite.

## IV. Recherche des causes d'erreur.

Ce point est bien entendu fondamental, et la dispersion observée des résultats de mesures d'une même grandeur sera un bon point de départ pour essayer de les recenser.

Si, après les avoir recensées, on est capable d'en éliminer certaines (par exemple en remplaçant le chronométrage manuel par un chronométrage électronique avec cellules photoélectriques, en atténuant les ronflements par une disposition convenable des masses en électricité, en employant des fils blindés dans l'utilisation de l'oscilloscope, en changeant de méthode de dosage de manière à obtenir un changement de couleur plus rapide en chimie, etc.), on pourra essayer de comparer la dispersion des mesures après atténuation des causes d'erreur, et avant.

Nous n'essayerons pas de dresser une liste des erreurs possibles, qui serait fastidieuse, pas plus que nous ne tenterons de les classifier.

Il faudra cependant veiller aux erreurs systématiques.

En effet, telle erreur systématique qui est indécelable dans une mesure isolée où les causes d'erreurs sont nombreuses peut fort bien apparaître en faisant la moyenne d'une dizaine de mesures indépendantes où les erreurs de type aléatoire se compensent partiellement, tandis que les erreurs systématiques se conservent.

Ainsi, dans une mesure de g par un pendule simple, la for-

mule 
$$g = \frac{4 \pi^2 l}{T^2} \left( 1 + \frac{\vartheta_0^2}{16} \right)$$
, constitue une meilleure approxima-

tion que l'expression simplifiée 
$$g=\frac{4\pi^2 l}{T^2}$$
 que l'on emploie

habituellement.

Or, si  $\vartheta_o = 16^\circ$ , g est sous-estimé de 1 %, et la moyenne d'un grand nombre de mesures donnerait 9,71 au lieu de 9,81 m/s²; en fait, l'expérience prouve que l'incertitude  $\Delta g$ , au niveau de confiance 95 %, élaborée à partir d'une dizaine de mesures indépendantes, est dans une classe normale de Terminale, inférieure, en moyenne à 0,1. Autrement dit, il faut soit tenir compte de  $\vartheta_o$ , soit travailler avec des amplitudes maximales  $\vartheta_o$  très petites.

Si la grandeur mesurée est connue, il n'est pas rare qu'un nombre de mesures indépendantes supérieur à 25 fasse apparaître une erreur systématique qu'il n'est pas toujours facile de localiser.

#### V. Normalité de la distribution des mesures.

(Ce point ne revêt pas, bien entendu, l'importance des précédents, et l'on ne voit pas comment on pourrait l'aborder avec les élèves.) Nous avons supposé (hypothèse de travail du § C. I) que les échantillons de mesures étaient tirés de populations mères « normales ».

Une telle hypothèse est facile à tester lorsque l'effectif des mesures est assez élevé (Test chi-carré,  $\chi^2$ , de Pearson pour n > 50)

ou lorsqu'il est de l'ordre de 30 (méthode graphique connue sous le nom de droite de Henry ou méthode numérique de Shapiro-Wilk, etc.).

Lorsque le nombre des mesures est faible, il est évident qu'il devient difficile de répondre à la question : « Ces mesures sontelles tirées d'une population normale ? »

On peut, cependant, tenter un test graphique qui dépiste au moins les cas où cette hypothèse doit être rejetée de manière flagrante. Par expérience, la méthode que nous indiquons pour cela, semble optimiste. Elle conduit plus souvent à accepter l'hypothèse alors qu'elle est fausse (risque de seconde espèce) qu'à la rejeter alors qu'elle est vraie (risque de première espèce).

Le principe de la méthode est le suivant :

On a tabulé les résultats moyens qu'on obtient en tirant n nombres au hasard dans une population normale réduite ( $\sigma = 1$ ) et centrée (m = 0), et en rangeant ensuite ces nombres par tailles décroissantes.

Les valeurs moyennes obtenues se répartissent symétriquement autour de zéro si bien qu'on n'a indiqué que les valeurs moyennes des n/2 premiers tirages si n est pair, des (n+1)/2 premiers si n est impair.

Procéder alors comme suit (Norme X 06-50):

- a) Classer les n valeurs  $X_1$ ,  $X_2$ , ...  $X_n$  par valeurs décroissantes.
- b) Placer sur un graphique les points d'abscisse  $\xi_{i,n}$  données dans le tableau ci-après et d'ordonnées  $X_{(i)}$ .
- c) Tracer, si cela semble possible, la meilleure droite paraissant représenter l'ensemble des points (c'est la droite de Pearson).
- d) Dans le cas où cet ajustement paraît acceptable (il faut être très exigeant), on peut estimer la moyenne  $\overline{\mathbf{X}}$  de la population par l'ordonnée du point d'abscisse 0 et l'écart-type par la demi différence des points d'abscisses 1 et -1.

Nous donnons, à titre d'exemple les résultats d'un tel tracé (confortés par les techniques de la régression linéaire) dans un cas théorique et dans un cas rencontré en classe.

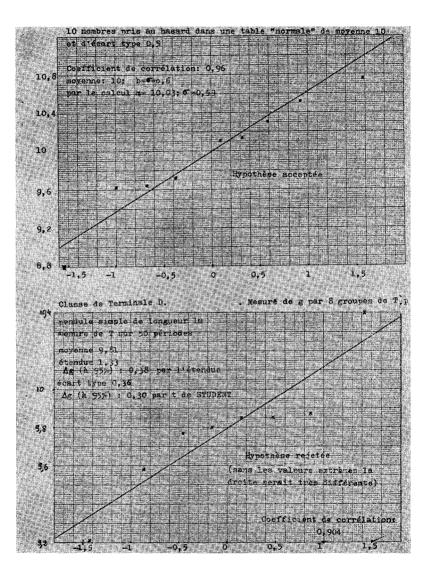
# Espérances mathématiques de variables normales réduites ordonnées

$$\xi_{i,n} = \mathbf{E} \left[ \frac{x_{(i,n)} - m}{\sigma} \right]$$

# (Tables Biometrika)

, P	\$	6	7	8	9	10	11	12	13	14	15	16	17	18	19	20	21
1	1,16	1,27	1,35	1,42	1,49	1,54	1,59	1,63	1,67	1,70	1,74	1,77	1,79	1,82	1,84	1,87	1,89
2	0,50	0,64	0,76	0,85	0,93	1,00	1,06	1,12	1,16	1,21	1,25	1,28	1,32	1,35	1,38	1,41	1,43
3	0,00	0,20	0,35	0,47	0,57	0,66	0,73	0,79	0,85	0,90	0,95	0,99	1,03	1,07	1,10	1,13	1,16
4			0,00	0,15	0,27	0,38	0,46	0,54	0,60	0,66	0,71	0,76	0,81	0,85	0.89	0,92	0,95
5					0,00	0,12	0,22	0,31	0,39	0,46	0,52	0,57	0,62	0,66	0,71	0,75	0.78
6							0,00	0,10	0,19	0,27	0,34	0,40	0,45	0,50	0,55	0,59	0,63
7									0,00	0,09	0,17	0,23	0,30	0,35	0,40	0.45	0,49
8											0,00	0,08	0,15	0,21	0,26	0,31	0,36
9										}			0.00	0.07	0,13	0,19	0,24
10								İ						<i>'</i>	0,00	0,06	0,12
11										l				1			0,00
		L						L		Ĺ	<u> </u>				L		

