REVUE DE STATISTIQUE APPLIQUÉE

A. KOHN

J. ULMO

Quelques considérations statistiques sur la mesure de l'activité de substances radioactives

Revue de statistique appliquée, tome 6, nº 2 (1958), p. 9-55

<http://www.numdam.org/item?id=RSA_1958__6_2_9_0>

QUELQUES CONSIDÉRATIONS STATISTIQUES SUR LA MESURE DE L'ACTIVITÉ DE SUBSTANCES RADIOACTIVES

par

A. KOHN
Chef du Service de Métallographie
Spéciale à l'I. R. S. I. D.

et

J. ULMO
Chef du Service Statistique
à l'I. R. S. I. D.

SOMMAIRE

QUELQUES CONSIDÉRATIONS STATISTIQUES SUR LA MESURE DE L'ACTIVITÉ DE SUBSTANCES RADIOACTIVES

par

A. KOHN

et

J. ULMO

Les techniques radioactives constituent un instrument d'étude particulièrement précieux dans de nombreux domaines de la science pure et appliquée et leur emploi a connu un développement considérable au cours de ces dernières années. Dans la plupart des cas, ces applications nécessitent l'estimation de l'activité d'échantillons placés dans une position déterminée par rapport à un instrument de mesure convenable. Mais la grandeur dont on cherche à déterminer la valeur est d'une nature différente des grandeurs qui interviennent dans les observations habituelles; ces dernières possèdent intrinsèquement une valeur déterminée et l'erreur que l'on peut commettre dans son estimation dépend uniquement de l'imperfection de la méthode de mesure utilisée. Ici, il s'agit d'une grandeur aléatoire, qui ne peut être complètement définie que par sa loi de probabilité, laquelle est une loi de Poisson, et le but de la mesure est de fournir une estimation du paramètre inconnu qui caractérise la distribution de la grandeur étudiée.

Seuls quelques articles étrangers[1], pas toujours facilement accessibles, examinent de façon plus ou moins complète certains des problèmes posés par la détermination de cette grandeur aléatoire. C'est pourquoi il a paru utile de présenter cette étude qui envisage d'un point de vue à la fois théorique et pratique le problème de l'estimation de l'activité d'échantillons radioactifs. Les auteurs espèrent, en effet, que certaines applications des techniques de la statistique à la mesure d'une grandeur physique aléatoire qui obéit

---

(1) Voir par exemple JARRET A.A. - U.S. Atomic Energy Commission Mon P 126 (1946)

RAINWATER L.J. et WU.C.C. Nucleonics 1 (Octobre 1947), p.60

TAYLOR D. La mesure des isotopes radioactifs (Mathew and Co Ltd, London).

à une loi de Poisson, pourront intéresser certains des lecteurs habituels de cette revue; mais ils pensent que les questions traitées ici pourront également retenir l'attention de tous ceux qui sont amenés à se servir de techniques radioactives. C'est à l'intention de ces derniers - qui ne possèdent pas toujours une connaissance approfondie des méthodes statistiques - qu'un premier chapitre est consacré au rappel des principales notions fondamentales employées par la suite, et que le but, les limites et la signification d'un test statistique sont précisés en détail dans un autre chapitre.

La présente étude évoque plusieurs problèmes qui concernent uniquement les mesures d'activité avec un appareil du type "numérateur"; les auteurs ont délibérément laissé de côté les questions qui se posent à propos du temps mort du compteur ou des mesures avec un appareil du type "intégrateur". Certaines des méthodes exposées, comme l'emploi du test de χ^2 par exemple, sont d'une application facile et peuvent rendre divers services; d'autres pourront paraître d'un intérêt pratique plus discutable, mais leur présentation a moins pour but de fournir une recette utile que d'attirer l'attention sur certaines conséquences peu évidentes qui découlent de l'examen correct de la question envisagée. A ce sujet, il ne faut pas oublier que tout examen de la validité ou de l'interprétation d'une mesure d'activité doit être considéré en termes probabilistes; vouloir l'exprimer avec les notions habituelles du cadre déterministe pourrait conduire à des conclusions totalement erronées.

I. - NOTIONS DE CALCUL DES PROBABILITÉS

PRINCIPALES LOIS DE DISTRIBUTION D'UNE VARIABLE ALÉATOIRE

RAPPEL DE QUELQUES NOTIONS ELEMENTAIRES -

Le calcul des probabilités et la statistique sont basés sur la notion de probabilité. Il s'agit là d'un concept dont on se fait facilement une représentation intuitive mais dont il est difficile de donner une définition entièrement satisfaisante. On imagine aisément que la probabilité d'apparition d'un évènement est le rapport théorique du nombre des cas favorables à l'apparition de l'évènement, au nombre total des cas possibles, lorsque ces derniers sont tous également vraisemblables.

On appelle variable aléatoire, une grandeur x susceptible de prendre au cours d'une observation l'une quelconque des valeurs différentes x_1, x_2, x_3, . . . x_n (nous supposons n fini pour la commodité de l'exposé), avec une probabilité déterminée p_1, p_2, p_3, . . . , p_n. On désigne sous le nom de population l'ensemble de toutes les valeurs qui seraient successivement prises par cette variable x, au cours de toutes les observations, en nombre infini, qu'il serait théoriquement possible d'effectuer. En fait, on ne peut réaliser qu'un nombre fini N d'observations qui constitue un échantillon de la population considérée. Soient X_1, X_2, . . . , X_N les valeurs successivement prises par la variable aléatoire x au cours de ces N observations; on appelle fréquence absolue F_k, ou nombre de répétitions de x_k, le nombre de fois que l'on a observé la valeur x_k et fréquence relative le rapport $f_k = \frac{F_k}{N}$.

L'ensemble des valeurs x_k et des fréquences relatives f_k correspondantes définit la loi de distribution de la variable x dans l'échantillon considéré. On définit la moyenne $\overline{x}$ et la variance s^2 de la distribution de la variable x dans cet échantillon par les expressions respectives suivantes :

$$\overline{x} = \frac{\sum_{k=1}^{N} X_k}{N} = \frac{\sum_{k=1}^{n} F_k x_k}{N} = \sum_1^n f_k x_k$$

$$s^2 = \frac{\sum_1^N (X_k - \overline{x})^2}{N} = \frac{\sum_1^n F_k (x_k - \overline{x})^2}{N} = \sum_1^n f_k (x_k - \overline{x})^2$$

De même, la loi de distribution de la variable x dans la population, également appelée loi de probabilité de x, est définie par l'ensemble des valeurs x_1, $x_2, \ldots, x_n$ et des probabilités correspondantes $p_1, p_2, \ldots, p_n$ d'observer les valeurs $x_1, x_2, \ldots, x_n$ dans le cas théorique où le nombre des observations serait infini. On définit encore la moyenne m et la variance σ^2 (la grandeur σ s'appelle l'écart-type) de x dans la population par les expressions :

$$m = \sum_{k=1}^{n} p_k x_k \qquad \sigma^2 = \sum_{k=1}^{n} p_k (x_k - m)^2$$

On démontre (théorème de Bernoulli, plus généralement appelé loi des grands nombres), que la fréquence relative f_k de la valeur x_k, dans l'échantillon de N observations prélevé dans la population considérée, converge en probabilité vers la probabilité p_k dans la population, lorsque le nombre N des observations augmente indéfiniment. On démontre également, que la moyenne et la variance (ou l'écart-type) d'un échantillon convergent en probabilité vers la moyenne et la variance (ou l'écart-type) de la population, lorsque la taille de l'échantillon augmente indéfiniment.

Le calcul des probabilités dont nous allons utiliser quelques applications repose sur la définition et les deux axiomes fondamentaux suivants.

- Indépendance en probabilité : Deux évènements sont dits indépendants en probabilité si la probabilité d'observer l'un d'entre eux est la même, que l'autre évènement se soit produit ou non.

- Axiome des probabilités totales : Si $p_1, p_2, p_3, \ldots, p_k$ sont les probabilités respectives de plusieurs évènements incompatibles $E_1, E_2, E_3, \ldots, E_k$, la probabilité pour que l'un quelconque de ces évènements se produise est égale à la somme $p_1 + p_2 + p_3 + \ldots + p_k$. Comme la somme des probabilités de tous les évènements possibles est égale à 1, il en résulte que si p est la probabilité d'un évènement E_1, la probabilité d'un évènement contraire E_2 est 1 - p.

- Axiome des probabilités composées : Si $p_1, p_2, p_3, \ldots, p_k$ représentent les probabilités respectives de k évènements indépendants en probabilité, la probabilité d'observer simultanément tous ces évènements est égale au produit $p_1 \times p_2 \times p_3 \times \ldots \times p_k$.

DISTRIBUTION BINOMIALE (Problème de Bernoulli) -

Considérons le cas d'une urne contenant un nombre A de boules toutes identiques, mis à part le fait que certaines sont noires et les autres blanches : soit p la proportion inconnue des boules noires et q celle des boules blanches (p + q = 1). Les boules ayant été convenablement mélangées, nous en tirons une au hasard; la probabilité pour que celle-ci soit noire est par définition p, puisque parmi les A cas, tous également possibles, il y en a seulement pA de favorables.

Effectuons une série de n tirages successifs (après chaque tirage, nous remettrons dans l'urne la boule tirée); au cours de ces tirages nous voyons apparaître x fois une boule noire. Si nous recommençons cette série de n tirages, nous observerons une nouvelle valeur du nombre x d'apparitions des boules noires; ce nombre est une variable aléatoire, attachée à la population constituée par l'ensemble, en nombre infini, de toutes les séries de n tirages successifs qu'il serait théoriquement possible d'effectuer dans cette urne, où la proportion des boules noires est p.

On peut calculer quelle est la probabilité P_k de voir apparaître un nombre k de boules noires au cours d'une série de n tirages:

La probabilité de faire sortir une boule noire au cours d'un tirage est p, et celle de faire sortir une boule blanche est $(1 - p) = q$. La probabilité de faire sortir k boules noires, et par conséquent $(n - k)$ boules blanches, dans un ordre déterminé à l'avance, est, en vertu du théorème des probabilités composées, égale au produit des probabilités de faire apparaître individuellement chacune de ces boules, puisque chaque tirage est indépendant des autres. Cette probabilité est donc :

$$p^k (1 - p)^{(n-k)}$$

Il existe plusieurs ordres de succession distincts les uns des autres par lesquels on peut faire apparaître k boules noires et n - k boules blanches au cours de n tirages successifs. Le nombre de ces ordres de succession est égal au nombre des combinaisons suivant lesquelles on peut grouper k objets choisis parmi n d'entre eux; il est égal à :

$$\frac{n!}{k!\,(n - k)!}$$

Puisque nous ne considérons pas l'ordre dans lequel les k boules noires sortent de l'urne, la probabilité d'en voir apparaître k au cours de n tirages, est, en vertu du théorème des probabilités totale :

$$P_k = p^k(1 - p)^{(n-k)} \frac{n!}{k!\,(n - k)!} = p^k q^{(n-k)} \frac{n!}{k!\,(n - k)!}$$

On voit que la valeur de la probabilité P_k est exprimée par le terme en p^k du développement suivant les puissances croissantes de p de l'expression $(p + q)^n$; c'est la raison pour laquelle la loi de répartition de la variable aléatoire x considérée ici est appelée distribution binomiale. On démontre que les valeurs de la moyenne et de la variance de cette distribution sont respectivement :

$$m = np \qquad \sigma^2 = npq$$

DISTRIBUTION DE POISSON -

Nous venons de voir que la loi de probabilité de la distribution binomiale pouvait s'exprimer par la relation :

$$\boxed{P(x) = p^x(1 - p)^{(n-x)} \frac{n!}{x!\,(n - x)!}}$$

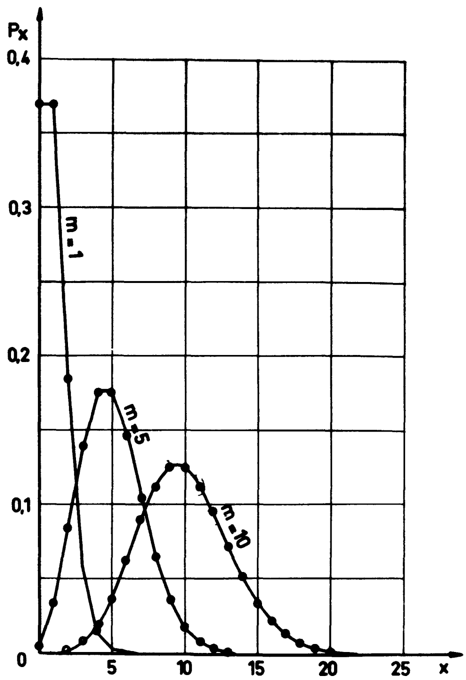

Figure 1
Représentation de lois de Poisson de moyennes 1, 5 et 10.

où x peut prendre toutes les valeurs entières comprises entre 0 et n.

Supposons que la probabilité p soit très petite et que le nombre n soit très grand, le produit de ces deux nombres conservant cependant une valeur finie np = m. On démontre que si p tend vers zéro et n vers l'infini, le produit np demeurànt égal à m, la distribution binomiale tend à la limite vers la distribution de Poisson, dont la loi de probabilité est :

$$\boxed{P(x) = \frac{m^x}{x!} e^{-m}}$$

où x est un nombre entier qui peut prendre toute valeur comprise entre zéro et l'infini. La loi de Poisson est caractérisée par le fait que sa variance σ^2 est égale à sa moyenne m.

On montre que la loi de Poisson jouit de la propriété d'additivité: La somme de plusieurs variables aléatoires, indépendantes en probabilité, qui suivent individuellement une loi de Poisson. obéit également à une loi de Poisson, dont la moyenne est égale à la somme des moyennes de chacune de ces variables.

Par contre, si une variable x suit une loi de Poisson, la variable y = hx, h étant un nombre certain, ne suit pas une loi de Poisson (on a, en effet, m_y = hm mais $\sigma_y^2 = h^2\sigma_x^2 = h^2 m = hm_y$).

DISTRIBUTION DE LAPLACE-GAUSS -

Nous avons admis jusqu'à présent qu'une variable aléatoire ne pouvait prendre que des valeurs discrètes, représentées par la suite des nombres entiers positifs, en quantité finie (cas de la distribution binomiale) ou infinie (cas de la distribution de Poisson).

Mais il existe des variables aléatoires qui sont susceptibles de varier de façon continue et de prendre n'importe quelle valeur positive ou négative. Dans ce cas, on appelle <u>probabilité élémentaire</u> p(x)dx de cette variable la probabilité qu'elle a de prendre une valeur variable comprise dans l'intervalle $(x - \frac{dx}{2}$, $x + \frac{dx}{2})$ et <u>fonction de répartition</u>, la probabilité $F(x) = \int_{-\infty}^{x} p(t)dt$ qu'elle a de prendre une valeur au plus égale à x. La probabilité qu'elle prenne une valeur comprise dans un intervalle (x_1, x_2) quelconque est alors

$$P(x_1, x_2) = \int_{x_1}^{x_2} p(x)dx = F(x_2) - F(x_1)$$

Une variable aléatoire continue, qui intervient dans la représentation de nombreux phénomènes, est celle dont la loi de probabilité est définie par la relation suivante :

$$\boxed{p(x)dx = \frac{1}{\sqrt{2\pi}\,\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx}$$

Cette distribution qui est entièrement définie par les valeurs m et σ de sa moyenne et de son écart-type, est connue sous les noms de distribution de Laplace-Gauss ou distribution normale. Elle est en particulier observée chaque fois qu'une certaine grandeur peut subir, sous l'influence de causes accidentelles, des fluctuations nombreuses, indépendantes les unes des autres, et dont chacune a une influence faible par rapport à leur effet global.

