INTRODUCTION

Un exemple de plan d'expérience : la pesée de trois objets avec le maximum de précision et le minimum d'expériences

Nous nous proposons de présenter la notion de plan d'expérience à propos du problème suivant. On veut déterminer les poids de trois objets O_1, O_2, O_3 à l'aide d'une balance où deux plateaux

1) en effectuant le minimum de pesées
2) en obtenant le maximum de précision

Analysons d'abord l'opération "pesée" et supposons que la balance comporte une aiguille permettant de faire la lecture. Lorqu'on place une objet sur la balance, la lecture $\hat{y}$ ne correspond pas à la valeur théorique y en raison de l'erreur expérimentale. Lorqu'on répète plusieurs fois la même pesée $\hat{y}$ varie mais la moyenne des lectures faites tend vers y. On dit que $\hat{y}$ est un estimateur de y :

$$\hat{y} = y + \varepsilon$$

ε est l'erreur expérimentale, sa valeur moyenne est nulle. Comme toutes les erreurs expérimentales ε est une variable aléatoire obéissant à une loi normale $\mathcal{N}[0,\sigma]$

Si l'on répète plusieurs fois des pesées différentes, notées i et j on peut écrire $\hat{y}_i = y_i + \varepsilon_i$

$$\hat{y}_j = y_j + \varepsilon_j$$

ε_i et ε_j sont deux variables aléatoires et on peut écrire :

$E(\varepsilon_i \varepsilon_j) = 0$
$E(\varepsilon_i) = E(\varepsilon_j) = 0$
$\text{var}(\varepsilon_i) = \text{var}(\varepsilon_j) = \sigma^2$

$\text{cov}(\hat{y}_i, \hat{y}_j) = 0$
qui entraine $E(\hat{y}_i) = y_i \quad E(\hat{y}_j) = y_j$
$\text{var}(\hat{y}_i) = \text{var}(y_j) = \sigma^2$

Nous retiendrons simplement les résultats suivants :

- l'espérance mathématique (valeur moyenne) d'une somme de variables aléatoires est égale à la somme des espérances mathématiques de ses termes.
- lorsque, quelque soit $i \neq j$, $cov(y_i, y_j) = 0$ la variance d'une somme est la somme des variances de ses termes. Rappelons aussi que si $var(\hat{y}) = \sigma^2$, $var(a\hat{y}) = a^2\sigma^2$

Nombre minimum d'opérations pour peser trois objets

Nous supposons que lorsque la balance n'est pas chargée, l'aiguille ne marque pas exactement 0 mais qu'on peut faire une lecture $\hat{y}_0$ variable aléatoire, $\hat{y}_0 = y_0 + \varepsilon_0$ correspondant à la pesée fictive d'un objet O_0. Pour chacun des objets O_1, O_2, O_3 il faudra faire une mesure. Le nombre minimum de pesées est de 4. Par souci d'homogénéité dans les notations, nous les noterons 0, 1, 2, 3. Les poids seront notés : p_0, poids de l'objet fictif O_0, puis $p_1(O_1)$, $p_2(O_2)$, $p_3(O_3)$

Quatre expérimentateurs vont résoudre le problème posé, en faisant chacun quatre pesées et nous évaluerons dans chaque cas la précision obtenue : nous allons donc comparer quatre plans d'expérience. Pour désigner chaque expérience, nous convenons de noter + un objet placé sur le plateau de droite, et de compter positivement son poids, − un objet placé sur le plateau de gauche et de retrancher son poids, et d'indiquer 0 si l'objet n'est pas sur la balance, son poids n'intervenant pas dans la mesure. L'objet fictif O_0 est supposé placé sur le plateau de droite.

