Évaluation des incertitudes en TP par simulation informatique

Un point de vue statistique, mais utilisable en pratique

par S. GAZAIX
Lycée Kléber - 67000 Strasbourg

RÉSUMÉ

Encore des «calculs d'erreurs» ! Diront les malheureux qui ont passé des heures à rédiger des «comptes-rendus de TP», en déclinant des formules magiques du genre : $\frac{\Delta R}{R} = \frac{\Delta U}{U} + \frac{\Delta I}{I}$.

Encore ! Diront ceux qui ont fait l'effort d'assimiler la théorie statistique des variables aléatoires, mais qui ont conclu qu'il est illusoire de présenter cette théorie dans la plupart de nos classes.

Nous proposons ici d'utiliser le micro-ordinateur pour faire «sentir» aux étudiants le caractère foncièrement probabiliste du phénomène, et pour effectuer l'évaluation réaliste de l'incertitude sur une mesure du type $y = f\ (a, b, c)$ *où a, b et c son des grandeurs expérimentales.*

Nous ne rechercherons pas ici la rigueur mathématique, car nous voulons présenter une démarche accessible à des élèves de niveau bac.

1. UN EXEMPLE PARLANT

Nous choisirons l'exemple (peu original) où la grandeur y est le produit de deux grandeurs mesurées z et Z ; supposons que les valeurs retenues pour z et Z soient un ; reste à décrire le phénomène d'incertitude, notamment le fait que plusieurs mesures indépendantes de la même grandeur ne donnent pas le même résultat.

Le point fondamental est que cette incertitude est de nature probabiliste ; en conséquence, nous admettrons qu'on peut *simuler* cette incertitude en tirant *aléatoirement N* valeurs de z et Z, avec une valeur moyenne 1 et un écart-type proportionnel à l'incertitude. Nous choisirons une distribution gaussienne pour z et Z ; la valeur moyenne des N valeurs de z est $< z > = \frac{1}{N} \sum_{i=1}^{N} z_i$; l'écart-type σ de la distribution en z est défini par $\sigma^2 = \frac{1}{N} \sum (z_i - <z>)^2 = <z^2> - <z>^2$. Dans ce paragraphe, nous avons choisi un écart-type de 0,1 pour z et Z. La densité de probabilité pour une distribution gaussienne centrée sur $<z>$ et d'écart-type σ s'écrit :

$$p(z) = \frac{1}{\sigma \sqrt{2\pi}} \exp\left(-\frac{(z - <z>)^2}{2\,\sigma^2}\right)$$

rappelons que cela signifie que $p(z)\,dz$ représente la probabilité pour qu'un tirage aléatoire donne une valeur comprise entre z et $z + dz$.

La probabilité $P(\lambda)$ pour qu'un tirage de z soit dans l'intervalle :

$$[<z> - \lambda \cdot \sigma, <z> + \lambda \cdot \sigma]$$

est 68 % pour $\lambda = 1$, 95 % pour $\lambda = 1{,}96$, 99,7 % pour $\lambda = 3$.

Le lien entre incertitude et écart-type est donc le suivant : la probabilité pour qu'une mesure isolée de z se trouve dans l'intervalle centré autour de $<z>$, de demi-largeur 2 σ, est de l'ordre de 95 %. Si on veut une probabilité supérieure à 99 %, il faut un intervalle de demi-largeur 3 σ.

L'algorithme pour obtenir une suite de nombres répartis suivant une distribution gaussienne (ou normale) sera présenté en annexe. Le logiciel a été écrit en utilisant le langage Maple, qui présente l'inconvénient d'être très lent et gourmand en mémoire ; le même algorithme avec un langage compilé tel que Turbo Pascal a tourné vingt fois plus vite ! L'avantage de Maple est qu'on dispose d'une bibliothèque adaptée aux statistiques (tracés d'histogramme, distributions statistiques usuelles déjà programmées...).

L'histogramme (avec dix intervalles) des valeurs de z pour deux mille tirages (moyenne 1, écart-type 0,1) est représenté sur la figure 1 ; les valeurs portées en ordonnée représentent le nombre de tirages tels que la valeur de z s'est située dans l'intervalle considéré.