L'équation y = p(x) représentant la "densité de probabilité" de la distribution normale, est figurée par une courbe en cloche, connue sous le nom de courbe en cloche de Gauss (Fig. 2). La courbe de répartition représentative de la fonction de répartition d'une distribution normale est connue sous le nom de courbe en S (Fig. 2).

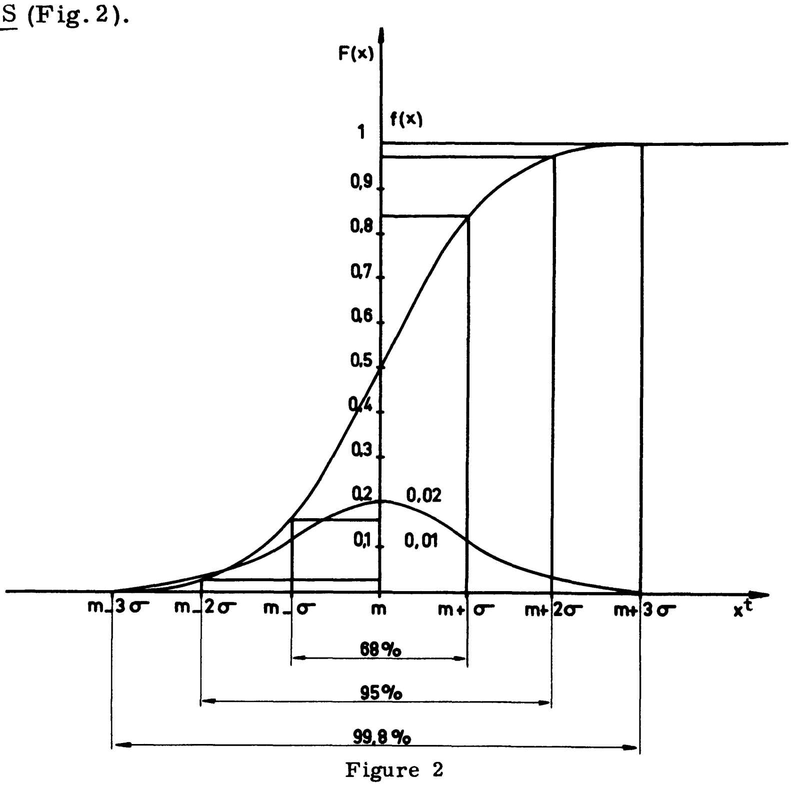

Figure 2

Distribution de Laplace - Gauss

"Courbe en cloche" f(x) et "courbe en S" F(x), intégrale de la précédente avec indications des intervalles contenant 68%, 95% et 99,8% des valeurs de x.

Il existe une double infinité de distributions normales, suivant les valeurs présentées par le couple des grandeurs m et σ, caractéristiques de chacune de ces distributions. On appelle distribution normale réduite la distribution particulière dont la densité de probabilité est :

$$\boxed{y(t) = \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}}}$$

Cette distribution normale est caractérisée par une moyenne nulle et un écart-type égal à l'unité.

On voit que toute distribution de Laplace-Gauss peut être déduite de la distribution normale réduite, par une translation parallèle à l'axe Ox d'amplitude m, suivie d'une affinité suivant l'axe Oy de rapport σ. La probabilité $P(x_1, x_2)$ pour qu'une variable aléatoire x suivant une loi normale caractérisée par les deux paramètres m et σ prenne une valeur comprise dans l'intervalle (x_1, x_2) peut être facilement déterminée au moyen de tables qui donnent les valeurs de la fonction de répartition de la variable normale réduite en fonction de la valeur v prise par cette variable, suivant la relation suivante:

On a en effet : $$F_o(v) = \int_{-\infty}^{v} \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt$$

$$P(x_1, x_2) = F(x_2) - F(x_1) = F_o\left(\frac{x_2 - m}{\sigma}\right) - F_o\left(\frac{x_1 - m}{\sigma}\right)$$

La loi de Laplace-Gauss jouit, comme la loi de Poisson, de la propriété d'additivité. En outre, si la variable x suit une loi de Laplace-Gauss, il en est de même de la loi de $y = hx$ où h est un nombre certain (les paramètres de la loi de y sont $m_y = hm_x$, $\sigma_y = h\sigma_x$).

DISTRIBUTION GAMMA -

C'est la distribution d'une variable continue <u>positive ou nulle</u> dont la densité de probabilité est :

$$p(x) = \frac{e^{-x}\, x^{m-1}}{\Gamma(m)} \qquad 0 \leqslant x \leqslant \infty \qquad m > 0$$

m est un paramètre positif, et $\Gamma(m)$ désigne l'intégrale définie suivante :

$$\Gamma(m) = \int_0^{\infty} e^{-t} t^{m-1} dt$$

$$\Gamma(m) = (m - 1)! \text{ si m est entier.}$$

La fonction de répartition de x est :

$$F(x) = \int_0^{x} \frac{e^{-t}\, t^{m-1} dt}{\Gamma(m)} = \frac{\Gamma_x(m)}{\Gamma(m)} = I(x, m)$$

La fonction $\Gamma_x(m) = \int_0^x e^{-t} t^{m-1} dt$ est appelée fonction Gamma incomplète, d'où le nom de distribution Gamma donné à la distribution de la variable x.

La distribution Γ est, comme la distribution de Poisson, caractérisée par le fait que sa moyenne, qui est égale au paramètre m, est aussi égale à sa variance. Elle jouit, comme la distribution de Poisson de la propriété d'additivité : la somme de plusieurs variables aléatoires indépendantes en probabilité qui suivent individuellement des distributions gamma, suit également une distribution gamma dont la moyenne est égale à la somme des moyennes de chacune de ces variables.

Comme dans le cas de la distribution de Poisson, si x a une distribution Γ, la variable hx, où h est un nombre certain, n'a pas une distribution Γ.

DISTRIBUTION DE χ^2 -

C'est la distribution d'une variable χ^2 qui est liée à une variable x ayant une distribution Γ de paramètre $m = \frac{\nu}{2}$, (où ν est un nombre entier positif) par la relation $\chi^2 = 2x$.

La densité de probabilité de χ^2 est donc :

$$p(\chi^2) = \frac{1}{2^{\frac{\nu}{2}}\,\Gamma(\frac{\nu}{2})}\, e^{-\frac{\chi^2}{2}} (\chi^2)^{\frac{\nu}{2}-1} \qquad 0 \leqslant \chi^2 \leqslant \infty$$

Le paramètre $\nu = 2m$ dont elle dépend est un nombre entier positif appelé nombre de degrés de liberté de la distribution.

La fonction de répartition de χ^2 est :

$$F(\chi^2) = \int_0^{\chi^2} \frac{1}{2^{\frac{\nu}{2}}\,\Gamma(\frac{\nu}{2})}\, e^{-\frac{t}{2}}\, t^{\frac{\nu}{2}-1}\, dt = I\left(\frac{\chi^2}{2}, \frac{\nu}{2}\right)$$

En raison de l'importance pratique de la distribution de χ^2, c'est la fonction de répartition de χ^2 qui figure dans les tables plutôt que celle de la distribution Γ.

De la relation $\chi^2_\nu = 2x_{\frac{\nu}{2}}$ liant χ^2 à ν degrés de liberté à une variable x ayant une distribution Γ, de paramètre $\frac{\nu}{2}$, on déduit immédiatement que :

a) la moyenne de χ^2 est égale à ν et sa variance à 2ν;

b) la distribution de χ^2 jouit comme la distribution Γ de la propriété d'additivité : le nombre de degrés de liberté de la somme de plusieurs variables χ^2 indépendantes en probabilité étant égal à la somme des nombres de degrés de liberté des variables composantes;

c) mais il convient de noter que le produit d'une variable distribuée suivant la loi de χ^2 par un nombre certain h n'est pas distribué suivant la loi de χ^2 (si $h = \frac{1}{2}$, il a une distribution Γ de paramètre $m = \frac{\nu}{2}$).

Le très grand intérêt pratique de la distribution de χ^2 résulte de ce qu'elle possède certaines propriétés particulières, dont nous ne citerons que les suivantes :

1) Le carré d'une variable ayant une distribution de Laplace-Gauss réduite suit une loi de χ^2 avec un degré de liberté. Si donc, x a une distribution de Laplace-Gauss de moyenne m et d'écart-type σ, $\frac{(x-m)^2}{\sigma^2}$ est distribué comme χ^2_1 (χ^2 avec un degré de liberté).

Il résulte de la propriété d'addivité des variables χ^2 que si les variables indépendantes $x_1, x_2, \dots, x_k$ ont des distributions de Laplace-Gauss de moyennes respectives m_i et d'écarts-types σ_i la grandeur $\sum_{i=1}^{k} \frac{(x_i - m_i)^2}{\sigma_i^2}$ suit une loi de χ^2_k (χ^2 avec k degrés de liberté).

En particulier, si $x_1, x_2, \dots, x_k$ constituent un échantillon prélevé dans une population distribué suivant la loi de Laplace-Gauss

$$\sum_{i=1}^{k} \frac{(x_i - m)^2}{\sigma^2} = \chi^2_k$$

On démontre, en outre, que si $\overline{x} = \frac{1}{k} \sum_{i=1}^{k} x_i$ désigne la moyenne des valeurs x_i, $\frac{\sum_{i=1}^{k} (x_i - \overline{x})^2}{\sigma^2}$ est distribuée comme χ^2 avec k - 1 degrés de liberté.

2) Dans le cas où $x_1, x_2, \ldots, x_k$ représente un échantillon extrait d'une distribution de Poisson dont le paramètre m est au moins de l'ordre de 5, la quantité $\sum_{i=1}^{k} \frac{(x_i - \overline{x})^2}{\overline{x}}$ est distribuée comme χ^2 avec k - 1 degrés de liberté(1). Nous exposerons aux chapitres IV et V plusieurs applications pratiques de ce résultat qui présente le grand avantage de ne pas faire intervenir le paramètre m généralement inconnu de la distribution.

ASSIMILATION DE LA LOI DE POISSON A UNE LOI NORMALE -

La loi de Poisson est dissymétrique par rapport à la valeur de la moyenne m, cette dissymétrie étant d'autant moins accusée que la valeur de m est plus élevée (voir figure 1). Lorsque m est élevé (par exemple supérieur à 20), on peut considérer que cette distribution est symétrique autour de la moyenne m, tout au moins pour les valeurs qui ne s'écartent pas trop de cette moyenne. On montre que l'on peut alors assimiler la distribution de Poisson à une distribution normale dont la moyenne est égale à la variance. Cette assimilation n'est pas valable pour les valeurs très éloignées de la moyenne mais comme ces valeurs ont une probabilité extrêmement faible d'être observées, on peut généralement les négliger du point de vue pratique.

A ce sujet, il convient de remarquer que les lois de Poisson et de Laplace-Gauss sont de nature différente : en effet, la première correspond au cas d'une variable discontinue, tandis que la seconde représente le cas d'une variable continue. L'assimilation que nous faisons revient donc à admettre que la probabilité p_k qu'a une variable obéissant à une loi de Poisson de moyenne m de prendre une valeur entière x_k peut s'écrire :

$$p_k = \int_{x_{k-\frac{1}{2}}}^{x_{k+\frac{1}{2}}} \frac{1}{\sqrt{2\pi m}} e^{-\frac{(x-m)^2}{2m^2}} dx$$

Cette valeur peut être facilement calculée à l'aide des tables de la loi normale réduite.

ASSIMILATION D'UNE DISTRIBUTION Γ A UNE DISTRIBUTION NORMALE -

La distribution Γ présente, à valeur égale de la moyenne, une dissymétrie plus accusée que la distribution de Poisson(2). Cette dissymétrie diminue néan-

----

(1) La démonstration de ce résultat repose sur la propriété suivante qui est à la base du test d'ajustement d'une distribution théorique à une distribution observée dit test de χ^2 de Karl PEARSON.

Si x est une variable aléatoire discrète prenant les valeurs $x_1, x_2, \ldots, x_k$ avec les probabilités $p_1, p_2, \ldots, p_k (p_1 + p_2 + \ldots + p_k = 1)$ et si $F_1, F_2, \ldots, F_k$ $(F_1 + F_2 + \ldots + F_k = n)$ désignent les nombres de répétitions des valeurs $x_1, x_2, \ldots, x_k$ dans un échantillon d'effectif n, la quantité $X^2 = \sum_{i=1}^{k} \frac{(F_i - F'_i)^2}{F'_i}$ où $F'_i = np_i$ désigne le nombre de répétitions "théorique" de la valeur x_i, tend vers une distribution de χ^2 avec k - 1 degrès de liberté lorsque l'effectif n de l'échantillon augmente indéfiniment. En fait, cette limite est pratiquement atteinte quand les nombres de répétitions théoriques sont relativement faibles (de l'ordre de 5).

moins quand sa moyenne m augmente et pour des valeurs de m suffisamment élevées (par exemple, supérieures à 100 si on ne s'intéresse pas aux valeurs très éloignées de la moyenne) elle peut, tout comme la loi de Poisson, être assimilée à une distribution normale de moyenne et de variance égales à m.

Il résulte des propriétés de la distribution normale, que si une variable x a une distribution de Poisson ou une distribution Γ de moyenne m assimilable à une loi normale, la variable hx où h est un nombre certain a une distribution assimilable à une distribution normale de moyenne hm et de variance h^2m.

On peut également montrer que si x a une distribution Γ de moyenne m supérieure à 400, la variable $y = \frac{Am}{x}$ où A est un nombre certain a également une distribution assimilable à une distribution normale de moyenne $m_y = A$ et de variance $\sigma_y^2 = \frac{A^2}{m}$ (1).

Note (2) de la page précédente.