i	22	23	24	25	26	28	30	32	34	36	38	40	42	44	46	48	50
1	1,91	1,93	1,95	1.97	1,98	2,01	2,04	2,07	2,09	2,12	2,14	2,16	2,18	2,20	2.22	2,23	2.25
2	1,46	1.48	1,50	1,52	1,54	1,58	1,62	1.65	1,68	1,70	1.73	1,75	1,78	1,80	1.82	1,84	1,85
3	1,19	1,21	1,24	1,26	1.29	1,33	1,36	1,40	1,43	1,46	1,49	1,52	1,54	1,57	1,59	1,61	1,63
4	0,98	1,01	1,04	1,07	1,09	1,14	1,18	1,22	1,25	1,28	1,32	1,34	1,37	1,40	1,42	1,44	1,46
5	0,82	0,85	0,88	0,91	0,93	0,98	1,03	1,07	1,11	1,14	1,17	1,20	1.23	1,26	1,28	1,31	1,33
6	0,67	0,70	0,73	0,76	0,79	0.85	0,89	0,94	0,98	1,02	1,05	1,08	1,11	1,14	1,17	1,19	1,22
7	0,53	0,57	0,60	0,64	0,67	0,73	0,78	0,82	0,87	0,91	0,94	0,98	1,01	1,04	1,07	1,09	1,12
8	0,41	0,45	0,48	0,52	0,55	0,61	0,67	0,72	0,76	0,81	0,85	0,88	0,91	0,95	0,98	1,00	1,03
9	0,29	0,33	0,37	0,41	0,44	0,51	0,57	0,62	0,67	0,71	0,75	0,79	0,83	0,86	0.89	0,92	0,95
10	0,17	0,22	0,26	0,30	0,34	0,41	0,47	0,53	0,58	.0,63	0,67	0,71	0,75	0,78	0,81	0,84	0,87
11	0,06	0,11	0,16	0,20	0,24	0,32	0,38	0,44	0.50	0,54	0.59	0,63	0,67	0,71	0,74	0,77	0.80
12		0,00	0,05	0,10	0,14	0,22	0,29	0,36	0,41	0,47	0,51	0,56	0,60	0,64	0,67	0,70	0,74
13				0,00	0.05	0,13	0,21	0,28	0,34	0.39	0,44	0,49	0,53	0,57	0,60	0,64	0,67
14						0,04	0,12	0,20	0,26	0,32	0,37	0,42	0,46	0.50	0,54	0.58	0,61
15							0,04	0,12	0,18	0,24	0,30	0,35	0,40	0,44	0,48	0,52	0,55
16						1		0,04	0,11	0,17	0.23	0,28	0,33	0,38	0,42	0,46	0,49
17	1								0,04	0,10	0,16	0,22	0,27	0,32	0,36	0,40	0,44
18										0,03	0.10	0,16	0,21	0,26	0,30	0,34	0,38
19				İ	}		l '	1			0,03	0,09	0,15	0,20	0,25	0,29	0,33
20				١,			١.				]	0,03	0,09	0,14	0,19	0,24	0,28
21													0,03	0,09	0,14	0,18	0,23
23														0,03	0,08	0,13	0,18
24															0,03	0,08	0,13
24																0,03	0,07
2.5																	0,02



# VI. Comparaison de deux échantillons de mesures.

Nous nous sommes fixés, au § A., des buts tels que : comparaison de deux méthodes, étude critique du travail comparé de deux classes, etc.

Si, bien souvent, on peut traiter ces problèmes de manière intuitive, il n'est pas inutile de savoir que des méthodes objectives existent, et qu'on peut y avoir recours.

Le problème essentiel est de comparer les dispersions, mais, si l'on fait procéder par deux classes à la détermination d'une même grandeur, on peut prévoir que chacune d'entre elles proposera un estimateur différent de cette grandeur.

Nous allons rapidement aborder ce problème :

Soient deux échantillons de mesures d'une même grandeur :

 $n_1$  mesures :  $X_1^{(\alpha)}$ ,  $X_2^{(\alpha)}$ , ...  $X_{i_1}^{(\alpha)}$  ...  $X_{n_1}^{(\alpha)}$  de moyenne  $\overline{X}_1$  et d'étendue  $r_1$ ,

 $n_2$  mesures :  $X_1^{(2)}$ ,  $X_2^{(2)}$ , ...  $X_{n_2}^{(2)}$  de moyenne  $\overline{X}_2$  et d'étendue  $r_2$ .

## Comparaison des mesures.

Si  $\overline{X}_1 \neq \overline{X}_2$ , peut-on accepter l'hypothèse qu'il n'y a eu aucune variation dans la grandeur mesurée par les deux groupes utilisant le même matériel?

Une première réponse est donnée en procédant de la manière

suivante : on calcule la quantité 
$$t = \frac{\overline{X}_1 - \overline{X}_2}{s\sqrt{\frac{1}{n_1} + \frac{1}{n_2}}}$$
 avec :

$$s^{2} = \frac{1}{(n_{1} + n_{2} - 2)} \cdot \left[ \sum_{1}^{n_{1}} (\mathbf{X}_{i}^{(0)} - \overline{\mathbf{X}}_{1})^{2} + \sum_{1}^{n_{2}} (\mathbf{X}_{j}^{(2)} - \overline{\mathbf{X}}_{2})^{2} \right]$$

On traite alors t comme une variable de STUDENT à  $n_1 + n_2 - 2$  degrés de liberté (voir remarque du § C. II).

Pour  $n_1 = n_2 = 10$ , il y a par exemple 5 chances sur 100 pour que t, du seul fait du hasard, soit supérieur à 2,09.

Si on trouve t=2,2, on rejette l'hypothèse de la constance de la grandeur entre les deux séries de mesures, au seuil 5 %.

Deuxième méthode basée sur la seule considération des étendues. Nous nous limiterons au cas où  $n_1 = n_2$ .

On calcule le rapport 
$$d = \frac{\overline{X}_1 - \overline{X}_2}{0.5 \cdot (r_1 + r_2)}$$
 et on compare la

valeur d aux valeurs du tableau ci-dessous qui ont la signification suivante :

Pour n=10, par exemple, il y a 10 % de chances pour que la quantité d, du seul fait du hasard, prenne une valeur supérieure ou égale à 0,25, mais il n'y a plus qu'une chance sur 100 pour que cette valeur soit supérieure à 0,419.

Si on trouve, toujours pour  $n_1 = n_2 = 10$ , d = 0.29, on peut rejeter l'hypothèse de la non variation de la grandeur entre les deux séries de mesures au seuil 10 %, mais non au seuil 5 %.

Effectif commun	seuils de signification (probabilité pour que du seul fait du hasard, d soit dépassé									
n	10 %	5 %	1 %							
5 6 7 8 9 10 12 14 16 18 20	0.49 0.40 0.35 0.31 0.28 0.25 0.21 0.19 0.17 0.15 0.14	0.61 0.50 0.43 0.37 0.33 0.30 0.26 0.23 0.21 0.19	0.90 0.71 0.60 0.52 0.46 0.42 0.36 0.31 0.28 0.25 0.23							

Exemple: mesure de g avec des pendules simples par huit binômes de la même classe opérant une première fois avec des pendules de 1 m de long, une deuxième avec des pendules de 0.36 m.

première série : 10,49 9,68 9,85 9,85 9,16 9,77 9,87 9,79 deuxième série : 9,87 9,87 9,87 9,89 9,80 9,39 9,87 9,74 
$$\overline{g}_1 = 9,81 \qquad r_1 = 1,33 \qquad g_2 = 9,72 \qquad r_2 = 0,48$$
 
$$d = \frac{9,81 - 9,72}{0.91} = 0,10$$

n = 8 la différence n'est absolument pas significative.