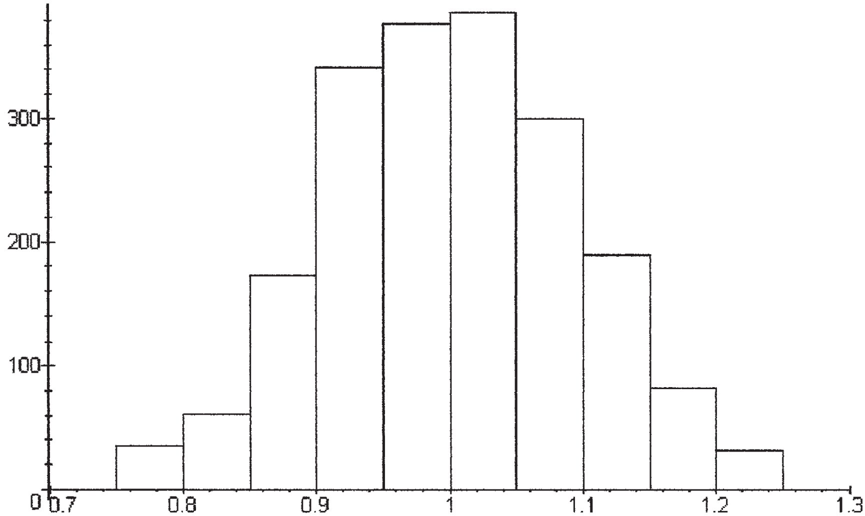

Figure 1 : Histogramme de z (deux mille tirages suivant une loi gaussienne d'écart-type 0,1).

Le résultat du calcul de $y = z \cdot Z$ est présenté sur la figure 2.

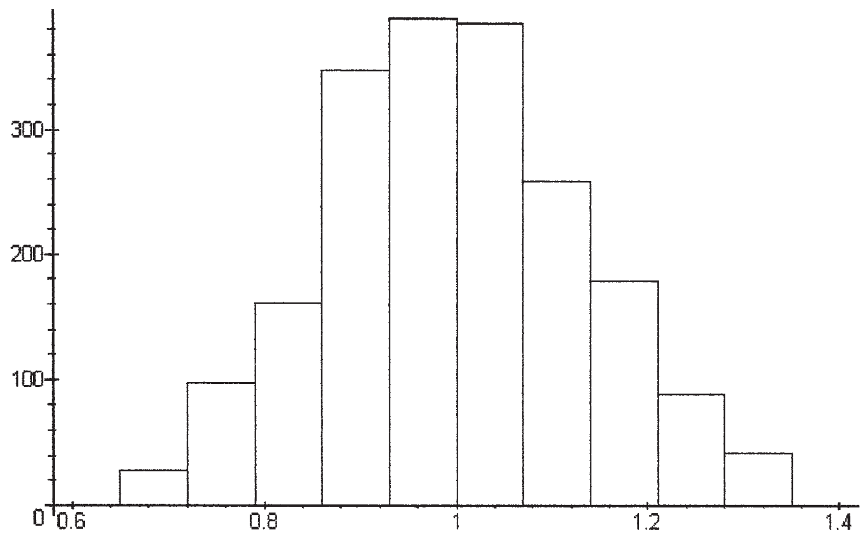

Figure 2 : Histogramme de $z \cdot Z$ ($N = 2000$).

Le point fondamental est que l'écart-type calculé sur $z \cdot Z$ est de 0.14, alors que l'application de la «formule magique» $\frac{\Delta\ (z \cdot Z)}{z \cdot Z} = \frac{\Delta z}{z} + \frac{\Delta Z}{Z}$ donnerait 0.2.

La raison en est simple : il est peu *probable* que deux mesures de z et Z présentent simultanément l'erreur maximale, dans le même sens ; bien sûr, cela peut arriver, et il n'est pas aberrant de se «couvrir» en prenant comme incertitude maximale 0.2. Il y a un choix à faire, qui sera différent suivant l'utilisation de la mesure.