(2) Les lecteurs habitués au calcul statistique vérifieront sans peine que les coefficients de dissymétrie γ_1 et d'aplatissement γ_2 de Fisher qui pour une loi normale sont nuls, sont égaux respectivement à $\gamma_1 = \frac{2}{\sqrt{m}}$ et $\gamma_2 = \frac{6}{m}$ tandis que pour la loi de Poisson ils valent $\gamma_1 = \frac{1}{\sqrt{m}}$ et $\gamma_2 = \frac{1}{m}$

(1) A l'intention de ces mêmes lecteurs, on trouvera ci-dessous le calcul des moments de la distribution de y, dont la probabilité élémentaire $p(y) = g(y)dy = + f(x)dx$. On a en effet, pour le moment d'ordre i :

$$m_i = \int_0^\infty y^i g(y)dy = (mA)^i \int_0^\infty \frac{1}{x^i} f(x)dx$$

soit

$$m_i = \frac{(mA)^i}{\Gamma(m)} \int_0^\infty x^{m-1-i} e^{-x}dx = \frac{(mA)^i \Gamma(m-i)}{\Gamma(m)} = \frac{(mA)^i}{(m-1)\ldots(m-i)}$$

d'où l'on déduit :

$$m_1 = \frac{m A}{m-1} \simeq A \qquad \sigma_y^2 = m_2 - m_1^2 = \frac{m^2A^2}{(m-1)^2(m-2)} \simeq \frac{A^2}{m}$$

$$\gamma_1 = 4\sqrt{\frac{m-2}{(m-3)^2}} \simeq \frac{4}{\sqrt{m}} \quad \text{et} \quad \gamma_2 = \frac{6(5m-11)}{(m-3)(m-4)} \simeq \frac{30}{m}$$

II .- LE PHÉNOMÈNE DE LA RADIOACTIVITÉ EXAMINÉ DU POINT DE VUE STATISTIQUE

A. MESURES EN PRÉTEMPS (DURÉE DÉTERMINÉE)

NATURE ALEATOIRE DU RESULTAT D'UN COMPTAGE EN PRETEMPS -

Une substance radioactive est constituée d'atomes susceptibles de subir spontanément, à un moment ou à un autre, une transformation nucléaire qui se traduit par l'émission d'un rayonnement que l'on peut en général déceler. Expérimentalement, on compte le nombre dN des atomes qui se désintègrent au cours d'un intervalle de temps petit $d\theta$. On constate que la variation du nombre des atomes transformés au cours de l'unité de temps se fait en fonction du temps θ suivant une loi exponentielle de la forme :

$$\frac{dN}{d\theta} = - k\, e^{-\lambda\theta}$$

En calculant l'intégrale de cette expression, on obtient le nombre N des atomes non désintégrés présents à l'instant θ :

$$N = - k\int_{\theta}^{\infty} e^{-\lambda t}\, dt = \frac{k}{\lambda}\, e^{-\lambda\theta}$$

La valeur de la constante k, déterminée expérimentalement, peut être exprimée en fonction du nombre N_0 des atomes présents à un instant θ_0 , choisi comme instant origine ($\theta_0 = 0$). A ce moment, en effet : $N_0 = \frac{k}{\lambda}$. On en tire :

$$N = N_0\, e^{-\lambda\theta} \quad \text{et} \quad dN = - N\lambda d\theta$$

Cette dernière expression montre que le nombre dN des atomes qui se désintègrent au cours d'un intervalle de temps très petit $d\theta$, à un instant θ quelconque, est proportionnel au nombre N des atomes radioactifs présents à ce moment. Ce résultat indique que tous les atomes radioactifs présents ont la même probabilité de se désintégrer au cours de l'intervalle de temps considéré; cette probabilité est égale à $\lambda d\theta$. La constante λ représente la probabilité qu'a un atome quelconque de la substance radioactive de se désintégrer au cours d'un intervalle de temps égal à l'unité. La quantité $A = N\lambda$ représente l'activité de cette substance à l'instant θ.

Considérons une substance radioactive comportant N atomes radioactifs à l'instant considéré. On peut se proposer de calculer la probabilité P_n d'observer n désintégrations au sein de cette substance au cours d'un intervalle de temps t très petit par rapport à la période de décroissance de cet élément(1) (ce qui revient à admettre que le nombre N demeure pratiquement inchangé au cours de l'intervalle de temps t). Si on désigne par $p = \lambda t$ la probabilité qu'a un atome de se désintégrer pendant cet intervalle de temps, on peut appliquer ici un raisonnement similaire à celui qui a été utilisé pour établir la loi de probabilité d'une distribution binomiale.

En supposant que l'on puisse suivre individuellement le comportement d'un atome particulier, la probabilité pour que n atomes particuliers se désintègrent, les N - n autres atomes ne subissant pas de transformation, est $p^n(1-p)^{N-n}$. La probabilité pour que n quelconques des N atomes présents se désintègrent est donnée par l'expression de la loi binomiale :

$$P_n = p^n (1 - p)^{N-n} \frac{N!}{n!\,(N - n)!}$$

Mais puisque par hypothèse, nous avons supposé que le nombre des atomes désintégrés au cours de la durée d'observation t est très petit par rapport au nombre d'atomes présents, p est très petit par rapport à N; nous sommes donc dans le cas où la loi binomiale se confond avec la loi de Poisson et nous pouvons écrire :

$$P_n = \frac{m^n}{n!} e^{-m}$$

où m, qui désigne la valeur moyenne du nombre d'atomes désintégrés pendant la durée t de l'observation, est égal au produit $N\lambda t = At$.

Nous venons donc de démontrer que le nombre des atomes d'une substance radioactive qui se désintègrent au cours d'un intervalle de temps très petit par rapport à la période radioactive de cet élément, est une grandeur aléatoire qui obéit à une loi de Poisson.

D'un point de vue pratique, on ne considère pas l'ensemble des atomes qui se désintègrent, mais seulement ceux d'entre eux qui le font en émettant leur rayonnement dans une direction favorable pour permettre la formation d'une impulsion dans le compteur (nous négligeons l'influence du temps mort). Nous pouvons reprendre le raisonnement précédent en écrivant que tout atome qui se désintègre a une probabilité w (identique pour tous les atomes) d'émettre son rayonnement dans la direction favorable. En raison de l'axiome des probabilités composées, la probabilité élémentaire d'un atome donné de se désintégrer pendant l'intervalle de temps t, en provoquant le déclenchement d'une impulsion dans l'appareil de mesure est pw. La démonstration précédente s'applique complètement en considérant la probabilité "élémentaire composée" pw. On voit donc que le nombre de coups enregistrés par un compteur, pendant une durée déterminée, est une variable aléatoire qui suit une loi de Poisson; le nombre de coups enregistrés au cours d'une mesure est, en général, suffisamment grand pour que l'on puisse assimiler sa loi de probabilité à une loi normale.

Comme les mesures d'activité au compteur n'ont pas en général pour but de fournir des valeurs absolues mais des valeurs relatives, les échantillons étant

---

(Note (1) de la page précédente).

(1) Rappelons que la période d'un radioélément représente l'intervalle de temps au cours duquel le nombre des atomes initialement présents a diminué de moitié ou, ce qui revient au même, l'intervalle de temps au cours duquel l'activité diminue de moitié. Il s'ensuit que la relation entre la constante radioactive λ et la période T, exprimées avec la même unité de temps, est :

$$\lambda T = \text{Log } 2 = 0,693$$

disposés dans les mêmes conditions géométriques par rapport au compteur, nous conviendrons d'appeler, dans la suite de cet exposé, activité d'un échantillon, la grandeur physique inconnue A représentée par le produit $N\lambda w$ (N étant le nombre d'atomes radioactifs présents à l'instant de la mesure, λ la constante radioactive de l'élément considéré, et w un coefficient inférieur à 1 qui tient compte des conditions géométriques de la mesure).

En conclusion, on peut dire que le nombre de désintégrations comptées par un appareil de mesure sur un échantillon radioactif, au cours d'une mesure de durée déterminée t, est une variable aléatoire qui suit une loi normale dont la moyenne est égale à la variance. Sa loi de probabilité est représentée par l'équation :

$$\boxed{p(n) = \frac{1}{\sqrt{2\pi m}} e^{-\frac{(n-m)^2}{2m}}} \qquad \text{où } m = At$$

LOI DE PROBABILITE DES RESULTATS DE MESURES D'ACTIVITE EFFECTUEES EN PRETEMPS (durée déterminée) -

Nous venons de voir que la grandeur qui représente le nombre de désintégrations comptées au cours d'un intervalle de temps déterminé (mesure en prétemps) n'a pas une valeur bien déterminée, comme la plupart des grandeurs physiques usuelles, mais qu'elle est une variable aléatoire. Ceci signifie qu'indépendamment des erreurs qui pourraient provenir de l'imperfection de la méthode de mesure, toute détermination d'activité comporte une certaine incertitude due à la nature aléatoire même du phénomène sur lequel porte la mesure.

Si nous considérons une série de k mesures successives effectuées sur un même échantillon radioactif, pendant des durées égales t (la durée totale de ces mesures étant suffisamment petite par rapport à la période du radioélément pour que l'on puisse négliger l'influence de la décroissance de son activité), l'ensemble des valeurs $n_1, n_2, \ldots, n_k$ observées au cours de ces mesures constitue un échantillon de la population totale des valeurs qui peuvent représenter le nombre de désintégrations observables au cours d'un intervalle de temps t. On ne peut évidemment pas connaître exactement la valeur m qui correspond à la moyenne de cette population et qui définirait la valeur la plus probable du nombre de désintégrations. Mais on démontre que la meilleure estimation que l'on puisse faire de cette valeur est obtenue en calculant la moyenne $\overline{n}$ de l'ensemble des valeurs mesurées; cette valeur $\overline{n}$ représente également la meilleure estimation que l'on puisse faire de la variance.

En fait, la valeur que l'on considère n'est pas la valeur mesurée n mais la valeur $a_t = \frac{n}{t}$ rapportée à l'unité de temps (minute), qui représente l'activité inconnue A de l'échantillon étudié. Dans ce cas, si on effectue plusieurs mesures successives de même durée t, les valeurs $a_1 = \frac{n_1}{t}$ $a_2 = \frac{n_2}{t}$ ... $a_k = \frac{n_k}{t}$ de la variable aléatoire a_t se répartissent autour de la valeur moyenne $A = \frac{m}{t}$ suivant une loi de Gauss dont la variance est égale à $\sigma^2_{a_t} = \frac{m}{t^2} = \frac{A}{t}$ (Cf. chapitre I).

La meilleure estimation que l'on puisse faire de la valeur A est donc la valeur moyenne $\overline{a}_t = \frac{a_1 + a_2 + \ldots + a_k}{k} = \frac{\overline{n}}{kt}$ de l'ensemble des valeurs a_1, a_2. ..., a_k.

Les grandeurs $\bar{n}$ et $\bar{a}_t$ qui représentent les valeurs moyennes des résultats de mesures ou des activités déduites d'une série de k mesures de même durée t sont elles-mêmes des variables aléatoires dont les lois de probabilité sont également des lois de Laplace-Gauss. Les paramètres de ces lois sont :

moyenne de $\bar{n}$: $m = At$ moyenne de $\bar{a}_t$: $\frac{m}{t} = A$

variance de $\bar{n}$: $\frac{m}{k} = \frac{At}{k}$ variance de a_t : $\frac{m}{kt^2} = \frac{A}{kt}$

B. MESURES EN "PRÉCOMPTE"

(nombre de désintégrations fixé)

Nous avons jusqu'à présent considéré le cas où les mesures d'activité avaient une durée déterminée t (mesures en prétemps), ce qui est naturel dans le cas où la valeur de la grandeur mesurée est proportionnelle à la durée de la mesure. Quand on opère ainsi on déduit de la valeur observée de la variable aléatoire n (nombre de désintégrations pendant une durée t fixée) la valeur de la variable aléatoire $a_t = \frac{n}{t}$ dont nous venons d'indiquer la loi de probabilité.

En raison de certaines particularités attachées à la nature aléatoire du phénomène étudié que nous verrons plus loin, on est souvent amené à effectuer les mesures d'activité en comptant le temps nécessaire pour observer un nombre déterminé de désintégrations (mesures en précompte). Dans ce cas, la variable aléatoire n'est plus n mais la durée t de laquelle on déduit la valeur $a_n = \frac{n}{t}$. La grandeur a_n est donc une variable aléatoire différente de la variable aléatoire a_t.

Pour déterminer la loi de probabilité de a_n, il convient d'étudier la loi de probabilité de la variable aléatoire t temps nécessaire pour compter n désintégrations.

LOI DE PROBABILITE DE L'INTERVALLE DE TEMPS t AU BOUT DUQUEL SURVIENT LA n^e DESINTEGRATION (n étant petit par rapport au nombre N des atomes présents à l'instant initial).

On a vu que la probabilité qu'a un atome radioactif déterminé de se désintégrer pendant un petit intervalle de temps dt est λdt.

Si N_t est le nombre des atomes présents au début de cet intervalle, la probabilité d'observer une désintégration pendant cet intervalle est $N_t \lambda dt = Adt$, si nous posons $N_t \lambda = A$.

Si dt est choisi très petit par rapport à la valeur moyenne $\frac{1}{A - 1}$ de l'intervalle de temps séparant deux désintégrations successives, on conçoit que la probabilité $(Adt)^2$ d'observer deux désintégrations dans l'intervalle de temps dt est très faible (c'est un infiniment petit d'ordre 2 par rapport à dt). On peut donc considérer que pendant l'intervalle dt, il ne peut pas y avoir plus d'une désintégration.

Supposons que nous ayons observé n désintégrations, la n^e étant survenue dans l'intervalle de temps (t - dt, t). Comme il ne peut y avoir plus d'une désintégration pendant cet intervalle, c'est que les n - 1 premières ont eu lieu dans l'intervalle (0, t - dt).

La probabilité d'observer la n^e désintégration dans l'intervalle $(t - dt, t)$ est donc :

$f_n(t)dt$ = prob. de $(n - 1)$ désintégrations dans $(0, t - dt)$
x prob. d'une désintégration dans $(t - dt, t)$

La probabilité d'une désintégration dans $(t - dt, t)$ est, on l'a vu, Adt.

Pour calculer la probabilité d'observer $n - 1$ désintégrations dans l'intervalle $(0, t - dt)$, nous allons poser $t - dt = k\Delta t$, c'est-à-dire, considérer l'intervalle $(0, t - dt)$ comme la somme de k intervalles de durée Δt dont chacun a la probabilité $A\Delta t$ de correspondre à une désintégration (et $1 - A\Delta t$ de ne correspondre à aucune désintégration). La probabilité d'observer $n - 1$ désintégrations est alors égale à la probabilité que sur les k intervalles $n - 1$ correspondent à une désintégration. C'est donc :

$$C_K^{n-1} (A\Delta t)^{n-1} [1 - A\Delta t]^{K-(n-1)}$$

On sait que si Δt tend vers 0, le produit $k\Delta t$ restant constant et égal à $t - dt \simeq t$, cette expression tend vers celle de la loi de probabilité d'une variable de Poisson de paramètre $Ak\Delta t = At$, c'est-à-dire vers :

$$e^{-At} \frac{(At)^{n-1}}{(n-1)!}$$

On a donc :

$$\boxed{f_n(t)dt = e^{-At} \frac{(At)^{n-1}}{(n-1)!} Adt}$$

où $f_n(t)$ représente la densité de probabilité de la variable t.

On voit que la variable $At = x$ dont la probabilité élémentaire est : $p(x)dx = \frac{e^{-x} x^{n-1}}{(n-1)!} dx$ suit une distribution gamma de moyenne n.

Si n est au moins de l'ordre de mille, ce qui est en pratique, le cas, cette distribution est assimilable à une distribution de Laplace-Gauss de moyenne et de variance égales à n. Il s'ensuit que l'intervalle de temps t nécessaire pour compter n désintégrations et la variable $a_n = \frac{n}{t}$ sont des variables aléatoires, dont les distributions peuvent être assimilées à des distributions de Laplace-Gauss dont les paramètres sont :

$$\begin{cases} \text{moyenne de } t : \dfrac{n}{A} \\ \text{variance de } t : \dfrac{n}{A^2} \end{cases} \qquad \begin{cases} \text{moyenne de } a_n : A \\ \text{variance de } a_n : \dfrac{A^2}{n} \end{cases}$$

LOI DE PROBABILITE DES RESULTATS DE MESURES D'ACTIVITE EFFECTUEES EN PRECOMPTE (nombre de désintégrations déterminé).

Il résulte de ce qui précède que si nous considérons une série de k mesures successives effectuées en précompte sur un même échantillon radioactif (le précompte n étant tel que kn soit petit devant le nombre total N des atomes présents), l'ensemble des valeurs $a_1 = \frac{n}{t_1}$, $a_2 = \frac{n}{t_2}$ ... $a_k = \frac{n}{t_k}$ de la variable a_n déduites des résultats de ces mesures se répartit autour de la valeur inconnue A suivant une loi de Laplace-Gauss de variance égale à $\frac{A^2}{n}$.