Autre exemple: les impédances de six dipôles contenant une résistance et une bobine à noyau de fer en série ont été mesurées par six groupes de T.P. Chaque groupe, travaillant avec la même bobine a fait 2 mesures. L'une pour une tension efficace voisine de 5,5 V, l'autre pour une tension de l'ordre de 27 V (les

tables étaient alimentées à l'aide d'un même transformateur à prises intermédiaires.)

Grâce à des appareils magnétoélectriques à redresseur, sensibles aux seules valeurs moyennes, on a trouvé :

Première mesure:

U : 5.5
 5.5
 5.4
 5.8
 5.4
 5.5
 (V)

 I : 12.5
 13.5
 13.0
 13.0
 13.0
 12.5
 (mA)

 
$$\frac{U}{I}$$
 : 440
 407
 415
 446
 415
 440
 ( $\Omega$ )

Deuxième mesure:

« impédance » moyenne dans le premier cas :

$$\overline{Z}_1 = 425 \Omega$$
; étendue :  $r_1 = 33 \Omega$ ,

« impédance » moyenne dans le second :

$$\overline{\mathbf{Z}}_2 = 391 \ \Omega$$
; étendue :  $r_2 = 20 \ \Omega$ .

$$\frac{r_1+r_2}{2}=26.5 \overline{Z}_1-\overline{Z}_2=34 d=34/26.5=1.28.$$

Or, pour n=6, le seuil de signification à 1% est 0,71. Il s'est donc passé, entre les deux séries de mesures, quelque chose qui a modifié la valeur du rapport U/I mesuré dans les conditions définies ci-dessus.

(Le phénomène physique est complexe; les pertes fer augmentent avec la tension, plus rapidement que celle-ci. La résistance équivalente augmente donc; mais la courbe de magnétisme moyenne est légèrement incurvée, l'inductance moyenne diminue donc. Enfin, le phénomène de l'hystérésis n'étant pas linéaire, le courant et la tension n'ont pas la même forme, les indications des appareils magnétoélectriques sont dès lors difficiles à interpréter).

### Remarque.

Le cas particulier que nous venons d'étudier est justiciable d'un autre type d'étude, en effet, on peut apparier les deux valeurs du rapport U/I pour chaque bobine.

Si ce rapport était invariant dans l'opération, la moyenne des différences entre les deux mesures aurait dû être proche de zéro.

Calculons ces différences Zi - Zi2.

On trouve: 44 23 18 60 32 37  $(\Omega)$ .

Différence moyenne : 35,7  $\Omega$ ; étendue : 38  $\Omega$ .

Intervalle de confiance au niveau 99 % :  $35.7 \pm 0.63 \cdot 38$  soit (12; 59). Zéro ne fait pas partie de cet intervalle : le test est significatif.

## Comparaison des dispersions.

Exemple : en faisant manipuler ses élèves de Seconde avec beaucoup de soin, tel professeur, prenant en compte l'eau entraînée, etc. a obtenu de ses élèves, lors de la détermination de la chaleur massique du fer, les résultats suivants :

0.1104 0.1133 0.120 0.1153 0.1180 0.1170 0.1106 0.1169 0.1176 et 0.1141.

Soient 10 mesures de moyenne 0,1133 cal/g.°C et d'étendue 0.0096.

Dans une autre classe, mais dans d'autres conditions (les élèves avaient beaucoup plus de mesures à réaliser, et bien que pourvus du même matériel que les premiers, ils opéraient certainement avec moins de soin), on a procédé à la même mesure.

Voici les résultats de cette deuxième classe :

0.118 0.104 0.118 0.124 0.116 0.126 0.110.

Soient 7 mesures de moyenne 0,1166 cal/g.°C et d'étendue 0,020 cal/g.

On se demande si l'on peut affirmer que la deuxième série de mesures présente une dispersion significativement plus grande que la première.

Pour cela, on ne peut pas comparer les intervalles de confiance déduits des deux séries, puisqu'ils portent sur des nombres de mesures différents.

(Le premier intervalle, au niveau de confiance 95 %, vaut :

 $0.1133 \pm 0.0022$  (méthode de l'étendue).

Au même niveau de confiance, le second, toujours évalué par la méthode de l'étendue, vaut :

 $0.1166 \pm 0.0073$ .

Calculées à l'aide de la méthode de l'écart-type, les deux

incertitudes au niveau de confiance 95 % valent respectivement 0,0023 et 0,0071. Nous vérifions bien que les deux méthodes donnent, sur ces cas réels, des résultats tout à fait comparables.)

Les statisticiens apportent une réponse à la question que nous nous posons. SNEDECOR a tabulé les valeurs prises par le quotient  $s_1^2/s_2^2$ , les quantités  $s_1$  et  $s_2$  sont toujours définies par les relations :

$$s_1^2 = \frac{1}{n_1 - 1} \cdot \sum_{1}^{n_1} (\mathbf{X_1}^{(1)} - \overline{\mathbf{X}_1})^2$$
 et  $s_2^2 = \frac{1}{n_2 - 1} \cdot \sum_{1}^{n_2} (\mathbf{X_1}^{(2)} - \overline{\mathbf{X}_2})^2$ 

s<sub>1</sub> est par ailleurs le plus grand carré moyen.

Le tableau ci-après indique les valeurs de ce quotient, fonction de  $n_1$ , et de  $n_2$  qui, dans le cas où les lois suivies par les mesures sont normales, n'ont respectivement que 25 %, 10 % et 5 % de chances d'être dépassées du seul fait du hasard.

Bien entendu, les valeurs du quotient croissent lorsque le risque de tirer une conclusion erronée diminue; les coefficients relatifs aux seuils respectifs 25 %, 10 % et 5 % sont écrits de haut en bas.

	$n_1$	nombre	de mesur	es du pl	us grand	carré	moyen
$n_2$	5	6	8	10	16	20	40
5	2.0 4.1 6.3	1 4.0	5 3.98	2.08 3.94 6.00	2.08 3.87 5.86	2.08 3.84 5.80	3.80
6	1.8 3.5 5.1	3.4	5 3.37	1.89 3.32 4.74	1.89 3.24 4.62	1.88 3.21 4.56	3.16
8	1.7 2.9 4.1	6 2.8	8 2.78	1.69 2.72 3.68	1.68 2.63 3.51	1.67 2.59 3.44	1.66 2.54 3.34
10	1.6 2.6 3.6	9 2.6	1 2.51	1.59 2.44 3.18	1.57 2.34 3.01	1.56 2.30 2.94	1.54 2.23 2.83
15	1.5 2.3 <b>3.1</b>	9 2.3	2.19	1.47 2.12 2.65	1.44 2.01 2.46	1.43 1.96 2.39	1.41 1.89 2.27
20	1.4 2.2 2.8	7 2.1	8 2.06	1.41 1.98 2.42	1.38 1.86 2.23	1.37 1.81 2.16	1.34 1.73 2.03
40	1.4 2.0 2.6	9 2.00	1.87	1.34 1.79 2.12	1.30 1.66 1.92	1.28 1.61 1.84	1.24 1.51 1.69

Dans l'exemple que nous avons choisi, c'est la seconde série de mesures qui présente le plus grand carré moyen :  $s_1^2 = 5.83 \cdot 10^{-5}$ .

Le carré moyen correspondant à la première série de mesures vaut  $s_2^2 = 1.02 \cdot 10^{-5}$ .

Le rapport  $s_1^2/s_2^2$  suit la loi F de SNEDECOR, parfois appelée de FISCHER-SNEDECOR et vaut ici 5,71.

Or, par interpolation, nous voyons que pour  $n_1 = 7$  et  $n_2 = 10$ , le rapport F ne prend une valeur supérieure à 3,39 que dans 5 % des cas.