Remarque

L'écart-type 0.14 vérifie la loi dite «d'addition des variances relatives» :

$$\left(\frac{\Delta y}{y}\right)^2 = \left(\frac{\Delta z}{z}\right)^2 + \left(\frac{\Delta Z}{Z}\right)^2$$

valable si les variables aléatoires z et Z sont indépendantes ; mais répétons que notre propos est de présenter le point de vue statistique de manière «expérimentale» par la simulation informatique, sans insister sur la théorie des probabilités.

2. MESURE DE L'INDICE D'UN PRISME

Ce TP de niveau Bac + 1 est incontournable, ne serait ce que parce que les mesures avec plusieurs chiffres significatifs ne sont pas si fréquentes !

Considérons donc $n = \frac{\sin\left(\frac{A+D}{2}\right)}{\sin\left(\frac{A}{2}\right)}$; A vaut 60° avec un écart-type 2', et D 45° avec un écart-type de 3'. Effectuons le même type de simulation (figure 3). Rappelons le lien entre incertitude et écart-type : si on assimile l'incertitude au double de l'écart-type, alors l'incertitude représente *l'intervalle de confiance* à 95,5 %.

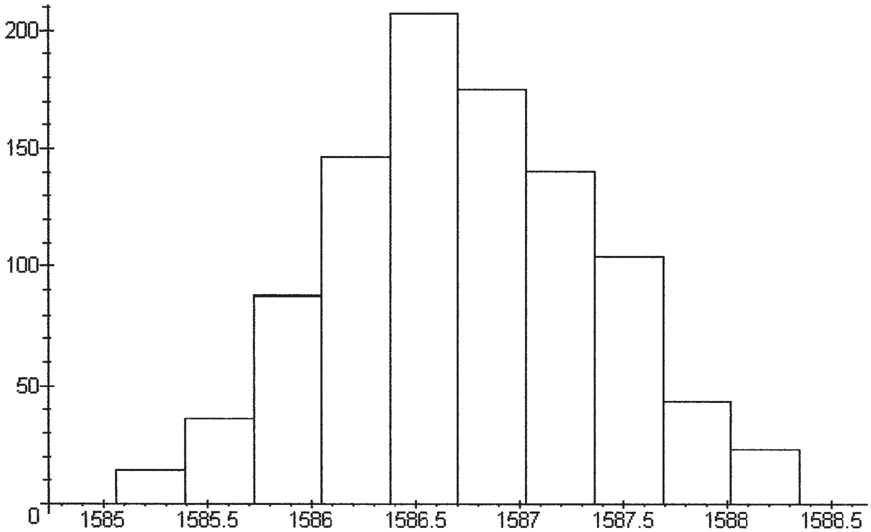

Figure 3 : Histogramme de l'indice * 1*e*3 (N = 1000) incertitude 3' sur *Dm* et 2' sur *A*.

Le programme permet de calculer la valeur moyenne des N ($N = 1000$) valeurs de l'indice n : $<n> = 1.5867$; ainsi que l'écart-type sur ces N valeurs : $6.6 \cdot 10^{-4}$; l'incertitude relative sur n est $4.1 \cdot 10^{-4}$. Notons que la relation simpliste :

$$<n> = \frac{\sin\left(\frac{<A> + <D>}{2}\right)}{\sin\left(\frac{<A>}{2}\right)}$$

est en général fausse, même si elle est acceptable pour de faibles écart-types relatifs.

Le gros avantage de cette approche par simulation est qu'on n'est plus obligé de recourir au fastidieux «calcul de petites variations» :

$$\left|\delta n\right| \leq \left|\frac{\partial n}{\partial A}\right| \cdot \left|\delta A\right| + \left|\frac{\partial n}{\partial D}\right| \cdot \left|\delta D\right|$$

En plus, la simulation permet de traiter le cas où les incertitudes relatives ne sont pas très inférieures à un, comme nous allons le montrer dans l'exemple suivant.

3. ANALYSE STATISTIQUE D'UNE MESURE EN PHYSIQUE DES HAUTES ÉNERGIES

Il est fréquent d'avoir à déterminer une grandeur en fonction d'un grand nombre de paramètres. Nous présentons ici une expérience conduite par Reines, Sobel et Pasierb en vue de savoir si le neutrino est stable, la réponse à cette question étant d'un grand intérêt du point de vue théorique. L'expérience consiste à déterminer le rapport *R* entre deux sections efficaces. Le neutrino est instable si R est inférieur à 0.42.