La meilleure estimation que l'on puisse faire de la valeur A est donc la valeur moyenne :

$$\overline{a}_n = \frac{n}{k}\left(\frac{1}{t_1} + \frac{1}{t_2} + \ldots + \frac{1}{t_k}\right)$$ de l'ensemble des valeurs a_1, $a_2, \ldots, a_k$.

La grandeur $\overline{a}_n$ est elle-même une variable aléatoire dont la loi de probabilité est une loi de Laplace-Gauss de moyenne A et de variance $\frac{A^2}{nk}$.

III. - LES ERREURS ALÉATOIRES DANS LES MESURES D'ACTIVITÉ

A. ERREUR SUR CETTE MESURE UNIQUE D'ACTIVITÉ

CONSIDERATIONS SUR LA SIGNIFICATION DU RESULTAT D'UNE MESURE UNIQUE D'ACTIVITE -

La mesure de l'activité d'un échantillon radioactif placé dans des conditions géométriques déterminées, par rapport à un certain appareillage de mesure, consiste à obtenir une appréciation aussi convenable que possible de la valeur A de son activité, d'après le nombre n de désintégrations comptées par l'appareil de mesure au cours d'un intervalle de temps de durée t minutes; le résultat de cette mesure est exprimé par le rapport $a = \frac{n}{t}$. Nous avons vu dans le chapitre précédent que les résultats de plusieurs mesures successives ne peuvent pas être identiques même si on fait abstraction des causes d'erreur liées à l'imperfection de l'appareil de mesure, car la grandeur mesurée est une variable aléatoire. Il est par conséquent impossible de connaître exactement la valeur de la grandeur inconnue A, mais nous avons vu au chapitre II, comment on pouvait l'estimer. Pour connaître l'incertitude attachée à cette estimation, il faut connaître la loi de probabilité de la variable aléatoire a.

On sait que cette loi de probabilité peut être assimilée à une loi normale dont la moyenne m est égale à la valeur inconnue A de l'activité de l'échantillon, et dont l'écart-type σ est également une fonction connue de A. Il peut, par suite, être estimé à partir de la valeur estimée de A.

Il suffit donc d'une seule mesure pour que l'on puisse estimer à la fois les deux paramètres A et σ_a qui définissent complètement la distribution normale de a. Ainsi le résultat d'une seule mesure permet d'estimer non seulement la valeur de A, mais aussi de caractériser l'incertitude attachée à cette estimation.

Pour effectuer cette mesure, on peut se fixer à l'avance la valeur de t (mesure en prétemps) : soit t_0 cette valeur. Dans ce cas, le nombre n des désintégrations est une variable aléatoire qui suit une loi de Poisson, et la valeur mesurée a_t constitue un échantillon de la population de toutes les mesures de même durée t_0 , en quantité infinie, qu'il serait théoriquement possible d'effectuer sur l'échantillon d'activité inconnue A. Les estimations des paramètres de cette population sont :

$$m'_{a_t} = a_t = \frac{n}{t_o} \qquad \sigma'_{a_t} = \sqrt{\frac{a_t}{t_o}}$$

Mais on peut aussi se fixer la valeur n à l'avance (mesure en précompte) : soit n_o cette valeur. Nous avons vu que dans ce cas, la grandeur At est une variable aléatoire dont la loi de probabilité est la distribution Γ : la valeur mesurée a_n constitue un échantillon de la population de toutes les mesures en quantité infinie, portant sur un même nombre n_o de désintégrations, qu'il serait théoriquement possible d'effectuer sur l'échantillon d'activité inconnue A. Les paramètres estimés de cette population sont :

$$m'_{a_n} = a_n = \frac{n_o}{t} \qquad \sigma'_{a_n} = \sqrt{\frac{a_n^2}{n_o}}$$

On peut également concevoir qu'on commence la mesure, sans se fixer a priori ni sa durée, ni le nombre d'impulsions n; on arrête la mesure au bout d'un temps quelconque t_1, alors que n_1 désintégrations ont été comptées. On peut considérer le résultat de la mesure $a_1 = \frac{n_1}{t_1}$, soit comme un échantillon de la population des mesures effectuées à temps constant $t = t_1$, soit comme un échantillon de la population des mesures effectuées à nombre de désintégrations constant $n = n_1$.

Cette valeur mesurée a_1 fournit une estimation de la valeur inconnue A avec une certaine "incertitude". Nous préciserons par la suite la signification de cette incertitude et nous verrons qu'elle est caractérisée par la valeur de l'écart-type σ_a. Cette incertitude doit être la même que nous considérions la valeur a comme appartenant à la population (a_t, σ_{a_t}) des mesures effectuées à temps constant, ou à la population (a_n, σ_{a_n}) des mesures effectuées à nombre de coups constant.

Il en résulte que l'estimation σ'_a que l'on fait de σ_a , d'après le résultat d'une mesure unique, doit être la même que l'on considère comme variable aléatoire la grandeur discontinue n ou la grandeur continue t. On doit donc avoir $\sigma'_{a_t} = \sigma'_{a_n}$. Il en est bien ainsi puisque l'on a :

$$\sigma'_{a_t} = \sqrt{\frac{a_1}{t_1}} \qquad \text{et} \qquad \sigma'_{a_n} = \sqrt{\frac{a_1^2}{n_1}}$$

$$\text{avec} \quad a_1 = \frac{n_1}{t_1}$$

En conclusion : en raison de la nature même du processus de désintégration radioactive, le résultat a_1 d'une mesure unique de durée t_1 au cours de laquelle on a compté n_1 désintégrations, permet à la fois d'estimer la valeur inconnue A de l'activité étudiée, et de caractériser l'incertitude qui résulte de cette estimation.

Comme cette mesure ne préjuge nullement des conditions dans lesquelles on pourrait la recommencer pour avoir d'autres valeurs de la variable aléatoire a, toutes les considérations qui peuvent être faites au sujet de l'incertitude de cette mesure aboutissent aux mêmes conclusions, que cette mesure soit effectuée en prétemps (temps fixé à l'avance) ou en précompte (nombre de désintégrations fixé à l'avance). Cette mesure unique permet de déterminer une valeur particulière de la variable aléatoire $a = \frac{n}{t}$, d'après l'observation d'un couple particulier de valeurs des deux grandeurs n et t, sans préciser laquelle d'entre elles est considérée comme variable aléatoire.

Ce choix ne devra être fait que si l'on désire connaître de nouvelles valeurs de la variable a, afin d'avoir une meilleure estimation de sa loi de probabilité, car les divers échantillons que constituent ces nouvelles valeurs doivent appartenir à une même population; il convient alors de définir celle-ci de façon plus précise. Mais si l'on effectue une mesure unique, il suffit de considérer que la valeur mesurée $a_1 = \frac{n_1}{t_1}$ est un échantillon particulier d'une population normale dont les paramètres estimés sont :

$$\boxed{m'_a = a_1 = \frac{n_1}{t_1}} \qquad \boxed{\sigma'^2_a = \frac{a_1}{t_1} = \frac{a_1^2}{n_1} = \frac{n_1}{t_1^2}}$$

DEFINITION DE L'ERREUR SUR UNE MESURE D'ACTIVITE -

Nous venons de voir que l'on peut considérer le résultat d'une mesure unique d'activité comme un échantillon d'une population normale dont les paramètres inconnus A et σ_a peuvent être estimés en fonction des valeurs n_1 et t_1 qui caractérisent cette mesure, l'une d'entre elles définissant les conditions de la mesure et l'autre représentant une valeur particulière prise par une certaine variable aléatoire. Au chapitre II, nous avons précisé que la valeur moyenne $\bar{x}$ d'un échantillon de k mesures tiré d'une population normale de paramètres m et σ n'est pas égale à la valeur moyenne m de cette population, et que la quantité $\bar{x}$ est elle-même une variable aléatoire suivant une distribution normale de moyenne m et d'écart-type $\sigma_{\bar{x}} = \frac{\sigma}{\sqrt{k}}$. Les tables de la fonction de répartition de la loi normale réduite permettent de calculer quelle est la probabilité P(h) pour que la quantité $\frac{|\bar{x} - m|}{\sigma_{\bar{x}}}$ soit inférieure à une valeur h fixée. Le tableau suivant reproduit quelques valeurs particulières tirées de ces tables :

h	0,674	1,00	$\sqrt{2}$	1,645	$1,96 \simeq 2$	2,576	3,09	3,891	4,417
P(h) en %	50	68	84	90	95	99	99,8	99,99	99,999

Dans le cas d'une mesure unique d'activité, on peut ainsi calculer quelle est la probabilité P(h) pour que la quantité $\frac{|a_1 - A|}{\sigma_a}$ soit inférieure à h ou, d'une façon plus pratique, estimer d'après cette table quelle est la probabilité P(h) qu'a l'écart $|a_1 - A|$, entre la valeur mesurée $a_1 = \frac{n_1}{t_1}$ et la valeur inconnue de l'activité A, d'être inférieur à $h\sigma_a \simeq h \frac{\sqrt{n_1}}{t_1}$. C'est cette probabilité P(h) qui définit ce que, dans un paragraphe précédent, nous avons appelé "incertitude" sur le résultat d'une mesure.

Pour reprendre un langage plus usuel, mais en précisant bien que l'on ne considère pas ici les causes d'erreur liées au dispositif de mesure mais seulement l'incertitude qui tient à la nature aléatoire de la grandeur mesurée, nous conviendrons d'appeler "erreur absolue à P%" la valeur e_p de la moitié de l'intervalle centré sur la valeur observée a qui a P chances sur 100 de contenir la valeur inconnue A et "erreur relative à P%" le quotient $\varepsilon_P = \frac{e_P}{A}$. L'examen du tableau précédent montre que si l'on écrit P = P(h), on a $e_P = h\sigma_a$.

Le choix de la valeur P = P(h), qui fixe à la fois la probabilité pour que l'écart entre la valeur théorique A et la valeur mesurée a soit compris à l'intérieur d'un certain intervalle, et l'étendue $h\sigma \simeq h\frac{\sqrt{n}}{t}$ de cet intervalle, dépend de considérations diverses déterminées par le problème étudié.

L'examen du tableau montre que si l'on désire être quasi certain que l'intervalle $a \pm e_P$ contient la valeur A, c'est-à-dire, si l'on se fixe une valeur de P très grande (99,99 ou 99,999%), l'intervalle e_P est très grand (h > 3). Dans le cas de certaines variables suivant une loi normale dont l'écart-type est petit par rapport à la moyenne, les statisticiens considèrent souvent l'erreur à 95% ($h \simeq 2$) ou à 99% (h = 2, 57). Dans le cas des mesures d'activité, de nombreux laboratoires considèrent l'erreur à 90% (h = 1, 64) et certains d'entre eux l'erreur à 84% ($h = \sqrt{2}$). Lorsqu'il s'agit de calculs théoriques, on considère plutôt "l'erreur-type" e correspondant à h = 1; cette erreur-type, qui est égale à l'écart-type σ_a, est celle que nous considérons dans la suite de cet exposé.

Le tableau précédent permet d'en déduire l'erreur à P%, par la formule $e_P = he$.

Si a est le résultat d'une mesure de durée t au cours de laquelle on a compté n désintégrations, les estimations de l'erreur-type absolue e_a et de l'erreur-type relative ε_a sur les résultats de cette mesure sont :

$$\boxed{e'_a = \sigma_a = \sqrt{\frac{n}{t^2}}} \qquad \boxed{\varepsilon'_a = \frac{\sigma_a}{a} = \frac{1}{\sqrt{n}}}$$

B. QUELQUES PROBLÈMES QUI PEUVENT SE POSER AU SUJET DES ERREURS SUR UN ENSEMBLE DE MESURES D'ACTIVITÉ

BASES DES CALCULS D'ERREURS SUR LES MESURES D'ACTIVITE -

Les mesures d'activité portent, en général, sur plusieurs échantillons, et l'on désire estimer une fonction simple (différence ou rapport en général) des activités inconnues des échantillons étudiés.

Le problème, qui se pose alors, est de déterminer l'erreur avec laquelle cette même fonction des résultats des mesures, représente la valeur cherchée.

Ces calculs d'erreurs sont basés sur les propriétés suivantes :

1) La somme ou la différence de deux variables aléatoires indépendantes en probabilité, qui suivent une loi normale, est une variable aléatoire suivant une loi normale de moyenne égale à la somme ou à la différence des moyennes de chacune des variables considérées, et de variance égale à la somme des variances de chacune de ces variables.

$$m_{a+b} = m_a + m_b \qquad \sigma^2_{a+b} = \sigma^2_a + \sigma^2_b$$

$$m_{a-b} = m_a - m_b \qquad \sigma^2_{a-b} = \sigma^2_a + \sigma^2_b$$

Il en résulte que les erreurs absolues sur la somme ou la différence des moyennes de deux variables aléatoires indépendantes suivant des lois normales sont :

$$e_{a+b} = \sqrt{e_a^2 + e_b^2} \qquad e_{a-b} = \sqrt{e_a^2 + e_b^2}$$

2) On peut montrer que le carré de l'erreur relative sur le produit ou le quotient de 2 variables aléatoires indépendantes en probabilité et suivant des lois normales est sensiblement égale à la somme des carrés des erreurs relatives de chacune des variables considérées

$$\boxed{\varepsilon_{ab}^2 \simeq \varepsilon_a^2 + \varepsilon_b^2} \qquad \boxed{\varepsilon_{a/b}^2 \simeq \varepsilon_a^2 + \varepsilon_b^2}$$

Les erreurs qui résultent de l'association de résultats de mesure d'activité indépendants sont calculées en portant dans les expressions précédentes les valeurs de l'erreur absolue ou de l'erreur relative indiquées au paragraphe précédent :

Erreur absolue sur une somme ou une différence :

$$e_{a+b} = e_{a-b} \simeq \sqrt{\frac{n_a}{t_a^2} + \frac{n_b}{t_b^2}} = \sqrt{\frac{a}{t_a} + \frac{b}{t_b}}$$

Erreur relative sur un produit ou un quotient :

$$\varepsilon_{ab} = \varepsilon_{a/b} \simeq \sqrt{\frac{1}{n_a} + \frac{1}{n_b}}$$

où a et b représentent les valeurs des activités, mesurées en comptant respectivement n_a et n_b désintégrations, pendant des durées t_a et t_b.

APPLICATION DU CALCUL D'ERREURS A L'ETUDE DE QUELQUES PROBLEMES -

Les résultats obtenus dans les paragraphes précédents permettent de faire un certain nombre de remarques :

1) <u>Erreur sur la valeur moyenne de k mesures effectuées dans des conditions identiques.</u>

Nous avons vu, au chapitre précédent, que l'écart-type de la loi de probabilité de la valeur moyenne $\bar{a}$ d'une activité, déduite du résultat de k mesures de même durée t, a pour expression : $\sigma_{\bar{a}} = \sqrt{\frac{A}{kt}}$. Au lieu de compter séparément les nombres n_i de désintégrations au cours de k mesures successives, on aurait pu compter globalement le nombre total des désintégration $N = \sum_{i=1}^{k} n_i = t \sum_{i=1}^{k} a_i$ au cours d'une mesure unique de durée T = kt. On aurait obtenu pour estimation de l'activité $\frac{N}{T} = \frac{t\Sigma a_i}{kt} = \bar{a}$, et pour erreur - type attachée à cette estimation $\sqrt{\frac{A}{T}} = \sqrt{\frac{A}{kt}}$. On retrouve l'expression ci-dessus.