Le test est donc significatif au seuil 5 % (il l'est même, pratiquement au seuil 1 %).

Deuxième exemple : dix observateurs différents ont évalué la masse d'un même objet, d'abord à l'aide de 10 pèse-lettres Posso, ensuite à l'aide de 10 balances ROBERVAL, en procédant par simple pesée (appareils non neufs). Voici les résultats obtenus (en grammes) :

203 197 199 202 201 196 207 209 197 193 moyenne : 200.4;  $s_1 = 5.02$ ;

199,2 199,8 199,9 198,4 199,5 102,4 201,2 100,1 199,3 198,8 avec les balances ROBERVAL.

Moyenne de cette dernière série : 199,86;  $s_2 = 1,18$ .

Le rapport  $s_1^2/s_2^2$  dans ce cas-là vaut 18,1, alors qu'au seuil 5 %, F ne vaut que 3,18. Le test est donc significatif à ce seuil.

Troisième exemple : six groupes d'élèves utilisent des voltmètres de calibre 30 V et des ampèremètres de calibre 100 mA.

Ils mesurent des impédances pratiquement identiques en utilisant leurs appareils dans la deuxième partie de leur graduation (28 V et 70 mA). Ils obtiennent 440 407 415 446 415 et 440.

En reprenant leurs mesures pour des tensions beaucoup plus faibles, sans changer de calibres, les appareils sont mal utilisés : (5 V et 13 mA).

Nous avons vu que le fait de changer de tension affectait certainement le rapport U/I, mais nous ne nous occupons présentement que de la seule dispersion. On a obtenu dans ce cas : 396 384 397 386 383 403, soit  $s_1^2 = 277$ , alors qu'on avait pour la première série :  $s_2^2 = 68$ .

Le rapport  $s_1^2/s_2^2$  vaut 4,1, et la table de la loi F nous indique que le test choisi (voir la norme N.F. X 06-065) est significatif au seuil 10 % (3.45) mais non au seuil 5 %.

# VII. Incertitudes sur les indications d'un appareil.

Mis à part les appareils de mesures électriques, portant l'indication de leur classe, on ne possède pas d'information sur l'ordre de grandeur des erreurs que l'on peut commettre en utilisant un appareil. Sa graduation n'est vraiment pas, en effet, un élément d'appréciation suffisant.

Il est bon, pourtant, comme nous l'avons vu au § IV, d'expliquer la dispersion des résultats calculés à partir des mesures « primaires » réalisées grâce aux appareils.

Puisque, comme nous l'avons dit dans notre introduction, nous ne connaissons pas suffisamment chaque appareil de la collection pour estimer ses erreurs, nous pouvons procéder de manière statistique pour déterminer l'intervalle d'incertitude moyen sur les mesures faites avec un appareil pris au hasard dans la collection.

# Prenons quelques exemples:

Premier exemple: 12 ampèremètres de la collection, placés en série, et alimentés à l'aide d'une pile et d'une résistance ont fourni les indications suivantes (exprimées en mA; le calibre utilisé était 100 mA et la classe de précision nominale: 1,5): 81,4 80,2 80,2 78,4 79,4 79,6 80,4 79,0 80,6 77,8 79,6 80,3.

Première constatation: l'intervalle déduit de la classe des appareils ne peut être un intervalle de certitude puisque la différence entre la plus grande indication et la plus petite, soit 3,6 mA, est supérieure à deux fois l'incertitude nominale, soit  $1.5 \cdot 2 = 3$  mA. Dès lors, on peut se demander, lorsqu'on fait une mesure avec l'un quelconque de ces appareils, si l'on peut lui affecter un intervalle de confiance à un niveau de confiance donné.

Deuxième exemple : on a réalisé des mesures calorimétriques (détermination de chaleurs massiques) en utilisant, entre autres appareils, des thermomètres au dixième de degré. La dispersion des résultats impliquant la présence d'erreurs, on cherche à savoir quelle part est imputable aux mesures de températures.

Pour cela, on place les dix thermomètres de la collection dans un même calorimètre dont le contenu est homogénéisé. On relève, au même instant, leurs dix indications. On trouve (en °C): 32.9 33.0 32.7 32.8 32.85 32.9 32.85 32.85 32.7 32.9.

On se demande, là encore, si on peut déterminer, pour les thermomètres de cette collection, une sorte de classe de précision (l'I.S.O., organisme international dont fait partie l'AFNOR, définit l'écart-type des indications d'un grand nombre d'appareils

ayant les mêmes caractéristiques nominales, et mesurant la même grandeur. Pour certains appareils de mesures concernant les fluides, cet écart  $2\sigma$  constitue même la classe des appareils, selon la norme ISO/R 541).

Troisième exemple : on va procéder à la mesure de g. On se pose, auparavant, la question de l'incertitude dont seront entachées les mesures de durées conduisant à l'évaluation de la période T.

Dans un premier temps, et sans préparation particulière (§ C. III 1°), chaque élève, muni d'un chronomètre, mesure la durée correspondant à 20 périodes d'un même pendule.

Il est bien évident que les résultats ne dépendent pas uniquement des chronomètres, mais aussi de l'habileté des expérimentateurs qui pourront ainsi comparer leurs performances à celles des chronométreurs officiels d'une réunion d'athlétisme...

On trouve: 40.8 40.8 40.4 40.3 40.8 40.5 40.5 40.9 40.6 40.6 40.5 41.0. Quel est, dans ces conditions, l'erreur à craindre sur une telle mesure?

Pour tous ces exemples, il faut se fixer un niveau de confiance, c'est-à-dire la probabilité pour que l'erreur à craindre soit inférieure à la limite que la série statistique va nous permettre de déterminer. 95 % est le niveau standard, mais on peut fort bien choisir 99 %.

Ensuite, il suffit d'estimer l'écart-type de la série des mesures que l'on aurait obtenue, non pas en faisant porter le test sur 8 ou 10 appareils mais sur un grand nombre d'entre eux.

Cette estimation peut se faire soit par le paramètre s déjà souvent rencontré, dont nous rappelons ci-dessous la définition :

$$s = \sqrt{\frac{\sum (X_i - \overline{X})^2}{n-1}} = \sqrt{\frac{\sum X_i^2 - n \overline{X}^2}{n-1}},$$

soit en utilisant l'étendue de l'échantillon (Annexe II).

Selon les résultats du § C. I, on pourra alors prendre  $2\sigma$  pour l'erreur à craindre si l'on se satisfait du niveau 95 %, ou si l'on tient au niveau 99 %, : 2,57  $\sigma$ , etc.

Dans le cas où l'on choisit, pour évaluer cette erreur à craindre, la méthode de l'étendue, nous donnons ci-après un tableau des coefficients b qui serviront à l'évaluation.

On calculera d'abord l'étendue  $r = X_M - X_m$  de la série obtenue. L'erreur à craindre vaut  $\Delta X = b \cdot r$  où le coefficient b

dépend du nombre n d'appareils analysés, et du niveau de confiance souhaité.

n	b <sub>95 %</sub>	b <sub>99 %</sub>	
3	1.18	1.52	
4	0.98	1.26	
5	0.86	1.11	
6	0.79	1.02	
7	0.74	0.95	
8	0.70	0.90	
9	0.67	0.86	
10	0.65	0.84	

n	b 95 %	b 99 %
12	0.61	0.78
14	0.59	0.76
16	0.57	0.73
18	0.55	0.71
20	0.53	0.68
30	0.49	0.63
40	0.46	0.59
50	0.44	0.57

# Applications.

Premier exemple: n=12; s=1.0; r=81.4-77.8=3.60. Au niveau de confiance 95 % l'erreur à craindre, évaluée par la première méthode, vaut 2 mA.