Experimentateur n° 1 Il utilise la méthode - le plan - de "Mr. tout le monde"

Il commence par lire la position de l'aiguille lorsque la balance n'est pas chargée puis il pèse successivement les trois objets en les mettant sur le plateau de droite. Sa démarche est parfaitement définie par le tableau ci dessous appelé matrice d'expérience D_1

Experience n°	O_1	O_2	O_3	lectures de l'aiguille
0	0	0	0	$\hat{y}_0$
1	1	0	0	$\hat{y}_1$
2	0	1	0	$\hat{y}_2$
3	0	0	1	$\hat{y}_3$

matrice d'expérience D_1

Ce tableau comporte autant de lignes que d'expériences, et autant de colonnes que d'objets à peser. Nous avons indiqué en outre les résultats $\hat{y}$ obtenus pour chacune des pesées.

On peut écrire les égalités suivantes

Exp n° 0 $\quad \hat{y}_0 = \hat{r}_0$

Exp n° 1 $\quad \hat{y}_1 = \hat{r}_0 + \hat{r}_1$

Exp n° 2 $\quad \hat{y}_2 = \hat{r}_0 + \hat{r}_2$

Exp n° 3 $\quad \hat{y}_3 = \hat{r}_0 + \hat{r}_3$

Système de quatre équations à 4 inconnues d'une résolution élémentaire

$$\hat{r}_0 = \hat{y}_0$$
$$\hat{r}_1 = \hat{y}_1 - \hat{y}_0$$
$$\hat{r}_2 = \hat{y}_2 - \hat{y}_0$$
$$\hat{r}_3 = \hat{y}_3 - \hat{y}_0$$

En appliquant les règles de probabilité données en p 2 il vient

$E(\hat{p}_0) = E(\hat{y}_0) = y_0 = p_0$

$E(\hat{p}_1) = E(\hat{y}_1 - \hat{y}_0) = E(\hat{y}_1) - E(\hat{y}_0) = p_0 + p_1 - p_0 = p_1$

et de même $E(\hat{p}_2) = p_2$ $E(\hat{p}_3) = p_3$

$\hat{p}_1$ $\hat{p}_2$ et $\hat{p}_3$ sont des estimateurs non biaisés des poids p_1, p_2 et p_3

Calculons maintenant les variances, qui traduisent la dispersion des résultats. Il vient immédiatement

$$\begin{aligned} var(\hat{p}_0) &= \sigma^2 \\ var(\hat{p}_1) &= 2\sigma^2 \\ var(\hat{p}_2) &= 2\sigma^2 \\ var(\hat{p}_3) &= 2\sigma^2 \end{aligned}$$

Rappelons que σ est l'écart type des lectures de la balance. On peut donc dire, dans le cas de ce plan d'expérience que si les pesées sont faites avec une erreur standard de 1g, les poids sont déterminés avec une erreur standard de 1,4g.

- Avant de passer à un autre plan d'expérience nous allons donner pour celui ci, bien que ce soit parfaitement inutile dans ce cas, une solution matricielle qui nous aidera par la suite. Le système des quatre equations à resoudre s'écrit :

$$P_1 \times \hat{p} = \hat{y} \quad \text{qui se résoud en } \hat{p} = P_1^{-1} \hat{y}$$

$\hat{p}$ designe le vecteur colonne des inconnues $\hat{p}_0, \hat{p}_1, \hat{p}_2, \hat{p}_3$

P_1 la matrice de leurs coefficients dans les quatre équations et $\hat{y}$ le vecteur colonne des résultats. P_1^{-1} est la matrice inverse de P. Ecrivons la matrice P_1

$$P_1 = \left|\begin{array}{cccc} 1 & 0 & 0 & 0 \\ 1 & 1 & 0 & 0 \\ 1 & 0 & 1 & 0 \\ 1 & 0 & 0 & 1 \end{array}\right|$$

matrice d'experience

On voit qu'elle se construit facilement à partir de la matrice d'experiences en ajoutant, à gauche une colonne de 1 (coefficients de $\hat{p}_0$)

La matrice P_1^{-1}, et plus généralement l'inverse d'une matrice 4×4 se calcule très facilement à l'aide d'une petite machine à calcul.