<u>L'erreur sur la valeur moyenne des résultats de k mesures de même durée t est égale à l'erreur sur une mesure unique de durée totale T = kt.</u>

----

(1) Cette formule n'est valable que si ε_b n'est pas trop grand (s'il est inférieur à 1/5).

Dans le cas où les mesures sont effectuées en précompte (nombre n fixé à l'avance), comme nous l'avons vu au chapitre précédent, la moyenne $\bar{a}$ des résultats de k mesures de même précompte suit une loi de Laplace-Gauss de moyenne A et d'écart-type $\sigma_{\bar{a}} = \sqrt{\frac{A^2}{nk}}$.

Cette expression représente aussi l'écart-type du résultat d'une mesure unique portant sur N = kn désintégrations. On en déduit :

L'erreur sur la valeur moyenne des résultats de k mesures effectuées avec un même précompte n est égale à l'erreur sur une mesure unique portant sur un précompte N = kn de désintégrations.

2) Mesure des activités de plusieurs échantillons avec une erreur absolue constante.

Il est parfois utile de déterminer les activités de plusieurs échantillons avec une même erreur absolue. C'est par exemple le cas lorsque la valeur de l'activité de chaque échantillon dépendant de la valeur d'une certaine grandeur qui caractérise cet échantillon, on désire tracer la courbe reliant les variations de l'activité aux variations de cette grandeur, dont les valeurs sont connues de façon plus précise que les valeurs de l'activité ; l'erreur sur la position exacte de chaque point figuratif sera alors la même.

Comme les expressions de l'erreur absolue sont :

$$e^2 = \frac{A}{t} \qquad \text{mesures en prétemps}$$

$$e^2 = \frac{A^2}{n} \qquad \text{mesures en précompte,}$$

on voit que la valeur de l'erreur absolue demeure constante dans les mesures en prétemps si la durée de la mesure est proportionnelle à la valeur de l'activité mesurée et dans les mesures en précompte , si le nombre de coups compté est proportionnel au carré de l'activité mesurée.

3) Mesure de l'activité de plusieurs échantillons avec une erreur relative constante.

Dans de nombreux autres cas, il est préférable de mesurer l'activité de divers échantillons avec une même erreur relative. L'expression de l'estimation de l'erreur relative pour une mesure unique $\varepsilon'^2 = \frac{1}{n}$ montre que l'estimation de l'erreur relative est indépendante de la durée de la mesure. Il faut donc effectuer toutes les mesures en comptant un même nombre de désintégrations si l'on désire avoir une même erreur relative(1).

C'est en raison de cette propriété particulière que la mesure en précompte est souvent utilisée.

4) Erreur sur la mesure d'une activité dont on a déduit le mouvement propre.

Soit $a = \frac{n_a}{t_a}$ le résultat d'une mesure sur un échantillon radioactif et

----

(1) Dans ces conditions, ce ne sera plus seulement l'estimation de l'erreur relative mais aussi la valeur réelle de cette erreur qui sera constante. En effet, dans les mesures en précompte :

$$\sigma_a^2 = \frac{A^2}{\sqrt{n}} \quad \text{soit} \quad \varepsilon_a = \frac{1}{\sqrt{n}}$$

$m = \frac{n_m}{t_m}$ la valeur du mouvement propre, déterminée par une mesure séparée. La valeur A - M de l'activité de l'échantillon est estimée d'après la différence (a - m).

A ce sujet, on peut se poser la question suivante : Comment choisir les durées respectives t_a et t_m des mesures de l'activité A d'un échantillon, et du mouvement propre M pour que leur durée totale T soit minimum, l'estimation de l'activité cherchée A - M étant effectuée avec une certaine précision fixée à l'avance.

L'erreur sur l'estimation de A - M est :

$$e_{a-m} = \sqrt{\frac{A}{t_a} + \frac{M}{t_m}}$$

On a : $T = t_a + t_m$; en différenciant on obtient $dT = dt_a + dt_m$

On peut également différencier l'expression représentant le carré de l'erreur :

$$de^2_{a-m} = -\left(\frac{A dt_a}{t_a^2} + \frac{M dt_m}{t_m^2}\right)$$

Les expressions de dT et de de^2 sont nulles si l'on suppose que T est minimum et que la valeur de e est constante (fixée à l'avance).

$$dt_a + dt_m = 0 \qquad \frac{A dt_a}{t_a^2} + \frac{M dt_m}{t_m^2} = 0$$

En identifiant terme à terme ces deux égalités, on obtient :

$$\boxed{\frac{A}{t_a^2} = \frac{M}{t_m^2}}$$

Ce résultat montre dans quelles conditions il faut réaliser les mesures sur l'échantillon et le mouvement propre pour réduire au minimum la durée totale des mesures; la durée de chaque mesure doit être proportionnelle à la racine carrée de l'activité mesurée. Il indique également de façon indirecte qu'il n'est pas nécessaire de connaître le mouvement propre avec une grande précision, si celui-ci est petit par rapport à l'activité mesurée sur l'échantillon.

On peut également écrire :

$$\frac{\sqrt{A}}{t_a} = \frac{\sqrt{M}}{t_m} = \frac{\sqrt{A} + \sqrt{M}}{T}$$

En portant les valeurs déduites de cette égalité dans l'expression de l'erreur, on obtient :

$$e = \frac{\sqrt{A} + \sqrt{M}}{\sqrt{T}} \simeq \frac{\sqrt{a} + \sqrt{m}}{\sqrt{T}}$$

qui montre que la valeur de l'erreur est inversement proportionnelle à la racine carrée de la durée totale T des mesures.

6) Erreur dans l'estimation du rapport de deux activités.

Dans de très nombreux cas, on est amené à déterminer le rapport existant entre les activités de deux échantillons. On peut se poser la question suivante : Comment effectuer les mesures avec une durée totale minimum pour que l'erreur relative attachée à l'estimation de ce rapport ait une valeur déterminée, le mouvement propre étant supposé négligeable?

On a $$\varepsilon_{a/b} = \sqrt{\frac{1}{At_a} + \frac{1}{Bt_b}}$$

On va encore écrire que les expressions représentant les différentielles de l'erreur et de la durée T sont nulles.

$$d\varepsilon^2 = -\left[\frac{dt_a}{At_a^2} + \frac{dt_b}{Bt_b^2}\right] = 0 \qquad dT = dt_a + dt_b = 0$$

En identifiant terme à terme, il vient :

$$\boxed{At_a^2 = Bt_b^2}$$

En écrivant cette relation sous la forme $\frac{\sqrt{A}}{t_b} = \frac{\sqrt{B}}{t_a} = \frac{\sqrt{A}+\sqrt{B}}{T}$, on obtient :

$$\varepsilon_{a/b} = \frac{1}{\sqrt{T}}\left[\frac{1}{\sqrt{A}} + \frac{1}{\sqrt{B}}\right] \simeq \frac{1}{\sqrt{T}}\left[\frac{1}{\sqrt{a}} + \frac{1}{\sqrt{b}}\right]$$

Ces résultats montrent que l'erreur est encore inversement proportionnelle à la racine carrée de la durée totale des mesures, <u>mais la durée de chaque mesure doit être inversement proportionnelle à la racine carrée de l'activité mesurée. La valeur moyenne du nombre des désintégrations est alors directement proportionnelle à la racine carrée de l'activité mesurée.</u>

En conclusion, on voit que le choix de la durée des mesures ou des précomptes permettant de réaliser une série de mesures en imposant certaines conditions aux erreurs, dépend de la nature du problème posé. Les durées ou les précomptes doivent être choisis en fonction des activités à mesurer (dont on peut déjà connaître a priori l'ordre de grandeur au moyen d'une mesure antérieure ou d'une mesure rapide de très courte durée) de façon à satisfaire au mieux aux relations que nous résumons dans le tableau ci-dessous, où nous rappelons également les expressions des erreurs attachées à une mesure unique.

	Mesures en prétemps	Mesures en précompte
Erreur absolue constante	$\frac{A}{t}$ constant	$\frac{A^2}{n}$ constant
Erreur relative constante	At constant	n constant
Estimation de la somme ou de la différence de deux activités: Durée totale des mesures minimum	$\frac{t^2}{A}$ constant	$\frac{n^2}{A^3}$ constant(1)
Estimation du rapport de deux activités : Durée totale des mesures minimums.	At^2 constant	$\frac{n^2}{A}$ constant(1)
Erreur absolue sur une mesure unique	$e = \sqrt{\frac{A}{t}} \simeq \sqrt{\frac{a}{t}} = \sqrt{\frac{n}{t^2}}$	$e = \sqrt{\frac{A^2}{n}} \simeq \sqrt{\frac{a^2}{n}} = \sqrt{\frac{n}{t^2}}$
Erreur relative sur une mesure unique	$\varepsilon = \frac{1}{\sqrt{At}} \simeq \frac{1}{\sqrt{at}} = \frac{1}{\sqrt{n}}$	$\varepsilon = \frac{1}{\sqrt{n}}$

---

(1) Dans les mesures en précompte, la durée totale des mesures est une variable aléatoire. Ce n'est donc pas elle mais sa valeur moyenne que l'on rend minimum.

C. QUELQUES PROBLÈMES CONCERNANT LA PÉRIODE D'UN RADIOÉLÉMENT

ESTIMATION D'UNE PERIODE, LONGUE PAR RAPPORT AUX DUREES DES MESURES -

Pour connaître de façon précise la valeur de la période d'un radioélément, il suffit d'effectuer une mesure très précise d'activité à deux époques différentes Θ_1 et Θ_2. Si les durées t_1 et t_2 des mesures sont très petites par rapport à la période T, l'estimation T' de sa valeur est obtenue directement à partir des valeurs des activités mesurées a_1 et a_2 par l'équation classique :

$$T' = 0,301 \frac{\Theta_2 - \Theta_1}{\log a_1 - \log a_2}$$

Mais, en pratique, on préfère exécuter toute une série de mesures, à des époques différentes, et ajuster ensuite par une méthode statistique une droite aux points figuratifs des résultats de ces mesures dans un système de coordonnées semi-logarithmiques. L'examen de l'alignement des points obtenus, par une méthode graphique ou par l'emploi d'un test statistique permet, en effet, de vérifier que la dispersion des points autour de la droite obtenue est due uniquement au caractère aléatoire du phénomène mesuré, et non à des facteurs extérieurs (par exemple, présence d'une impureté radioactive de période différente).

Pour procéder ainsi, il est souhaitable que les erreurs absolues sur les valeurs des logarithmes des activités soient égales. Or, on démontre que l'erreur absolue sur le logarithme népérien d'une variable aléatoire est égale, au second ordre près, à l'erreur relative sur la valeur de la variable elle-même. Autrement dit, si l'on considère les logarithmes décimaux et que l'on écrive $\alpha = \log a$, on a : $e_\alpha = M \varepsilon_a$ où $M = \log_{10} e = 0,4343$.

On voit donc que dans ce cas, les mesures de l'activité de l'échantillon à différentes époques de sa décroissance doivent être effectuées en précompte (nombre de désintégrations compté constant).

On peut alors calculer par la méthode des moindres carrés une valeur estimée m de la pente $-M\lambda$ de la droite d'équation :

$$\alpha = \alpha_0 - M\lambda\Theta$$

exprimant la décroissance de l'activité du radioélément en fonction de la constante radioactive λ par l'expression classique du coefficient de régression :

$$m = \frac{\Sigma\Theta_i\alpha_i - K\bar{\Theta}\,\bar{\alpha}}{\Sigma\Theta_i^2 - K\bar{\Theta}^2}$$

où α_i ($i = 1, 2 \ldots k$) représente le logarithme décimal de l'activité a_i, mesurée à l'époque Θ_i. On en déduit alors une estimation :

$$T' = \frac{0,301}{m} \text{ de la période.}$$

ESTIMATION D'UNE PERIODE COURTE, PAR RAPPORT A LA DUREE DES MESURES -

Lorsque la période est trop courte pour qu'il soit possible de réaliser des mesures qui soient à la fois précises et de durée petite par rapport à la valeur de cette période, on peut déterminer celle-ci au moyen de seulement deux mesures, à condition que celles-ci soient de même durée t.

Soit $A = A_0 e^{-\lambda\Theta}$ l'expression de la décroissance de l'activité de l'échantillon, et τ l'intervalle de temps séparant les époques des mesures. Les valeurs des nombres de coups n_1 et n_2 que l'on peut compter au cours de ces mesures sont des variables aléatoires dont les valeurs moyennes inconnues N_1 et N_2 ont pour expression :

$$N_1 = \int_0^t A_0 e^{-\lambda\Theta}\, d\Theta = \frac{A_0}{\lambda}(1 - e^{-\lambda t})$$

$$N_2 = \int_{t+\tau}^{2t+\tau} A_0 e^{-\lambda\Theta} = \frac{A_0}{\lambda}(1 - e^{-\lambda t})e^{-\lambda(t+\tau)}$$

On en tire $e^{-\lambda(t+\tau)} = \frac{N_2}{N_1}$. La valeur T de la période ($T = \frac{0,693}{\lambda}$) peut donc être estimées directement à partir des résultats n_1 et n_2 des mesures, par la valeur T' telle que :

$$e^{-\frac{0,693(t+\tau)}{T'}} = \frac{n_2}{n_1} \quad \text{c'est-à-dire} \quad T' = 0,301 \frac{t + \tau}{\log n_1 - \log n_2}$$

ESTIMATION DE L'ACTIVITE INSTANTANEE D'ELEMENTS DE COURTE PERIODE CONNUE -

Dans de nombreux cas, la période du radioélément est connue mais on désire connaître la valeur de son activité à une époque déterminée, choisie comme époque origine (en vue, par exemple, de comparer cette activité à celle d'un autre échantillon de la même substance).

Lorsque la période du radioélément est longue par rapport à la durée de la mesure, cette détermination ne présente aucune difficulté, mais il n'en est pas de même s'il s'agit d'un élément de très courte période, car la décroissance du radioélément se produit au cours même de la mesure. Dans ce cas, il faut déterminer d'abord l'activité réelle, que nous appellerons "activité instantanée", existant à un moment donné (celui-ci peut être, par exemple, l'instant du début de la mesure). Soit A_0 cette activité instantanée inconnue. Au cours de la mesure de durée t, on compte un nombre de désintégrations n, qui est une variable aléatoire obéissant, en raison de la propriété d'additivité des lois de Poisson, à une loi de Poisson de moyenne N et d'écart-type $\sigma_n = \sqrt{N}$.

La loi de décroissance de l'échantillon permet de calculer N en fonction de A_0:

$$N = \int_0^t A_0 e^{-\lambda x}\, dx = \frac{1 - e^{-\lambda t}}{\lambda} A_0$$

Cette relation permet d'obtenir une estimation a_0 de la valeur de l'activité instantanée A_0, en fonction de la valeur mesurée n :

$$a_0 = \frac{n\lambda}{1 - e^{-\lambda t}}$$

La variable aléatoire a_0 se déduit de la variable poissonienne n, en multipliant celle-ci par le nombre certain $\rho = \frac{\lambda}{1 - e^{-\lambda t}}$.

Comme on l'a vu au chapitre I, la loi de probabilité de la variable a_0 peut donc être assimilée à une loi normale de moyenne A_0, dont on peut estimer l'écart-type par :

$$\sigma'_{a_0} = \rho \sqrt{n} = \frac{\lambda}{1 - e^{-\lambda t}} \sqrt{n}$$

ESTIMATION DES ACTIVITES RESPECTIVES DE DEUX RADIOELEMENTS DONT LES PERIODES SONT CONNUES -

Un cas qui se présente souvent est celui qui consiste à déterminer les valeurs, à l'époque origine, des activités respectives d'un mélange de deux radioéléments ayant des périodes différentes connues. FREYLING et BUNNEY(1), ont indiqué une méthode graphique simple permettant d'obtenir directement les valeurs cherchées avec une bonne précision.