Par l'étendue, elle vaut :  $3.6 \cdot 0.61 = 2.2 \text{ mA}$ .

Si on mesure, avec l'un de ces appareils une intensité donnée, et que l'on trouve  $\overline{I}=85$  mA, on écrira par exemple :

$$82.8 \le I \le 87.2.$$

L'incertitude 1,5 mA, correspondant à la classe de l'appareil, vaut une fois et demie l'écart-type estimé, puisque s=1,0 mA.

Or, la probabilité pour qu'une variable normale soit comprise entre la moyenne m et  $m \pm 2\sigma$  vaut 86 %.

Le niveau de confiance correspondant à la classe nominale serait donc ici de 86 %.

Deuxième exemple : (10 thermomètres mesurent la même température) :

première méthode:

 $\Sigma \vartheta_i^2 = 10788,02$ ;  $\vartheta = 32,85$ ; n = 10;  $n \cdot \vartheta^2 = 10787,9$  soit s = 0,093 degré Celsius.

$$\Delta \vartheta_{95 \%} = 0.19 \,^{\circ}\text{C}$$
  $\Delta \vartheta_{99 \%} = 0.24 \,^{\circ}\text{C}$ 

deuxième méthode:

$$r = 0.3 \,^{\circ}\text{C}$$
;  $b_{(95,10\%)} = 0.65$ ;  $\Delta \vartheta_{95\%} = 0.20 \,^{\circ}\text{C}$ ;  $\Delta \vartheta_{99\%} = 0.25 \,^{\circ}\text{C}$ .

Troisième exemple : (mesure d'une même durée par 12 chronométreurs) :

première méthode:

s = 0.22 seconde;  $\Delta t_{95\%} = 0.44$  seconde

deuxième méthode:

r = 0.7 seconde;  $\Delta t_{95 \text{ oz}} = 0.7 \cdot 0.61 = 0.43$  seconde.

# VIII. Ajustements linéaires.

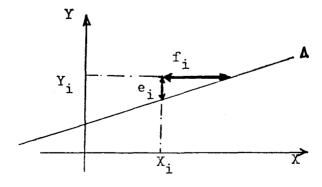
Il arrive souvent qu'on ait besoin, pour accéder à la mesure de certaines grandeurs (f.é.m. et résistance interne d'une pile, coefficient d'augmentation de pression d'un gaz dont le volume est maintenu constant, caractéristique d'une diode à semiconducteur, de type exponentielle mais rendue linéaire après anamorphose, etc.), de tracer une droite dans un système d'axes orthogonaux.

Dans certains cas, on ne s'intéresse qu'au seul coefficient directeur de la droite, dans d'autres on désire de plus estimer par exemple l'ordonnée à l'origine.

Il arrive qu'on soit alors tenté d'appliquer les méthodes dites « de la régression linéaire ». Or, ces méthodes, pour être fondées, exigent qu'on possède sur la variabilité des différents points qui ont été « collectés » des renseignements qu'on ne connaît que très rarement. De plus, il existe différentes droites de « régression » et nous allons montrer qu'il est difficile, parfois, en Physique, de savoir celle qu'il faut prendre.

La méthode la plus connue est la suivante :

Etant donnés les points expérimentaux de coordonnées  $X_i$ ,  $Y_i$ , on cherche la droite d'équation Y = aX + b qui minimise le mieux la somme des carrés des écarts  $e_i$  « parallèles à OY » (figure ci-après) :



En posant encore  $\overline{X} = \frac{\sum X_i}{n}$ ,  $\overline{Y} = \frac{\sum Y_i}{n}$  et  $x_i = X_i - \overline{X}$ ,  $y_i = Y_i - \overline{Y}$ , on trouve:

$$a = \frac{\sum x_i y_i}{\sum x_i^2}, \quad b = \overline{\mathbf{Y}} - a \, \overline{\mathbf{X}}.$$

L'équation de la droite dite de « régression de Y selon X » est donc :

$$Y = \overline{Y} + a(X - \overline{X}).$$

Mais on aurait également pu chercher la droite qui minimise  $\Sigma f_i^z$ , c'est-à-dire les écarts parallèles à **OX**.

Le coefficient directeur de cette nouvelle droite serait :

$$a' = \frac{\sum x_i y_i}{\sum y_i^2}.$$

Si tous les points sont rigoureusement alignés, alors a=a'; sinon, le coefficient  $c=\sqrt{a\,a'}<1$ , appelé coefficient de corrélation est une mesure de la relation réciproque entre les deux variables X et Y.

$$c = \frac{\sum x_i y_i}{\sqrt{(\sum x_i^2) \cdot (\sum y_i^2)}}$$

Souvent, en physique, on n'a pas de raison primordiale pour minimiser la somme des  $e_i^2$  au lieu de celle des  $f_i^2$ ; autrement dit, on n'a pas plus de raisons de tracer la droite représentant Y = f(X), que celle qui rend compte de  $X = f^{-1}(Y)$ .

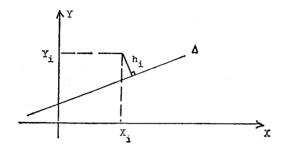
Dans ces conditions, on peut se demander si on ne pourrait pas tracer la droite qui minimise les  $h_i^2$  (fig. ci-après) où les  $h_i$  sont les distances des points  $(X_i, Y_i)$  à la droite.

Cette droite est appelée droite de régression orthogonale.

Malheureusement, non seulement les formules déterminant une telle droite à partir des données  $(X_i,Y_i)$  sont compliquées, mais encore le caractère d'orthogonalité sur la base duquel est construite cette dernière droite ne résiste pas à un simple changement d'échelle. Or, on ne peut admettre, par exemple, que la loi de la relation linéaire entre la longueur d'une tige et sa température dépende de l'échelle choisie pour représenter les grandeurs en présence!

Bien que cette insuffisance de la régression linéaire disparaisse lorsque les variables étudiées sont homogènes et représentables à la même échelle, on ne cherchera pas dans ce genre de détermination une précision qui serait illusoire.

De toute manière, la régression linéaire selon la méthode des moindres carrés n'est qu'une technique qu'on ne peut, et pour cause, parer de toutes les vertus.



Par conséquent, pour évaluer le coefficient angulaire d'une droite exacte inconnue, rendant compte d'un phénomène physique unique, bien déterminé, le mieux sera de demander aux élèves, à partir de leurs mesures individuelles portant sur le même objet physique, d'estimer « à l'œil » la meilleure droite qu'ils peuvent tracer, puis d'en déduire les caractéristiques [coefficient angulaire (a) et ordonnée à l'origine (b)].

On raisonnera ensuite sur les  $(a_i)$  et les  $(b_i)$  selon l'une des méthodes exposées au § C. II, en vue de déterminer, pour ces grandeurs, une moyenne et un intervalle de confiance.

Il sera bon toutefois de fixer les abscisses des points que les élèves devront relever : pour le tracé u = f(i) d'une pile par exemple, on pourra convenir de relever les tensions relatives aux intensités 100, 200, 300, 400, et 500 mA. Ainsi, pour chaque valeur de i imposée à l'avance, le professeur disposera d'autant de valeurs indépendantes de u qu'il y aura de groupe d'élèves. Il sera alors à même, s'il le désire, mais évidemment à titre personnel, de se livrer à une analyse de régression en ayant cette fois des renseignements précieux sur les incertitudes que l'on peut attacher à chacun des points moyens... Si, pour la valeur i = 200 mA, par exemple, les élèves indiquent les valeurs suivantes: 4.15 4.10 3.95 4.01 3.90 3.93 4.10 V. la valeur movenne correspondante est 4,03 tandis que l'écart-type estimé de la variable aléatoire correspondante vaut 0,10 V. On montre alors que les « poids » dont on doit affecter chacun des points « moyens » tels que i = 0.2 A;  $\overline{u} = 4.03$  V, sont inversement proportionnels au carré de l'écart-type (voir Annexe IV).