Dans le cas présent

$$\begin{vmatrix} \hat{r}_0 \\ \hat{r}_1 \\ \hat{r}_2 \\ \hat{r}_3 \end{vmatrix} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ -1 & 1 & 0 & 0 \\ -1 & 0 & 1 & 0 \\ -1 & 0 & 0 & 1 \end{vmatrix} \times \begin{vmatrix} \hat{y}_0 \\ \hat{y}_1 \\ \hat{y}_2 \\ \hat{y}_3 \end{vmatrix} \Rightarrow \begin{aligned} \hat{r}_0 &= \hat{y}_0 \\ \hat{r}_1 &= \hat{y}_1 - \hat{y}_0 \\ \hat{r}_2 &= \hat{y}_2 - \hat{y}_0 \\ r_3 &= \hat{y}_3 - \hat{y}_0 \end{aligned}$$

<u>Expérimentateur n° 2</u> Comme le précédent il commence par observer la balance non chargée (exp 0). Il place ensuite deux par deux les objets sur le plateau de droite. Avec trois objets on peut constituer 3 couples différents d'où la matrice d'expérience D_2 et la matrice P_2. Le calcul de P_2^{-1} donne immédiatement le résultat

$$D_2 = \begin{array}{c} \begin{matrix} O_1 & O_2 & O_3 \end{matrix} \\ \begin{vmatrix} 0 & 0 & 0 \\ 1 & 1 & 0 \\ 1 & 0 & 1 \\ 0 & 1 & 1 \end{vmatrix} \end{array} \qquad P_2 = \begin{vmatrix} 1 & 0 & 0 & 0 \\ 1 & 1 & 1 & 0 \\ 1 & 1 & 0 & 1 \\ 1 & 0 & 1 & 1 \end{vmatrix} \qquad P_2^{-1} = \begin{vmatrix} 1 & 0 & 0 & 0 \\ -0,5 & 0,5 & 0,5 & -0,5 \\ -0,5 & 0,5 & -0,5 & 0,5 \\ -0,5 & -0,5 & 0,5 & 0,5 \end{vmatrix}$$

Les estimateurs $\hat{r}$ sont calculés ci dessous :

$$\hat{r}_0 = \hat{y}_0$$

$$\hat{r}_1 = \frac{1}{2}(-\hat{y}_0 + \hat{y}_1 + \hat{y}_2 - \hat{y}_3)$$

$$\hat{r}_2 = \frac{1}{2}(-\hat{y}_0 + \hat{y}_1 - \hat{y}_2 + \hat{y}_3)$$

$$\hat{r}_3 = \frac{1}{2}(-\hat{y}_0 - \hat{y}_1 + \hat{y}_2 + \hat{y}_3)$$

On vérifie facilement que ce sont des estimateurs non biaisés des poids r_i et que $\mathrm{var}(\hat{r}_0) = \mathrm{var}(\hat{r}_1) = \mathrm{var}(\hat{r}_2) = \mathrm{var}(\hat{r}_3) = \sigma^2$

En utilisant ce plan, on a amélioré la précision. Si l'erreur standard sur une pesée est de 1 g, l'erreur standard sur la détermination d'un poids est aussi de 1 g.

Experimentateur n° 3 A-t-il appris, par induction, qu'en mettant sur la balance deux objets au lieu d'un on augmentait la précision d'un facteur $\sqrt{2}$? Toujours est-il qu'il place systématiquement, et dès le début, les trois objets sur les plateaux. Il réalise le plan modélisé par la matrice d'expérience D_3.
P_3 est la matrice des coefficients des variables μ que l'inverse P_3^{-1} permet de calculer à partir des mesures expérimentales y

$$D_3 = \begin{array}{c} \begin{matrix} O_1 & O_2 & O_3 \end{matrix} \\ \begin{vmatrix} -1 & -1 & -1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{vmatrix} \end{array} \qquad P_3 = \begin{vmatrix} 1 & -1 & -1 & -1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix} \qquad P_3^{-1} = \begin{vmatrix} -0,5 & 0,5 & 0,5 & 0,5 \\ -0,5 & 0,5 & 0 & 0 \\ -0,5 & 0 & 0,5 & 0 \\ -0,5 & 0 & 0 & 0,5 \end{vmatrix}$$

$$\hat{\mu}_0 = \frac{1}{2}\left(-\hat{y}_0 + \hat{y}_1 + \hat{y}_2 + \hat{y}_3\right)$$
$$\hat{\mu}_1 = \frac{1}{2}\left(-\hat{y}_0 + \hat{y}_1\right)$$
$$\hat{\mu}_2 = \frac{1}{2}\left(-\hat{y}_0 + \hat{y}_2\right)$$
$$\hat{\mu}_3 = \frac{1}{2}\left(-\hat{y}_0 + \hat{y}_3\right)$$