Soit A_0 et B_0 les activités des substances à l'époque origine, α et β leurs constantes radioactives respectives, et M l'activité mesurée à l'époque Θ sur le mélange de ces substances.

$$M = A_0 e^{-\alpha\Theta} + B_0 e^{-\beta\Theta}$$

En supposant $\beta > \alpha$, cette égalité peut s'écrire :

$$M\, e^{\alpha\Theta} = A_0 + B_0 e^{-(\beta-\alpha)\Theta}$$

Si sur un graphique, on porte en ordonnées les diverses valeurs calculées de l'expression $M\, e^{\alpha\Theta}$, d'après les résultats des mesures faites à diverses époques, et en abscisses les valeurs correspondantes de l'expression $e^{-(\beta-\alpha)\Theta}$, les points obtenus doivent être sensiblement alignés sur une droite; l'ordonnée à l'origine de cette droite représente la valeur A_0 de l'activité de la première substance à l'époque origine, et la pente représente la valeur B_0 de l'activité de la deuxième substance.

IV. - RECHERCHE DES ERREURS DUES A DES CAUSES ACCIDENTELLES

SIGNIFICATION D'UN TEST STATISTIQUE -

Nous avons vu que les différentes grandeurs qui interviennent dans les problèmes de radioactivité sont des variables aléatoires, dont nous avons défini les lois théoriques de probabilité. Les valeurs obtenues au cours d'une série de mesures, effectuée dans des conditions déterminées, constituent un échantillon de la population des valeurs que peut prendre la variable considérée. Ces valeurs présentent entre elles des écarts, et il importe de savoir si ces écarts correspondent parfaitement à ceux qu'on peut s'attendre à observer entre les valeurs d'une variable qui suit la loi de probabilité supposée; sinon, on sera tenté de penser qu'aux fluctuations aléatoires prévues, se sont ajoutées des fluctuations accidentelles dues à des causes extérieures. Pour répondre à cette question on utilise un test statistique.

Un test statistique n'indique pas de façon catégorique, si oui ou non, la distribution expérimentale résultant des observations correspond à la loi théorique que le phénomène observé est supposé suivre. Il permet seulement, soit de conclure qu'il n'y a pas concordance entre la distribution expérimentale et la

----

(1) E.C. FREYLING et L.R. BUNNEY - Nucleonics, Sept. 1956, p. 112.

loi théorique, et en énonçant cette affirmation on prend un certain risque de la formuler à tort, soit de dire qu'il n'y a pas de raison de penser que cette concordance n'est pas réalisée, sans que l'on ait aucune certitude à ce sujet. Les conclusions que l'on peut tirer d'un test ont donc une certaine probabilité d'être inexactes, et il convient de se fixer à l'avance le risque que l'on prend de se tromper.

Un test statistique est basé sur la connaissance de la loi de répartition d'une certaine variable aléatoire, que nous désignerons par Ω, qui caractérise les écarts qu'il serait possible de constater entre une distribution théorique et les distributions observables sur des échantillons prélevés au hasard dans une population totale correspondant effectivement à cette distribution théorique. La connaissance de la loi de répartition de la variable Ω permet d'établir des tables indiquant, en fonction d'une valeur ω prise par la variable Ω, quelle est la probabilité d'observer une valeur égale ou supérieure à ω. L'application du test consiste à calculer, à partir des valeurs expérimentales obtenues, la valeur ω' d'une variable Ω' définie de la même façon que la variable Ω, qui caractérise les écarts entre la distribution théorique et la répartition expérimentale à tester, puis à voir d'après les tables quelle serait la probabilité d'observer une valeur au moins égale à ω' dans le cas où la variable Ω' suivrait effectivement la loi de répartition de Ω. Lorsque cette probabilité est inférieure à une certaine valeur appelée "seuil de signification", la variable Ω a très peu de chances de prendre une valeur aussi élevée que ω'; on en conclut que la quantité Ω' ne suit pas la loi de la variable Ω et que par conséquent, la distribution expérimentale ne correspond pas à la distribution théorique supposée. Si cette probabilité $P(\Omega \geqslant \omega')$ est supérieure à la valeur du seuil de signification, on en conclut qu'une valeur de Ω aussi élevée que ω' étant tout à fait plausible, il n'y a pas de raison de penser que la variable Ω' ne suit pas la loi de Ω et, par conséquent, d'estimer que la distribution expérimentale n'obéit pas à la loi théorique supposée. Dans le premier cas, on dit qu'il y a un écart significatif entre la distribution expérimentale et la loi qu'elle est supposée suivre; dans le second, on dit que l'écart n'est pas significatif.

Le choix du seuil de signification est dicté par diverses considérations basées sur l'expérience, dont certaines peuvent évidemment présenter un caractère arbitraire : dans de nombreux cas, on considère les seuils à 5% ou 1%. Choisir un seuil à 5% signifie que, pour une distribution expérimentale correspondant parfaitement à la loi théorique, on a 5 chances sur 100 d'affirmer que l'écart est significatif, et de penser ainsi à tort que la distribution expérimentale ne correspond pas à la distribution théorique. Afin de nuancer son appréciation, on peut convenir par exemple de dire qu'un écart est significatif s'il est significatif au "risque" 5% et pas au risque 1%, et hautement significatif s'il est significatif à la fois aux risques 5% et 1%.

On voit que la conclusion tirée d'un test statistique a un caractère "probabiliste"; il n'en demeure pas moins que l'emploi judicieux d'un tel test peut fournir des éléments d' information précieux.

L'application pratique d'un test statistique se ramène à calculer la valeur ω' de la quantité Ω' résultant des observations expérimentales, et à la comparer à la valeur ω qui correspond au seuil de signification choisi; si elle lui est supérieure, l'écart est jugé significatif.

Il y a d'ailleurs lieu de distinguer plusieurs sortes de tests notamment :

Les tests d'ajustement qui ont pour but de vérifier que l'échantillon constitué par l'ensemble des valeurs obtenues peut effectivement provenir d'une population suivant la loi de probabilité supposée.

Les tests de rejet d'une valeur aberrante qui ont pour but de vérifier s'il est justifié d'écarter le résultat d'une mesure particulière, que l'on a des raisons de supposer entaché d'une erreur accidentelle.

TEST D'AJUSTEMENT DE χ^2 APPLIQUE AUX MESURES D'ACTIVITE -

Le test de χ^2 dû à Karl PEARSON, consiste à calculer à partir des valeurs observées, la valeur d'une variable X^2 qui doit avoir une distribution de χ^2 (définie au chapitre I) si la distribution observée concorde avec la distribution théorique.

Le test de χ^2 est habituellement utilisé dans des conditions que nous ne développerons pas ici sous leur forme générale : en effet les variables qui interviennent dans les mesures d'activité ont des distributions de Poisson, ce qui nous permet d'envisager l'emploi de ce test, sous une forme particulière, propre à la nature de ces variables.

Le problème consiste en effet à tester si les résultats de k mesures effectuées en prétemps suivent une loi de Poisson, ou si les résultats de k mesures effectuées en précompte, correspondent bien à la distribution que nous avons définie au chapitre I.

Nous avons indiqué au chapitre I que dans le cas d'une distribution de Poisson, la quantité $X^2 = \frac{\Sigma (x_i - \bar{x})^2}{\bar{x}}$ avait une distribution de χ^2 avec k - 1 degrés de liberté.

Dans le cas des mesures en prétemps, on sera donc amené à former la quantité $X^2 = \frac{\sum_{i=1}^{k}(n_i - \bar{n})^2}{\bar{n}}$ et à vérifier dans les tables de la distribution de χ^2 dont nous donnons ici un extrait (tableau I) si la valeur trouvée est inférieure à la valeur χ_o^2, correspondant au seuil de signification choisi pour k - 1 degrés de liberté.

S'il en est ainsi, l'écart avec la loi de Poisson sera considéré comme non significatif et on pourra conclure que l'on n'a pas de raison de penser qu'un facteur extérieur a perturbé les résultats des mesures. Sinon, on pourra estimer qu'aux fluctuations aléatoires se sont ajoutées des erreurs dues à des causes extérieures.

On notera qu'en raison de la relation $a_t = \frac{n}{t}$, il revient au même d'appliquer le test de χ^2 à la quantité $\frac{t \sum_{i=1}^{k}(a_{t_i} - \bar{a}_t)^2}{\bar{a}_t}$.

Dans le cas des mesures en précompte, nous avons vu au chapitre II que les distributions des variables t et a_n pouvaient être assimilées à des distributions normales de caractéristiques :

$$\begin{cases} m_t = \dfrac{n}{A} \\ \sigma_t^2 = \dfrac{n}{A^2} = \dfrac{m_t^2}{n} \end{cases} \quad \text{et} \quad \begin{cases} m_{a_n} = A \\ \sigma_{a_n}^2 = \dfrac{A^2}{n} = \dfrac{m_{a_n}^2}{n} \end{cases}$$

Nous avons également indiqué au chapitre I que si x est une variable dont la distribution est normale, la quantité $\frac{\sum_{i=1}^{k}(x_i - \bar{x})^2}{\sigma^2}$ est distribuée comme χ^2 avec k - 1 degrés de liberté.

On peut montrer que si l'on estime la valeur inconnue σ de l'écart-type de l'une ou l'autre des variables t et a_n par l'expression $\sigma^2 \simeq \frac{\overline{x}^2}{n}$ (n étant grand), on obtient une variable dont la distribution est aussi très sensiblement celle de χ^2 avec k - 1 degrés de liberté(1).

On sera donc amené dans le cas de mesures en précompte à appliquer le test de χ^2 avec k - 1 degrés de liberté à l'une des deux quantités $n\frac{\Sigma(t_i - \overline{t})^2}{\overline{t}^2}$ ou ou $n\frac{\Sigma(a_{ti} - \overline{a}_t)^2}{\overline{a}_t^2}$.

COMPARAISON DES RESULTATS DE 2 MESURES D'ACTIVITE -

Le test de χ^2 peut être également employé pour comparer les résultats de deux mesures. Il s'agit de voir alors si la différence observée entre ces deux résultats est due uniquement aux fluctuations aléatoires ou si l'on peut penser qu'un autre phénomène peut en être responsable. On vérifie aisément que l'expression de X^2 (un seul degré de liberté) devient dans ce cas pour des mesures en prétemps :

$$\frac{(n_1 - \overline{n})^2 + (n_2 - \overline{n})^2}{\overline{n}} = \frac{(n_1 - n_2)^2}{n_1 + n_2}$$

et pour des mesures en précompte :

$$2n \frac{(t_1 - t_2)^2}{(t_1 + t_2)^2}$$

L'examen des tables montre que les valeurs de χ^2 qui correspondent à des seuils de signification de 95% et 99% sont respectivement 3,84 et 6,64.

Par suite si dans le cas de mesures en prétemps :

$\frac{(n_1 - n_2)^2}{n_1 + n_2} > 3,84$ l'écart est significatif;

$\frac{(n_1 - n_2)^2}{n_1 + n_2} > 6,64$ l'écart est hautement significatif.

EXEMPLES D'APPLICATION -

<u>1er exemple</u> :

<u>Au cours d'une même série de mesures d'activité en prétemps, on mesure</u>

----

(1) On peut, par exemple, considérer l'expression $y = \frac{\Sigma(x_i - \overline{x})^2}{\overline{x}^2/n}$ comme le quotient de deux variables aléatoires $u = \frac{\Sigma(x_i - \overline{x})^2}{\sigma^2}$ et $v^2 = \frac{\overline{x}^2}{n\sigma^2}$ indépendantes en probabilité (on montre que dans une distribution normale, les variables $\overline{x}$ et $\Sigma(x_i - \overline{x})^2$ sont effectivement indépendantes en probabilité). On a vu (chapitre I), que la variable u a une distribution de χ^2 avec k - 1 degrés de liberté, tandis que $v = \frac{\overline{x}}{\sigma\sqrt{n}}$ qui a une distribution normale de moyenne 1 et de variance $\frac{1}{kn}$ ($\overline{x}$ a pour moyenne $m = \sigma\sqrt{n}$ et pour variance $\frac{\sigma^2}{k}$) est une variable quasi certaine puisque n est de l'ordre de plusieurs milliers.

à titre de contrôle l'activité d'un étalon d'oxyde d'uranium. On veut savoir, d'après les résultats obtenus sur l'étalon, si les conditions de fonctionnement du compteur sont restées identiques pendant toute la durée des mesures.

Le tableau II donne l'ensemble des calculs effectués pour déterminer la valeur de χ^2. On trouve $\chi^2 = 28,2$. Le nombre des degrés de liberté est $22 - 1 = 21$. Le tableau I montre que pour 21 degrés de liberté, le seuil de signification à 5% correspond à un χ^2 de 32, 7.

La répartition des résultats correspond bien à une distribution de Poisson de moyenne 24, 576. Apparemment, aucun autre facteur que les fluctuations aléatoires du phénomène n'est intervenu au cours des mesures.

A titre indicatif, nous reproduisons au tableau III les résultats de calculs analogues effectués sur 14 séries de mesures. On voit que dans 9 des séries, la probabilité d'obtenir une valeur de χ^2 supérieure à la valeur observée est inférieure à 1%. On peut donc penser qu'au cours de ces essais, les conditions de fonctionnement du compteur ne sont pas restées identiques (fluctuations accidentelles de la tension d'alimentation, variations du rendement du compteur, etc.).

Le test de χ^2 ainsi appliqué aux mesures de l'activité d'un étalon effectuées au cours d'une période de plusieurs jours ou plusieurs semaines permet de tester que les mesures effectuées sur un échantillon au cours de cette période en vue de déterminer la courbe de décroissance de son activité ont été exécutées dans des conditions expérimentales comparables.

2ème exemple :

Convient-il d'effectuer des mesures au compteur, aussitôt sa mise en route ?

Le tableau IV montre l'ensemble des calculs effectués pour déterminer la valeur de χ^2 pour une série de 10 mesures successives de 3 minutes effectuées sur un étalon d'oxyde d'uranium. La valeur de χ^2 relative à l'ensemble des 10 mesures est de 82,46, valeur dont la probabilité d'être dépassée est nettement inférieure au seuil de 1% (9 degrés de liberté). L'examen des valeurs $(n - \bar{n})$ indique nettement que cet écart est négatif pour les deux premières mesures, et positif pour les 8 autres. Si on applique le test de χ^2 seulement à ces dernières, on obtient une valeur de 3, 11 qui n'est plus significative même avec un seuil de 5%.

Ces résultats montrent donc que les 8 dernières mesures correspondent bien à une distribution de Poisson de moyenne estimée à 23.964 tandis que les deux premières mesures présentent une fluctuation due à des causes extérieures. Ce résultat ayant été observé à plusieurs reprises, mais non de façon systématique, il est apparu prudent de faire fonctionner le compteur pendant une dizaine de minutes avant de commencer les mesures.

3ème exemple :

On utilise un certain dispositif pour comparer les activités de divers échantillons. On veut s'assurer que la géométrie est bien reproductible et que certains facteurs (existence par exemple d'un jeu entre deux pièces mobiles du support d'échantillon) ne risquent pas d'introduire des erreurs accidentelles.