Pour vérifier une loi (pour établir par exemple que  $\sin i$  est une fonction linéaire de  $\sin r$ , ou que la tension d'une pile dépend linéairement du courant débité, etc...), on pourra, pour gagner du temps, opérer comme nous l'avons signalé au § A. II. On pourra, en effet, procéder à des mesures sur des objets techniques différents. En ayant fixé à l'avance les abscisses des points à relever, il sera possible de raisonner sur un objet technique ayant des caractéristiques « moyennes ». Dans ce cas, en effet, le but de la manipulation n'est pas de procéder à des mesures, mais de s'assurer que deux grandeurs sont liées par une relation d'un certain type.

C'est donc l'alignement des points qui est à rechercher et non un intervalle de confiance pour une grandeur particulière, or le caractère linéaire sera plus probant en utilisant les résultats collectifs, comme il a été indiqué ci-dessus.

# IX. Un exemple d'utilisation de « l'esprit statistique ».

Avoir « l'esprit statistique », ce n'est pas forcément faire des moyennes, calculer des écarts-types, appliquer des tests, etc..., c'est vouloir utiliser les nombreuses mesures de la classe pour compenser ce que chacune d'elles a d'erroné et d'imparfait.

Nous proposons ci-dessous un exemple de cette disposition d'esprit, applicable (?) à la classe de seconde.

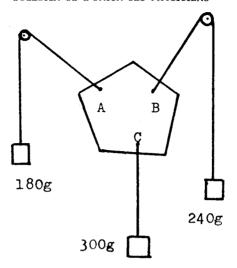
Dans le futur programme, comme dans l'actuel, les enfants auront à vérifier expérimentalement que la somme des vecteurs forces représentant les actions mécaniques localisées auxquelles est soumis un solide en équilibre est nulle.

Habituellement, chacun, pressentant le résultat, joue sur l'épaisseur des traits, l'inclinaison des vecteurs,... pour pouvoir proclamer : « elle est nulle, M'sieur ! »

On pourrait procéder différemment en avertissant les élèves qu'on va procéder à un travail collectif et que c'est seulement après, que les conclusions définitives seront tirées. Il importe donc que la plus grande « honnêteté » soit accordée à l'énoncé de leurs résultats.

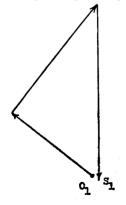
Pour tirer parti de la totalité des mesures, il faut toutefois que les situations étudiées soient identiques.

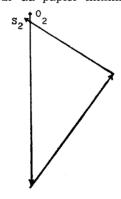
En utilisant des plaques de polystyrène expansé, de poids négligeables, on demandera donc aux élèves de réaliser un équilibre en prenant par exemple deux poulies, et trois masses marquées de 180, 240 et 300 g (ces trois valeurs devraient théoriquement conduire au tracé d'un triangle rectangle). Il faudra encore

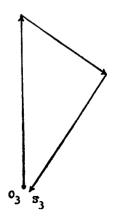


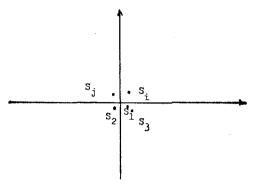
stipuler que la masse de 300 g soit celle qui est suspendue directement, tandis que celle de 180 g sera par exemple à gauche. Il faudra encore que les points d'ancrage soient disposés selon la même géométrie. Bref, toutes les causes de variabilité qui peuvent être recensées et évitées seront supprimées dans le double but de donner à chaque mesure la même importance, et de montrer que, quelles que soient les précautions prises, il reste un caractère aléatoire dans la réalisation d'une mesure. On demande alors aux élèves de représenter, avec la même échelle,

la somme des trois vecteurs forces  $\vec{F}_1 + \vec{F}_2 + \vec{F}_3$  à partir d'un point quelconque O ,sur du papier millimétré.









L'ordre dans lequel ces vecteurs sont tracés à partir de O, est bien entendu laissé au choix de l'élève (il faut croire aux propriétés — associativité et commutativité — de l'addition vectorielle!) Le professeur peut alors relever au tableau, en partant toujours du même point O la position de l'extrémité du

$$\text{vecteur } \overrightarrow{OS}_{\iota} \ = \ \overrightarrow{F}_{1} \ + \ \overrightarrow{F}_{2} \ + \ \overrightarrow{F}_{3}.$$

La figure obtenue, mettant en évidence la dispersion des points suggèrera mieux que ne saurait le faire un relevé individuel, que cette somme vectorielle est certainement nulle, mais que les erreurs de relevé, etc., sont à l'origine des défauts rencontrés.

#### ANNEXE I

# INCERTITUDE SUR UNE MESURE ISOLEE ET INCERTITUDE SUR LA MOYENNE DE PLUSIEURS MESURES DE LA MEME GRANDEUR

#### 1º Résultats mathématiques.

Soit X une variable aléatoire de moyenne  $\mu_X$  et d'écart-type  $\sigma_X$ . A tout échantillon  $(X_1, X_2, ... X_n)$  d'effectif n, on associe sa moyenne

$$\overline{\mathbf{X}}_n = \frac{1}{n} \; \Sigma \; \mathbf{X}_i.$$

On définit ainsi une nouvelle variable aléatoire  $\mathbb{Z}_n$  qui prend pour valeurs les  $\overline{\mathbb{X}}_n$ .

On démontre que : la moyenne  $\mu_Z$  de  $Z_n$  est égale à  $\mu_X$ .

l'écart-type 
$$\sigma_z$$
 vaut  $\frac{\sigma_x}{\sqrt{n}}$ 

Si X suit une loi de Gauss, Zn suit aussi une loi de Gauss.

Si X ne suit pas une loi de Gauss,  $Z_n$  suit pratiquement une loi de Gauss dès que n dépasse quelques unités, si la loi suivie par X n'est pas trop singulière (en particulier pas trop dissymétrique). De toute manière si n > 30,  $Z_n$  suit une loi très proche de celle de Gauss.

# 2° Applications.

On pourrait en déduire que si l'incertitude au taux de confiance  $\tau$ , est  $\Delta X$  sur une mesure d'une grandeur de vraie

valeur X, alors on va pouvoir la réduire à  $\frac{\Delta X}{4}$  en opérant 16 me-

sures consécutives de X chaque fois qu'on veut estimer celle-ci, et en prenant leur moyenne.

En fait, pour que ceci soit juste, il faudrait prendre les seize mesures avec des appareils différents, provenant même de constructeurs différents, pour éviter les erreurs systématiques liées à une fabrication en série, et que ces appareils possèdent cependant des précisions comparables.

Il faut en effet pouvoir admettre une répartition assez régulière des erreurs, de manière à ce qu'elles se neutralisent partiellement, pour utiliser valablement les résultats de la théorie mathématique.

(On retrouve le facteur  $\sqrt{n}$  en dénominateur de l'estimateur s au  $\S$  C. II. b), où l'on se pose le problème de l'estimation d'un intervalle de confiance.)

# ANNEXE II

# RELATION GENERALE ENTRE LES ESTIMATIONS BASEES SUR L'ECART-TYPE (LOI DE STUDENT, etc.) ET CELLES BASEES SUR L'ETENDUE

Le tableau du § C.V (Normalité d'une distribution) permet, entre autres choses, d'estimer l'écart-type à partir de l'étendue.