R s'écrit :

$$R = \frac{a}{\dfrac{d}{k^2 e}\,(b-c) - 2\left(1 - \dfrac{k^2 d}{ke}\right) a}$$

Les mesures ont donné :

$$\begin{array}{lll} a = 3.84 \pm 1.3 & c = 9.5 \pm 3 & e = 0.32 \pm 0.002 \\ b = 74 \pm 4 & d = 0.112 \pm 0.009 & k = 0.89 \end{array}$$

Pour évaluer l'incertitude sur *R*, il n'est pas possible d'utiliser l'approximation linéaire :

$$\delta R \approx \frac{\partial R}{\partial a}\,\delta a + \frac{\partial R}{\partial b}\,\delta b + \frac{\partial R}{\partial c}\,\delta c + \frac{\partial R}{\partial d}\,\delta d + \frac{\partial R}{\partial e}\,\delta e$$

car l'incertitude relative sur *a* et *c* est bien trop grande. Le seul moyen d'obtenir une évaluation fiable est la simulation statistique (appelée méthode Monte-Carlo dans les revues). Nous présentons sur la figure 4 l'histogramme obtenu.

On peut calculer directement la valeur moyenne $<R>$, en appliquant la relation $<R> = \dfrac{1}{N} \displaystyle\sum_{i=1}^{N} R_i$ ainsi que l'écart-type, sans approximations ni calculs difficiles. La courbe continue est la distribution normale construite à partir de la moyenne et de l'écart-type calculés sur la distribution *R* ; il est intéressant de noter que la distribution *R* n'est pas une gaussienne, alors que les distributions *a*, *b*, *c*, *d* et *e* le sont.