Pour faire cette vérification, nous avons pris deux échantillons dont nous avons mesuré alternativement l'activité, dix foix de suite. L'application du test de χ^2 à chacune des deux séries de mesures obtenues a donné les valeurs 11,7 et 15,2 (9 degrés de liberté). On peut également appliquer à la somme des χ^2

trouvés, soit 26,9, le test de χ^2 avec 18 degrés de liberté (additivité de la distribution de χ^2).

Ces résultats indiquent que le dispositif utilisé ne paraît pas introduire d'erreurs accidentelles dues à des variations dans la mise en place des porteurs.

L'interprétation d'un écart significatif aurait été plus délicate car il aurait fallu vérifier que cet écart était bien dû au dispositif employé, et non à d'autres causes extérieures, telles que celles indiquées dans le premier exemple.

4ème exemple :

On mesure dans des conditions géométriques identiques les activités des deux faces d'une rondelle d'acier contenant du radiophosphore. Peut-on considérer que le phosphore est réparti de façon uniforme dans l'épaisseur de l'échantillon (qui est beaucoup plus grande que le parcours du rayonnement β du phosphore dans l'acier)?

Des mesures de cinq minutes ont fourni les valeurs suivantes (mouvement propre non déduit) : $n_1 = 1214 \quad n_2 = 1366$

$$\frac{(n_1 - n_2)^2}{n_1 + n_2} = \frac{(152)^2}{2580} = 8,95$$

Cette valeur est supérieure à 6,64, valeur de χ^2 avec 1 degré de liberté correspondant au seuil 1%. Il existe donc une différence significative entre les activités mesurées sur les deux faces de l'échantillon.

CRITERES DE REJET D'UNE VALEUR ABERRANTE -

Il arrive que l'un des résultats provenant d'une même série de mesures paraisse anormalement écarté de la valeur moyenne. On peut alors se demander si cet écart n'est pas dû à l'influence d'une cause extérieure qui a perturbé la mesure considérée. L'emploi de tests statistiques permet de déterminer si le résultat incriminé est significativement aberrant.

On peut d'abord utiliser le test de χ^2. Supposons que celui-ci appliqué à l'ensemble des k valeurs de la série ait mis en évidence un écart significatif. On peut alors supprimer la valeur qui s'écarte le plus de la valeur moyenne, et appliquer le test de χ^2 à l'ensemble des k - 1 autres valeurs : si cette fois-ci, le test n'indique pas d'écart significatif, on peut penser que la valeur incriminée est aberrante. C'est ce qui a été fait dans le cas de l'exemple n°2 ci-dessus.

On peut également comparer l'écart $|x_e - \bar{x}|$ entre la valeur suspecte x_e et la moyenne $\bar{x}$ de tous les résultats à l'écart-type σ de la population, estimé à partir de la moyenne $\bar{x}$ de l'échantillon (cf. chapitre II). On regarde alors dans le tableau V si le rapport $\frac{x_e - \bar{x}}{\sqrt{\bar{x}}}$ observé ($\frac{n_e - \bar{n}}{\sqrt{\bar{n}}}$ dans le cas de mesures en prétemps) est supérieur à la valeur théorique correspondant au nombre k des résultats de la série et au niveau de signification choisi. S'il en est ainsi, on peut admettre que la valeur extrême est aberrante, mais il peut être bon de s'assurer que l'ensemble des autres résultats n'indique pas d'écart significatif avec la distribution théorique.

TABLEAU I

Valeurs de χ^2 correspondant à des seuils de signification P de 0, 10; 0, 05 et 0, 01

Degrés de liberté	P = 0, 10	P = 0, 050	P = 0, 010
1	2,71	3.84	6,64
2	4,61	5,99	9,21
3	6,25	7,82	11,34
4	7,78	9,49	13,28
5	9,24	11,07	15,09
6	10,64	12,59	16,81
7	12,02	14,07	18,48
8	13,36	15,51	20,09
9	14,68	16,92	21,67
10	15,99	18,31	23,21
11	17,28	19,68	24,73
12	18,55	21,03	26,22
13	19,81	22,36	27,69
14	21,06	23,69	29,14
15	22,31	25,00	30,58
16	23,54	26,30	32,00
17	24,77	27,59	33,41
18	25,99	28,87	34,81
19	27,20	30,14	36,19
20	28,41	31,41	37,57
21	29,62	32,67	38,93
22	30,81	33,92	40,29
23	32,01	35,17	41,64
24	33,20	36,42	42,98
25	34,38	37,65	44,31
26	35,56	38,89	45,64
27	36,74	40,11	46,96
28	37,92	41,34	48,28
29	39,09	42,56	49,59
30	40,26	43,77	50,89

TABLEAU II

Application du test de χ^2 à l'exemple I

n	n - $\overline{n}$	(n - $\overline{n}$)^2
24 387	- 189	35 721
24 408	- 168	28 224
24 349	- 227	51 529
24 698	+ 122	14 884
24 724	+ 148	21 904
24 788	+ 212	44 944
24 109	- 467	218 089
24 420	- 166	27 556
24 494	- 82	6 724
24 668	+ 92	8 464
24 525	- 51	2 601
24 743	+ 167	27 889
24 461	- 115	13 225
24 441	- 135	18 225
24 677	+ 101	10 201
24 765	+ 189	35 721
24 635	+ 59	3 481
24 693	+ 117	13 689
24 768	+ 192	36 864
24 531	- 45	2 025
24 561	- 15	225
24 841	+ 265	70 225
540 676		692 410

$$\overline{n} = \frac{540\ 676}{22} = 24\ 576$$

$$X^2 = \frac{\Sigma(n - \overline{n})^2}{n} = \frac{692\ 410}{24\ 576} = 28{,}2$$

avec 21 degrés de liberté (non significatif).

TABLEAU III

Résultats de l'application du test de χ^2 aux valeurs de l'activité d'un échantillon d'oxyde d'uranium mesurées au cours de diverses séries d'essais (mesures en prétemps).

Repère de l'essai	Nombre de mesures	Valeur observée de χ^2	Valeur de χ^2 correspondant aux seuils :		L'écart est-il hautement significatif ?
			5%	1%	
A	24	18,7	35,2	41,6	non
B	12	15,2	19,7	24,7	non
C	12	24,7	19,7	24,7	oui
D	13	30,3	21,0	26,2	oui
E	22	28,2	32,7	38,9	non
F	19	26,2	28,9	34,8	non
G	17	89	26,3	32,0	oui
H	11	69	18,3	23,2	oui
I	12	51,1	19,6	24,7	oui
J	12	28,6	19,6	24,7	oui
K	16	16,5	24,9	30,6	non
L	12	42,5	19,7	24,7	oui
M	13	43,0	21,0	26,2	oui
N	12	25,8	19,7	24,7	oui

TABLEAU IV

Exemple 2 :

Application du test de χ^2 pour étudier la période de mise en route d'un compteur

$n - \overline{n}_1$	$(n - \overline{n}_1)^2$	n	$n - \overline{n}_2$	$(n - \overline{n}_2)^2$
- 1,219	1.485.961	22.553		
- 321	103.041	23.451		
+ 259	67.081	24.031	+ 67	4.489
+ 86	7.396	23.858	- 106	11.236
+ 103	10.609	23.875	- 89	7.921
+ 331	109.561	24.103	+ 139	19.321
+ 158	24.964	23.930	- 34	1.156
+ 146	21.316	23.918	- 46	2.116
+ 342	116.964	24.114	+ 150	22.500
+ 1116	13.456	23.888	- 76	5.776
	1.960.349			74.515

Pour les 10 mesures :

$\overline{n}_1 = 23.772$

$$X^2 = \frac{\Sigma(x - \overline{x}_1)^2}{\overline{x}_1} = 82,46$$

avec 9 degrés de liberté

(hautement significatif)

Pour les 8 dernières mesures :

$\overline{n}_2 = 23.964$

$$X^2 = \frac{\Sigma(x - \overline{x}_2)^2}{\overline{x}_2} = 3,11$$

avec 7 degrés de liberté

(non significatif)

TABLEAU V

Valeurs de $\frac{|x_e - \overline{x}|}{\sigma} \simeq \frac{|x_e - \overline{x}|}{\sqrt{\overline{x}}}$ correspondant aux seuils de signification

$p = 5\%$ et 1%(1).

k	p = 5%	p = 1%
3	1,95	2,40
4	2,16	2,62
5	2,30	2,76
6	2,41	2,87
7	2,49	2,95
8	2,56	3,02
9	2,61	3,07
10	2,66	3,12
11	2,70	3,16
12	2,74	3,20
13	2,78	3,23
14	2,80	3,26
15	2,84	3,29
16	2,86	3,31
17	2,89	3,33
18	2,91	3,36
19	2,93	3,38
20	2,95	3,39
21	2,96	3,41
22	2,98	3,43
23	3	3,44
24	3,01	3,45
25	3,03	3,47

----

(1) D'après GRUBBS : "Sample criteria for testing outlying observations" Annals of Mathem. Stat. 21, p. 27 (Mars 1950).

L'échantillon est supposé extrait d'une distribution normale d'écart-type σ et la table que nous donnons ici est celle des valeurs de $\frac{|x_e - \overline{x}|}{\sigma}$. Nous pensons qu'étant donné que $\overline{x}$ constitue une bonne estimation de σ^2, la probabilité pour que, en l'absence de tout résultat aberrant, $\frac{x_e - \overline{x}}{\sqrt{\overline{x}}}$ ou $-\frac{x_e - \overline{x}}{\sqrt{\overline{x}}}$ dépasse les valeurs données dans la table ne doit pas différer beaucoup de 5% ou 1%, tout en restant inférieure à ces nombres (les valeurs de $\frac{|x_e - \overline{x}|}{s}$ avec $s = \sqrt{\frac{1}{n} \Sigma (x_i - \overline{x})^2}$ correspondant aux mêmes risques sont inférieures d'environ 0,1 à 0,3 à celles de $\frac{x_e - \overline{x}}{\sigma}$).

TABLEAU VI

Etude du caractère suspect d'un résultat particulier

	Résultat	Ecart à la moyenne		Résultat	Ecart à la moyenne
1	6 815	- 134	11	6 943	- 6
2	6 784	- 165	12	7 017	+ 68
3	6 815	- 134	13	6 982	+ 33
4	7 034	+ 85	14	6 678	- 271
5	6 907	- 42	15	6 900	- 49
6	7 161	+ 212	16	7 050	+ 101
7	6 898	- 51	17	7 110	+ 161
8	6 845	- 104	18	7 058	+ 99
9	7 066	+ 117	19	7 021	+ 72
10	6 997	+ 48			
			Total :	132 041	

Le total est 132.041 et la moyenne : 6949

L'écart-type est : $\sqrt{6.949} = 83,4$

Pour 19 résultats, l'écart est significatif si :

$$\frac{|x - \bar{x}|}{\sigma} > 2,71$$

Pour 19 résultats, l'écart est très significatif si :

$$\frac{|x - \bar{x}|}{\sigma} > 3,19$$

Le total des 18 résultats obtenus en excluant le 14^{e} est 125.363.

La moyenne est : 6.964 et l'écart-type : 83,5.

Exemple :

Peut-on considérer comme suspects, deux résultats de mesure qui s'écartent sensiblement de la valeur moyenne de la série à laquelle ils appartiennent ?

Au cours d'une série de mesures, on a régulièrement effectué des comptages sur un échantillon étalon, en vue de vérifier que les conditions de fonctionnement de l'appareil demeuraient identiques. Ces comptages étaient au nombre de 19 (voir tableau VI); leur valeur moyenne était 6. 949 et leur écart-type estimé $\sigma' = \sqrt{6.949} = 83,4$. Deux valeurs présentent un écart par rapport à la moyenne supérieur à $2\sigma'$: ce sont 7. 161 (écart de + 212) et 6.678 (écart de - 271). Peut-on considérer que ces valeurs indiquent un fonctionnement momentané incorrect de l'appareil de mesure ?

Nous allons d'abord appliquer le test à la valeur ayant l'écart le plus élevé.

Le rapport $\frac{|x_e - \bar{x}|}{\sigma}$ correspondant à ce résultat est $\frac{271}{83,4} = 3,25$.

Les valeurs indiquées dans le tableau V montrent que dans une série de 19 valeurs, l'une d'entre elles peut être considérée comme significativement aberrante si le rapport $\frac{|x_e - \bar{x}|}{\sigma}$ est supérieur à 2, 71 et comme très significativement aberrante si ce rapport est supérieur à 3, 19.

Nous voyons donc qu'il y a moins de 1 chance sur 100 d'observer une valeur dont l'écart par rapport à la moyenne est de 271 dans une série de 19 valeurs prises par une variable aléatoire obéissant à une loi de Poisson dont la moyenne est 6.949. Nous pouvons donc considérer la valeur de 6.678 comme aberrante et admettre que nous n'en tenons pas compte. Nous sommes alors amenés à considérer une série de 18 résultats dont la valeur moyenne est 6.964 et l'écart-type 83, 5. L'écart de la sixième valeur (7.161) par rapport à cette nouvelle moyenne est 197. $\frac{|x_e - \bar{x}|}{\sigma} = \frac{197}{83,5} = 2,36$.

L'examen du tableau V montre que l'on a plus de 5 chances sur 100 d'observer une valeur dont l'écart par rapport à la valeur moyenne est 197 dans une série de 18 valeurs prises par une variable suivant une loi de Poisson dont la moyenne est 6.964. Il n'y a donc pas lieu de considérer la valeur de 7.161 comme aberrante.

Ces calculs montrent que s'il y a lieu de considérer comme aberrante la valeur de 6.678 et de l'écarter, il n'y a pas lieu de tenir pour suspect le résultat de 7.161.

V. MINIMISATION DE L'INFLUENCE D'UNE CAUSE D'ERREUR EXTÉRIEURE DANS LA DÉTERMINATION DES ACTIVITÉS DE PLUSIEURS ÉCHANTILLONS

(Méthode des carrés gréco-latins)

Dans le chapitre précédent, nous avons vu que l'emploi du test de χ^2 permettait de déceler l'existence d'une cause d'erreur accidentelle survenue au cours d'une série de mesures : l'utilisation d'un test de rejet permet, en outre, de vérifier si un résultat de mesure jugé suspect peut être valablement écarté, les autres valeurs de la même série de mesures étant alors considérées comme n'ayant pas subi l'influence d'une erreur accidentelle. Mais il y a des cas où il n'est pas possible d'utiliser les résultats d'un ensemble de comptages, si l'un d'entre eux doit être écarté, et il est alors nécessaire de recommencer en totalité cette série de mesures. On peut donc avoir intérêt à réaliser les comptages suivant un certain plan, qui minimise l'influence d'une cause d'erreur accidentelle possible, et permet ainsi de tirer des indications valables de l'ensemble des résultats obtenus, malgré le caractère erroné de l'un d'entre eux.

Nous avons déjà signalé (voir chapitre III, p.) que l'erreur sur la valeur moyenne de k mesures de même durée t est égale à celle d'une mesure unique de durée kt; à condition de ne pas introduire de cause d'erreur extérieure (imprécision plus grande de la durée exacte de la mesure, par suite de l'erreur due au déclenchement du chronomètre, par exemple), on a intérêt à scinder une mesure de longue durée en plusieurs mesures de courte durée (par exemple, 6 mesures de 5 minutes, ou même 10 mesures de 3 minutes au lieu d'une mesure unique de 30 minutes. Si, pour une cause accidentelle, la mesure a été perturbée et que seul, l'un des résultats de comptage s'en trouve affecté (exécution, par exemple, d'une fausse manœuvre, erreur de lecture, etc. . .), il n'y aura pas lieu de recommencer ces mesures, puisque la moyenne des 5 ou 9 autres comptages fournira une valeur valable dont on peut d'ailleurs contrôler la validité par un test de χ^2.