Le modèle mathématique choisi est toujours celui d'une population normale, simulant la répartition des mesures autour d'une moyenne égale à la grandeur mesurée.

Soit m cette moyenne et  $\sigma$  l'écart-type de la distribution.

Le tableau indique que pour n=8, par exemple, la plus grande des valeurs d'une série de 8 mesures, vaut en moyenne  $m+1.42\,\sigma$ ; de la même manière, la plus petite de la série vaut

en moyenne  $m-1,42\,\sigma$ . C'est dire qu'en moyenne, l'étendue d'une série de 8 mesures vaut 2,84  $\sigma$ . On peut donc estimer l'écart-type de la variable aléatoire correspondante en calculant l'étendue r de cette série de 8 mesures, et en la divisant par 2,84.

Soit s' (nouvel estimateur de 
$$\sigma$$
) =  $\frac{r}{2.84}$ .

Il faut bien noter que s' n'est qu'une estimation de  $\sigma$ , et qu'une autre série de 8 mesures en aurait fourni une autre. Les nombres qu'on obtiendrait en recommençant souvent l'opération seraient assez dispersés. Pour en avoir une idée, il faut noter que pour n=8, le tableau en question nous apprend que la différence entre la plus grande des 8 mesures, soit  $X_M$ , et celle qui la suit immédiatement, soit  $X_M$ , vaut en moyenne :

$$m + 1.42 \sigma - (m + 0.85 \sigma) = 0.57 \sigma.$$

Le rapport  $(X_M - X'_M)/(X_M - X_m)$  doit donc, en moyenne, être voisin de 0,57/2,84 = 0,20 alors que le test de Dixon nous indique que 10 % de ces rapports (voir le tableau du § C. III) sont supérieurs à 0,47 !

L'estimation de  $\sigma$  par la quantité s donne des résultats légèrement moins dispersés lorsque n est faible, et beaucoup plus groupés lorsque n est grand (n>30), c'est en cela que les méthodes utilisant s, sont supérieures à celles qui utilisent l'étendue.

On peut d'ailleurs, comme le montre le tableau cité au début de ce paragraphe, estimer  $\sigma$  à partir d'autres couples de valeurs que les valeurs extrêmes.

Quoi qu'il en soit, chaque fois que nous avons fait figurer l'un à côté de l'autre deux tableaux, l'un portant sur le calcul de s, l'autre sur l'étendue r, il suffit, pour passer de l'un à l'autre, de remplacer s par le quotient  $r/2 \, \xi$ ,  $\xi$  étant le coefficient qui, dans le cas de n=8, vaut 2.84.

Cela est vrai pour les tableaux permettant de déterminer un intervalle de confiance; un des tests des valeurs aberrantes, et le tableau du § C. VII.

# ANNEXE III

# METHODE DE MONTE CARLO

Les tests, les tableaux qui ont été indiqués dans les pages qui précèdent, n'ont été accompagnés d'aucune démonstration de caractère mathématique. C'est que les justifications mathématiques en sont difficiles, et on peut se demander s'il faut vraiment commencer par l'étude des Statistiques et des Probabilités pour traiter les résultats des mesures physiques...

Par contre, on peut très facilement vérifier le bien-fondé des résultats que nous avons indiqués, en simulant des résultats de mesures par des échantillons de nombres au hasard, tirés de nombres à répartition normale (ou présentant d'autres caractères).

Dans l'ouvrage de DIXON et MASSEY: « Introduction to statistical analysis », Mc Graw Hill Book Cy, New-York and London (1951), on trouvera de telles tables de nombres à répartition normale, de moyenne m donnée et d'écart-type également stipulé.

L'échantillonnage à partir de table de nombres au hasard est une technique de grand intérêt, non seulement pour vérifier des résultats tels que ceux que nous avons indiqués, mais encore pour résoudre de difficiles problèmes de probabilités dont aucune solution mathématique n'est encore connue.

Ces techniques sont connues sous le nom de *méthode de Monte Carlo*. A cet effet, les ordinateurs ont des programmes pour créer leurs propres tables de nombres au hasard nécessaires aux calculs.

#### ANNEXE IV

#### PONDERATION DES RESULTATS DE MESURES

Il peut arriver que l'on dispose de plusieurs estimations de la même grandeur, chacune avec une incertitude différente, pour le même niveau de confiance.

Dans ce cas, on peut obtenir une estimation finale de cette grandeur, en prenant la moyenne pondérée des diverses estimations initiales, les poids étant proportionnels aux inverses des carrés des incertitudes.

C'est la conséquence de la théorie des moindres carrés. On sait que si l'on dispose de N nombres  $X_1, X_2, ... X_\ell, ... X_N$ , l'expres-

sion  $S(\mu) = \sum_{i}^{N} (X_{i} - \mu)^{2}$  est minimale quand la variable  $\mu$  est égale à la moyenne  $\overline{X}$  des N nombres  $X_{i}$ .

Supposons donc qu'on tire d'une même population-mère normale, des échantillons d'effectifs  $n_1$ ,  $n_2$ , etc. et soient  $m_1$ ,  $m_2$ , etc. les moyennes correspondantes de ces échantillons.

Si l'on pose  $N = n_1 + n_2 + ...$ , il est évident que la moyenne des N mesures  $X_1$ ,  $X_2$ , ...  $X_{n_1}$  ...  $X_{n_1 + n_2}$  ...  $X_N$  est égale à :

$$\overline{m} = \frac{n_1 m_1 + n_2 m_2 + ...}{n_1 + n_2 + ...}$$

Or, pour un niveau de confiance donné, l'incertitude sur la détermination de la moyenne m de la population-mère en prenant  $m_1$  comme estimation de celle-ci est égale (voir l'Annexe I) à

$$\frac{k\sigma}{\sqrt{n_1}}$$
.

(Le coefficient k dépend du niveau de confiance choisi, et vaut par exemple 1,96 pour le niveau de confiance 95 %.)

Si l'on pose 
$$\Delta m_1 = \frac{k \sigma}{\sqrt{n_1}}$$
, soit  $n_1 = \frac{k^2 \sigma^2}{(\Delta m_1)^2}$ , on voit que la

formule ci-dessus donnant la quantité  $\overline{m}$  (meilleur estimateur de m au sens des moindres carrés) en fonction des estimations initiales  $m_1$ ,  $m_2$ , etc. est bien obtenue en pondérant les  $m_1$  par

des coefficients  $n_i = \frac{k^2 \sigma^2}{(\Delta m_i)^2}$  qui sont bien inversement proportionnels au carrés des incertitudes  $\Delta m_i$ .

Quant à l'incertitude finale sur la moyenne  $\overline{m}$  portant en fait sur N mesures initiales, elle vaut, toujours au même niveau

de confiance caractérisé par 
$$k$$
 :  $\Delta \overline{m} = \frac{k \sigma}{\sqrt{N}}$ .

En remplaçant, dans l'expression de  $N = n_1 + n_2 + n_3 + ...$ 

les nombres  $n_i$  par leur valeeur  $\frac{k^2 \sigma^2}{(\Delta m_i)^2}$ , on arrive très facilement à l'expression :

$$\frac{1}{(\Delta \overline{m})^2} = \Sigma \frac{1}{(\Delta m_i)^2}.$$

# Application.

Six groupes d'élèves ont, à tour de rôle, procédé à deux évaluations successives d'une même résistance en gardant les mêmes calibres d'appareils. Ils ont trouvé :

Pour i = 50 mA:

141,2 140,4 142,6 138,5 138 141,6 movenne 140,4 Ω.