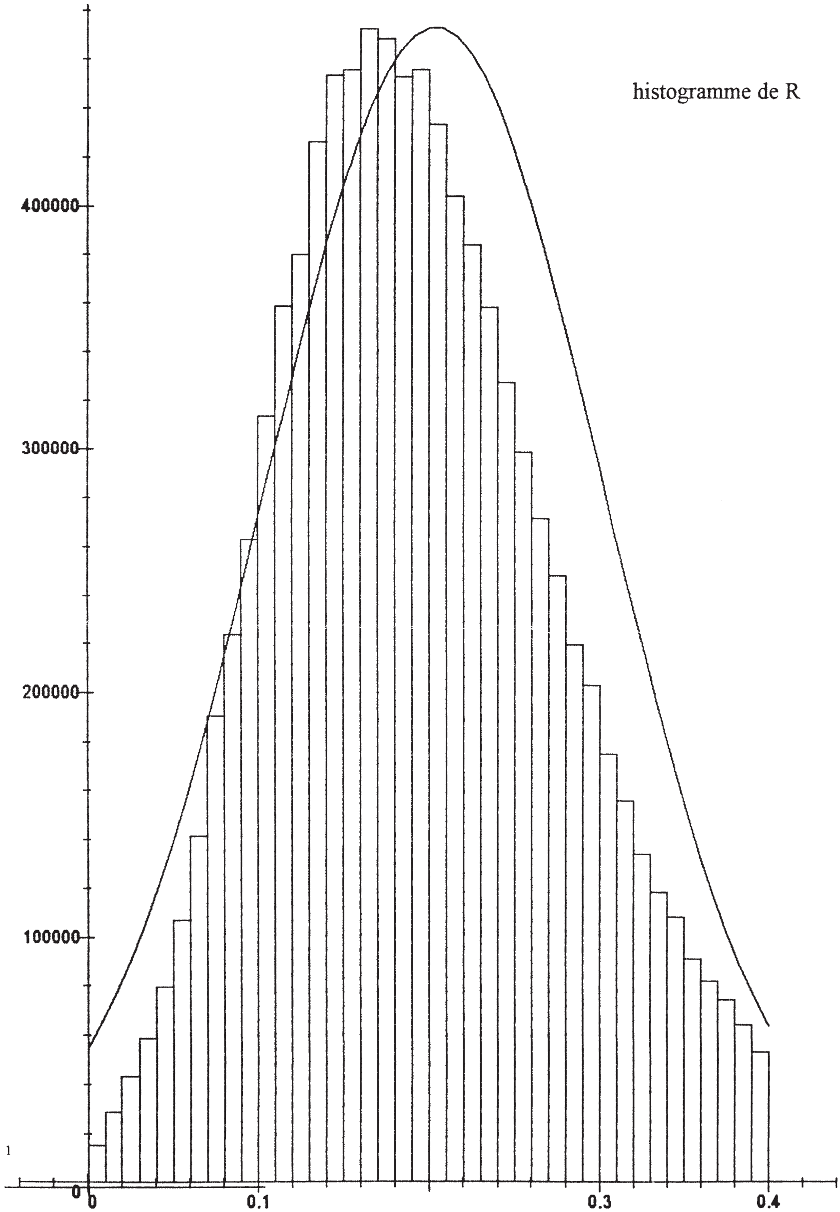

Figure 4 : Neutrino

CONCLUSION

Le but de cet article est de montrer qu'il est possible de trouver un compromis entre les deux approches généralement rencontrées lors de l'exploitation d'une série de mesures en TP :

– «Je sais bien que l'approche traditionnelle $\frac{\Delta R}{R} = \frac{\Delta U}{U} + \frac{\Delta I}{I}$ est trop pessimiste, mais c'est la méthode la plus simple».

– «J'utilise une approche probabiliste rigoureuse, mais la difficulté mathématique rebute les élèves, et les calculs n'aboutissent pas toujours (par exemple dans le cas du paragraphe 3 : $R = R\ (a, b, c, d, e)$)».

Dans l'exemple de la mesure de l'indice d'un prisme, on peut imaginer que les élèves puissent utiliser le logiciel «clé en main» pour déterminer l'écart-type sur la grandeur n en fonction des incertitudes sur A et D.

Signalons que la mise en œuvre de ce logiciel a donné lieu à une séance de TD informatique (niveau Math Spé) ; le but était de fournir aux élèves une vision intuitive des phénomènes probabilistes, en complément de l'approche abstraite développée dans les cours de mathématiques.

REMERCIEMENTS

Nicolas WEIBEL a contribué à la partie informatique, dans le cadre de son TIPE (année scolaire 1996/1997).

BIBLIOGRAPHIE

– H. GIÉ et R. MOREAU : *«Le calcul des incertitudes»* - BUP n° 691.
– P.-Y. NYZOU : *«Interprétation statistique des erreurs de mesure»* - BUP n° 679.
– F. JAMES : *«Determining the statistical significance of experimental results»* - Cern-Data Handling Division, february 1981.
– P. JAFFARD : *«Méthodes de la statistique et du calcul des probabilités»* - Masson, 1990.
– B. ESCOUBES : *«Cours de DEA de physique nucléaire»* - Strasbourg, 1994.
– Press et al : *«Numerical Recipes»* - Cambridge University Press, 1989.

FICHIER MAPLE

Il suffit de le demander à l'adresse :

`SGazaix@aol.com`

Annexe
Tirage aléatoire suivant une loi de probabilité normale

Tous les langages proposent une procédure (Random ou équivalent) permettant un tirage pseudo-aléatoire d'un nombre compris entre 0 et 1, avec une distribution uniforme, c'est-à-dire que la probabilité pour que le nombre x appartienne à l'intervalle dx est égale à dx entre 0 et 1, et 0 en dehors. A partir de cette distribution uniforme, il est possible de construire toutes les autres distributions. Pour la distribution normale, le moyen le plus rapide est le suivant : u_1 et u_2 étant deux nombres tirés aléatoirement entre 0 et 1 avec une loi de distribution uniforme, on montre que :

$$z_1 = \sin\ (2\pi\ u_1) \cdot \sqrt{-\ 2\ \ln\ (u_2)}$$

et :

$$z_2 = \cos\ (2\pi\ u_1) \cdot \sqrt{-\ 2\ \ln\ (u_2)}$$

sont deux variables aléatoires centrées sur zéro, avec une loi de distribution gaussienne d'écart-type 1. A partir de la variable z_1, on peut construire la variable z, centrée sur z_0, d'écart-type σ : $z = \sigma \cdot z_1 + z_0$.