Si l'on pense même que le fonctionnement de l'appareil peut présenter une variation lente dans le temps (les premiers et derniers comptages de la même série ne sont pas effectués dans des conditions absolument identiques en raison, par exemple, d'une mauvaise stabilisation de la haute tension), et que l'on désire cependant pouvoir connaître correctement les activités relatives de plusieurs échantillons, on aura intérêt à effectuer une première série de mesures comportant un comptage de courte durée sur chacun de ces échantillons, et à recommencer 4 ou 5 fois de suite cette série de comptages successifs. L'action de la cause d'erreur se répercutera de façon sensiblement égale sur la valeur moyenne des activités de chacun des échantillons, et son influence sur le quotient de deux de ces valeurs moyennes sera minimisée. En outre, en opérant ainsi, on supprime la nécessité d'intercaler, au cours des comptages, des mesures sur un échantillon étalon destinées à vérifier la constance des conditions de fonctionnement de l'appareillage. Cette manière de faire se traduit finalement par un gain de temps.

Pour effectuer les mesures, on peut réaliser chaque série de comptages en prenant chaque fois les échantillons dans le même ordre, ou dans un ordre déterminé par tirage au sort; on pourra, de cette façon, minimiser l'influence éventuelle d'une erreur accidentelle liée au temps : mais lorsqu'une telle cause d'erreur se sera manifestée (ce que l'on pourra vérifier par l'application du test de χ^2), on ne pourra pas contrôler si son influence sur les valeurs moyennes des activités a été effectivement très faible. Il peut être alors préférable, en vue de pouvoir faire cette vérification, de réaliser les mesures dans un ordre particulier, en utilisant, par exemple, la méthode connue sous le nom de "carrés gréco-latins".

Cette méthode que nous allons décrire ne peut s'appliquer de façon commode qu'à 3, 4 ou 5 échantillons; elle est l'extension aux mesures d'activité d'une méthode déjà utilisée depuis longtemps en agriculture pour équilibrer l'influence des variations de fertilité existant dans un même champ, où l'on veut comparer les propriétés de plants d'origine différente; elle a été préconisée par H. H. SELIGER(1), qui indique l'avoir utilisée avec succès pour plusieurs centaines de séries de mesures. Son intérêt n'est pas seulement de minimiser l'influence d'une cause d'erreur accidentelle, mais de permettre d'examiner, par l'application d'un test statistique, si cette influence est effectivement très faible.

PRINCIPE DE LA METHODE DES CARRES GRECO-LATINS -

Considérons un tableau de 4 rangées de 4 cases chacune, dans lequel on dispose 4 lettres A, B, C, D répétées chacune 4 fois, de telle manière qu'une même lettre n'apparaisse qu'une seule fois dans une même rangée ou une même colonne. Il existe plusieurs manières de réaliser une telle disposition appelée carré latin : le tableau 1 représente l'une d'entre elles.

Le tableau 2 représente une disposition analogue pour les 4 lettres grecques α, β, γ, δ; il présente de plus la particularité qu'à chacune des 4 cases occupées par une même lettre grecque, il correspond 4 cases du tableau 1 comportant les 4 lettres A, B, C, D.

Le tableau 3, où sont disposés les 4 chiffres romains I, II, III, IV, présente également la même particularité vis-à-vis des tableaux 1 et 2.

Le tableau 4, appelé carré gréco-latin, a été obtenu en superposant les trois tableaux précédents.

On voit qu'il est possible de grouper de quatre manières différentes les 16 cases de ce dernier tableau en formant chaque fois 4 classes telles que chacune d'entre elles contienne les 4 lettres A, B, C, D.

1) chaque classe est formée d'une colonne du tableau,

2) chaque classe est formée d'une rangée du tableau,

3) chaque classe est formée des cases contenant la même lettre grecque,

---

(1) H.H. SELIGER - Journ. of the Nat. Bur. of Standards 45 (Déc. 1950) p. 496.

4) chaque classe est formée des cases contenant le même chiffre romain.

Supposons que les lettres A, B, C, D représentent des échantillons radioactifs dont on a mesuré les activités à quatre reprises par des mesures de même durée pour un échantillon donné, les mesures ayant été effectuées dans l'ordre indiqué par les nombres inscrits entre parenthèses dans chaque case.

On voit immédiatement que chacune des classes du premier groupement est formée de mesures successives; une fluctuation anormale momentanée qui s'est produite au cours des mesures risque donc d'affecter plus particulièrement l'une de ces classes, si elle s'est manifestée pendant la durée de plusieurs mesures. Au contraire, les classes des trois autres groupements qui sont constituées de mesures non successives, réparties sur l'ensemble de l'intervalle de temps nécessaire pour effectuer les 16 mesures seront à peu près également affectées par une telle fluctuation (elles seront affectées de façon semblable si la fluctuation porte sur les 4 mesures d'une même colonne).

Une fois que les mesures ont été exécutées, on peut vérifier si la méthode adoptée a effectivement eu l'effet escompté, en appliquant le test de χ^2 à chacun des trois derniers groupements considérés : on s'assure que la somme des valeurs observées dans chaque classe égale à $n_A + n_B + n_C + n_D$ suit bien une loi de Poisson, ce qui conduit à effectuer, pour chaque groupement, un test de χ^2 avec 3 degrés de liberté. On peut montrer que dans les mêmes conditions, la somme des quantités X^2 (qui doivent être distribuées comme χ^2) correspondant aux 3 groupements doit être distribuée comme χ^2 avec 9 degrés de liberté. On appliquera donc le test de χ^2 (avec 9 degrés de liberté) à l'ensemble des 12 classes formées par les trois groupements.

Tableau 1

A	B	C	D
B	A	D	C
C	D	A	B
D	C	B	A

Tableau 2

α	β	γ	δ
γ	δ	α	β
δ	γ	β	α
β	α	δ	γ

Tableau 3

I	II	III	IV
IV	III	II	I
II	I	IV	III
III	IV	I	II

Tableau 4

(1) A α I 1	(5) B β II 1	(9) C γ III 1	(13) D δ IV 1
(2) B γ IV 2	(6) A δ III 2	(10) D α II 2	(14) C β I 2
(3) C δ II 3	(7) D γ I 3	(11) A β IV 3	(15) B α III 3
(4) D β III 4	(8) C α IV 4	(12) B δ I 4	(16) A γ II 4

Pour déterminer si des facteurs perturbateurs momentanés sont intervenus de façon sensible au cours de l'ensemble des mesures, on appliquera le test de χ^2 avec 3 degrés de liberté aux 4 classes du premier groupement, ou avec 12 degrés de liberté à l'ensemble des 16 classes correspondant aux quatre groupements.

MODALITES PRATIQUES D'UTILISATION -

Soit 4 échantillons A, B, C, D dont on veut déterminer les activités les unes par rapport aux autres.

On peut se fixer une même durée de comptage pour tous les échantillons, mais souvent on préfère choisir pour chaque échantillon une durée de mesure particulière qui correspond sensiblement à un même nombre de coups enregistrés pour tous les échantillons (erreurs relatives sur les activités du même ordre de grandeur).

Les mesures individuelles sont effectuées dans l'ordre suivant :

A B C D   B A D C   C D A B   D C B A

La valeur de l'activité de chaque échantillon est obtenue en faisant la moyenne des quatre valeurs obtenues au cours du comptage sur cet échantillon, que l'on divise par la durée de mesure correspondante.

Pour vérifier l'efficacité du plan d'expérience adopté, on classe les résultats des mesures (sans déduction du mouvement propre) dans un tableau de quatre rangées et de quatre colonnes, suivant l'ordre indiqué par les nombres entre parenthèses de chaque case du tableau 4.

On calcule ensuite la somme totale des 16 résultats que l'on divise par 4. Soit $\overline{N}$ la valeur ainsi obtenue. C'est la moyenne des valeurs $N = n_A + n_B + n_C + n_D$ des 4 classes de chaque groupement.

On calcule la somme N_c des 4 valeurs indiquées dans les cases d'une même colonne, et ceci pour chacune des 4 colonnes.

On calcule la somme N_r des 4 valeurs indiquées dans les cases d'une même rangée, et ceci pour chaque rangée.

On calcule la somme N_α des 4 valeurs indiquées dans les cases repérées par la même lettre grecque, et ceci pour les quatre lettre grecques.

On calcule la somme N_I des 4 valeurs indiquées dans les cases repérées par le même chiffre romain, et ceci pour les quatre chiffres romains.

On calcule le carré de l'écart de chacune de ces 16 sommes N par rapport à la valeur moyenne $\overline{N}$. On fait la somme de ces écarts que l'on divise par $\overline{N}$ et l'on obtient la valeur $X^2 = X_1^2 + X_2^2 + X_3^2 + X_4^2$ pour l'ensemble qui devrait être distribuée comme χ^2 avec 12 degrés de liberté (3 par groupement).

Si la valeur obtenue pour X^2 ne correspond pas à un écart significatif, on en conclut qu'il n'y a pas eu de perturbations extérieures significatives. Si l'écart est significatif, on applique le test de χ^2 avec 9 degrés de liberté à la valeur $X^2 = X_2^2 + X_3^2 + X_4^2$ obtenue à partir de l'ensemble des 12 sommes fournies par les rangées, les lettres grecques et les chiffres romains. Si ce nouveau test n'indique pas d'écart significatif, on peut en conclure que des perturbations extérieures ont eu une influence liée à la date des mesures (une ou plusieurs mesures successives ont subi une fluctuation due à des causes extérieures), mais que l'influence de ces perturbations momentanées a été de faible importance sur l'ensemble des mesures. (On peut, du reste, tester directement la signification de l'influence de ces perturbations en appliquant le test de χ^2 avec 3 degrés de liberté à la valeur X_1^2 obtenue à partir de l'ensemble des 4 sommes fournies par les colonnes). L'ordre adopté pour effectuer les mesures a donc permis de minimiser l'action de ces facteurs extérieurs.

Si ce nouveau test indique un écart significatif, on ne peut rien conclure sur la nature des perturbations extérieures qui ont affecté l'ensemble des résultats.

Exemple :

Nous avions à comparer 3 échantillons contenant du tantale, à un étalon à teneur connue en tantale, les échantillons et l'étalon ayant été irradiés ensemble dans un flux de neutrons lents. Des essais antérieurs avaient montré que les trois échantillons de provenance différente avaient des activités très voisines et du même ordre de grandeur que celle d'un étalon à 4% Ta.

En se basant sur les résultats d'une mesure rapide de 30 secondes, on a choisi d'effectuer les mesures pendant trois minutes pour les échantillons (assimilés aux lettres B, C et D du tableau) et trois minutes et demi pour l'étalon (assimilé à la lettre A), ce qui correspondait à environ 20. 000 coups.

Le tableau VII indique les résultats des mesures et les principaux calculs effectués. La première colonne indique l'ordre des mesures, et la deuxième colonne les échantillons mesurés. Les chiffres arabes, les lettres grecques et les chiffres romains des colonnes suivantes indiquent la classe dans laquelle est comptée la mesure considérée lors des différents groupements utilisés pour le calcul des écarts suivant la méthode du carré gréco-latin.

La valeur de X^2 pour les 4 groupements est de :

$$\frac{2\ 150\ 398}{82\ 383} = 26,10$$

La valeur de χ^2 avec 12 degrés de liberté qui correspond au seuil de signification de 5% est de 21, 03. Le test de χ^2 appliqué à l'ensemble des 4 groupements considérés met donc en évidence un écart significatif.

La valeur de X^2 correspondant aux trois premiers groupements (par rangées, par lettres grecques et chiffres romains) qui éliminent l'influence du moment des mesures est de :

$$\frac{1\ 243\ 872}{82\ 383} = 15,1$$

La valeur de χ^2 avec 9 degrés de liberté qui correspond au seuil de signification de 5% est 16, 9. Nous voyons que l'écart n'est plus significatif dans ce cas.

La valeur χ_1^2 correspondant au groupement par colonnes est 26,10 - 15, 10 = 11 alors que la valeur de χ^2 avec 3 degrés de liberté qui correspond au seuil de 5% est 7, 8; ce résultat indique nettement l'influence significative d'une perturbation liée à la date des mesures.

Dans l'exemple considéré, l'emploi de la méthode du carré gréco-latin a donc permis de minimiser l'influence d'une petite perturbation momentanée. L'examen des valeurs $N - \overline{N}$ et $(N - \overline{N})^2$ du tableau VII montre que cette perturbation s'est produite au cours des 4 dernières mesures.

Tableau VII

Application du test de χ^2 à des mesures effectuées suivant la méthode du carré gréco-latin

Ordre des mesures	Echantillons	Rangées	Lettres grecques	Chiffres romains	Nombre de coups enregistré
(1)	A	1	α	I	21 124
(2)	B	2	γ	IV	20 903
(3)	C	3	δ	II	21 316
(4)	D	4	β	III	19 518
(5)	B	1	β	II	20 884
(6)	A	2	δ	III	21 243
(7)	D	3	γ	I	19 761
(8)	C	4	α	IV	20 756
(9)	C	1	γ	III	20 738
(10)	D	2	α	II	19 531
(11)	A	3	β	IV	21 381
(12)	B	4	δ	I	20 772
(13)	D	1	δ	IV	19 519
(14)	C	2	β	I	20 686
(15)	B	3	α	III	20 459
(16)	A	4	γ	II	20 939
				Total	= 329 530

$$\overline{N} = \frac{329\ 530}{4} = 82\ 383$$

	$N = \Sigma n$	$N-\overline{N}$	$(N - \overline{N})^2$
1° ligne	82 265	118	13 924
2° ligne	82 363	20	400
3° ligne	82 917	534	285 156
4° ligne	81 985	398	158 404
			Total 457 884 = χ_1^2
α	81 870	513	263 169
β	82 469	86	7 396
γ	82 341	42	1 764
δ	82 850	467	218 089
			Total 490 418 = χ_2^2
I	82 343	40	1 600
II	82 670	287	82 369
III	81 958	425	180 625
IV	82 559	176	30 976
			Total 295 570 = χ_3^2
1° colon	82 861	478	228 484
2° colon	82 644	261	68 121
3° colon	82 422	39	1 521
4° colon	81 603	780	608 400
			Total 906 526 = χ_4^2

Remarque :

La méthode du carré gréco-latin ne s'applique pas aux mesures en précompte car elle est basée sur le fait que si X désigne le total $x_A + x_B + x_C + x_D$ des valeurs observées dans une classe d'un des 4 groupements, σ_X^2 n'est fonction que de la valeur moyenne $\underline{M}$ de X et peut, par suite, être estimée en fonction de la seule variable observée $\overline{X}$. Pour des mesures en précompte, si l'on considère l'activité a_n, $X = a_A + a_B + a_C + a_D$ et $\sigma_X^2 = \sigma^2_{a_A} + \sigma^2_{a_B} + \sigma^2_{a_C} + \sigma^2_{a_D} = \frac{A_A^2}{n_1} + \frac{A_B^2}{n_2} + \frac{A_C^2}{n_3} + \frac{A_D^2}{n_4}$

ce qui montre que même si on prend un même précompte pour toutes les mesures, σ_X^2 n'est plus fonction du seul $M = A_A + A_B + A_C + A_D$.

On notera qu'en considérant non plus l'activité a_n, mais la durée t de la mesure qui est telle que $\sigma_t^2 = \frac{[E(t)]^2}{n}$, on retrouve le même inconvénient.