Pour i = 100 mA:

A la première moyenne correspond une étendue de 4,1  $\Omega$ . A la seconde, l'étendue 2,0  $\Omega$ . Comme il y a le même nombre de mesures dans les deux séries, pour un niveau de confiance donné, les incertitudes sont proportionnelles aux étendues.

On pondèrera donc les estimations 140,4 et 140,0 par  $\frac{1}{(4,4)^2}$  et  $\frac{1}{\cdot}$ .

On trouve:

$$\overline{R} = \frac{140,4 \cdot 5,2 \cdot 10^{-2} + 140 \cdot 0,25}{0.302} = 140,1 \Omega.$$

En appliquant la méthode de l'étendue exposée au § C. II, on trouve pour incertitudes au niveau de confiance 99 %, dans le premier cas :  $\Delta R_1 = 0.63 \cdot 4.1 = 2.58 \Omega$ ; dans le second :  $\Delta R_2 = 1.26 \Omega$ .

D'après l'expression générale établie précédemment :

$$\frac{1}{(\Delta R)^2} = \frac{1}{(2,58)^2} + \frac{1}{(1,26)^2}$$

Soit  $\Delta R = 1.13 \Omega$ .

On remarque que les modifications apportées au résultat qu'auraient donné les seules mesures effectuées en bout de calibre, par celles qui ont été obtenues à demi calibre, sont très faibles.

#### ANNEXE V

# ERREURS DE MESURE ET CLASSE DE PRECISION DES APPAREILS ELECTRIQUES

Les appareils de mesures électriques portent l'indication de leur classe de précision. Rappelons brièvement la définition de la classe d'un appareil, en prenant l'exemple d'un voltmètre.

Supposons qu'on mesure une tension U à l'aide d'un voltmètre de calibre 30 V et de classe 1,5.

On lira la valeur  $\overline{U}$ , telle que  $\overline{U} = U + \delta U$ .

δU est l'erreur due à la construction de l'appareil; nous négligeons les autres (parallaxe, erreur de zéro, ronflements...).

Les règles de l'U.T.E. précisent que l'erreur commise dans les conditions normales de fonctionnement (conditions de référence) ne doit pas dépasser  $\pm$  1,5 % du calibre, c'est-à-dire ici  $\pm$  0,45 V.

Les constructeurs semblent donc présenter l'incertitude due à la construction des appareils, et décrite par leur classe, comme dotée du niveau de confiance 100 %. Cela ne peut être vrai que pour des appareils qui font l'objet d'un contrôle unitaire.

De plus, on est sans informations sur la forme de la loi de répartition des erreurs telles que  $\delta U$ .

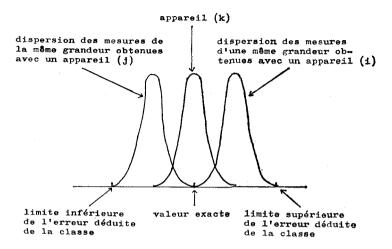
En outre, cette loi évolue avec le vieillissement des appareils, et il faudrait pour la connaître à un instant donné, pouvoir réunir un millier d'appareils du même âge, et du même type et procéder avec chacun d'eux à la mesure d'une même tension ou d'une même intensité.

Des informations sur la manière dont sont ajustés les appareils, jointes aux observations que l'on peut effectuer sur les différentes collections des établissements, permettent toutefois de se faire une opinion sur les lois de répartition de ces erreurs. (Il serait souhaitable d'être informé correctement de la situation actuelle par les fabricants). La réalisation d'un appareil de classe 1,5 nécessite une opération d'ajustage au cours de laquelle un étalonneur agissant sur un paramètre de l'appareil, doit amener l'aiguille d'un indicateur dans une plage correspondant à l'erreur tolérée par la classe de précision de l'appareil (une certaine marge de sécurité, correspondant à la présence, toujours possible, d'une erreur systématique, implique une plage légèrement plus resserrée.)

L'ajustage, par valeurs croissantes ou décroissantes, se fait de manière à amener l'aiguille de l'indicateur de réglage dans les limites de la plage. On ne cherche pas un réglage rigoureux (milieu de la plage), sauf dans le cas d'appareils d'excellente qualité, de classe inférieure ou égale à 0,5, pour ne pas allonger démesurément la durée d'expérimentation qui a une incidence directe sur le prix de revient.

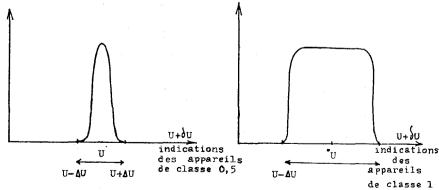
Sauf consigne spéciale, (c'est certainement le cas pour les appareils de précision), l'étalonneur arrête l'opération dès que l'aiguille arrive dans les limites de la plage.

Il résulte de ce qui précède que si l'on réalise, avec un même appareil de classe supérieure ou égale à 1, une série de mesures d'une même grandeur, parfaitement déterminée, et constante pendant la durée des mesures, on trouvera une largeur de dispersion de ces mesures très inférieure à l'incertitude déduite de la classe de l'appareil. Par contre, ces mesures ne seront pas centrées sur la valeur exacte de la grandeur mesurée.



Il est donc certain qu'en étalonnant un tel appareil et en corrigeant ses indications de l'erreur systématique de justesse, on obtient, comme certains auteurs du B.U.P. l'ont remarqué, un appareil plus précis que ce que laisse espérer sa classe.

Si maintenant, muni d'un grand nombre d'appareils identiques, nous procédons, avec chacun d'entre eux, à la mesure de la même grandeur, nous obtiendrons certainement l'une des dispersions représentées ci-dessous, qui seront cette fois toutes deux centrées sur la vraie valeur de la grandeur mesurée.



(ΔU est ici l'incertitude correspondant à la classe de l'appareil).

En conclusion, même pour les appareils de mesures électriques, pour lesquels le constructeur a donné des éléments d'appréciation des erreurs de mesure (valables seulement lorsque ces appareils n'ont pas été altérés par des manipulations intempestives qui font par exemple souffrir les shunts), on est loin de connaître assez de renseignements sur la nature des erreurs δU ou δI, leur fonction de répartition, les niveaux de confiance correspondant aux incertitudes annoncées, etc.

### CONCLUSION

On ne peut prétendre que l'ensemble des résultats indiqués rapidement dans les pages qui précèdent soient nécessaires à une première approche du problème.

Il nous semble toutefois que le paragraphe le plus important, à savoir le § C. II concernant la détermination d'un intervalle de confiance à un niveau de confiance donné peut être présenté en classe sous sa forme la plus simple, c'est-à-dire la méthode de l'étendue.

Le reste est certainement moins important, mais peut-être que ces considérations permettront de répondre à certaines questions qu'on peut se poser, ou que les élèves peuvent soulever.

Peut-être permettront-elles encore de réfléchir sur la validité de certains comportements spontanés, en matière d'élimination de résultats aberrants, de chiffres significatifs conservés, etc.

Je remercie tous ceux, et ils sont nombreux, qui, par les résultats qu'ils m'ont transmis, les discussions que j'ai eues avec eux, les critiques qu'ils ont bien voulu apporter aux premiers manuscrits, ou les renseignements qu'ils m'ont donnés, m'ont permis de mener à bien ce travail, et de l'illustrer par quelques exemples, et notamment M. NININ qui m'a aimablement communiqué la plupart des tables de nombres au hasard dont je dispose ainsi que MM. Dussarat et Prunet.

R. MOREAU (Bordeaux).