On voit immédiatement que $\hat{\mu}_i$ est un estimateur non biaisé de μ_i et que $\mathrm{var}(\hat{\mu}_0) = \sigma^2$, $\mathrm{var}(\hat{\mu}_1) = \mathrm{var}(\hat{\mu}_2) = \mathrm{var}(\hat{\mu}_3) = \frac{\sigma^2}{2}$
La précision est améliorée. Si l'erreur standard sur une pesée est de 1 g, l'erreur standard sur un poids est de 0,5 g (μ_0 ne nous intéresse pas particulièrement)

Expérimentateur n° 4 L'"intuition" de l'expérimentateur précédent a donné un bon résultat, mais il n'y a plus rien à faire dans cette voie, le maximum d'objets étant sur la balance On peut seulement changer une ou plusieurs expériences (et y a 8 façons de placer 3 objets à la fois sur les plateaux. Le quatrième expérimentateur choisit une solution que rien ne justifie à priori en modifiant la première pesée (exp n° 0)

Au lieu de placer les 3 objets sur le plateau de gauche, il les mit sur le plateau de droite. D'où la matrice d'expérience D_4 les matrices P_4 et P_4^{-1} et le résultat.

$$D_4 = \begin{vmatrix} 1 & 1 & 1 \\ 1 & -1 & -1 \\ -1 & 1 & -1 \\ -1 & -1 & 1 \end{vmatrix} \qquad P_4 = \begin{vmatrix} 1 & 1 & 1 & 1 \\ 1 & 1 & -1 & -1 \\ 1 & -1 & 1 & -1 \\ 1 & -1 & -1 & 1 \end{vmatrix} \qquad P_4^{-1} = \begin{vmatrix} 0,25 & 0,25 & 0,25 & 0,25 \\ 0,25 & 0,25 & -0,25 & -0,25 \\ 0,25 & -0,25 & 0,25 & -0,25 \\ 0,25 & -0,25 & -0,25 & 0,25 \end{vmatrix}$$

$$\hat{p}_0 = \frac{1}{4}\left(\hat{y}_0 + \hat{y}_1 + \hat{y}_2 + \hat{y}_3\right)$$
$$\hat{p}_1 = \frac{1}{4}\left(\hat{y}_0 + \hat{y}_1 - \hat{y}_2 - \hat{y}_3\right)$$
$$\hat{p}_2 = \frac{1}{4}\left(\hat{y}_0 - \hat{y}_1 + \hat{y}_2 - \hat{y}_3\right)$$
$$\hat{p}_3 = \frac{1}{4}\left(\hat{y}_0 - \hat{y}_1 - \hat{y}_2 + \hat{y}_3\right)$$

Les valeurs $\hat{p}_i$ sont des estimateurs non biaisés de p_i et $\operatorname{var} \hat{p}_i = \frac{\sigma^2}{4}$ $i = 0, 1, 2, 3$. Encore mieux !

Si les pesées sont faites avec une erreur standard de 1 g, l'erreur sur les poids est de 0,25 g. Peut-on faire encore mieux ?

Reprenons le problème posé : peser trois objets avec le minimum de pesées et, sous cette condition, le maximum de précision.

Nous avons vu que le nombre minimum de pesées est de 4.

On montre d'autre part que pour n'importe quel plan de pesée utilisant une balance à 2 plateaux, K objets et N pesées $(N \geq K+1)$ $\operatorname{var}(\hat{p}_i) \geq \frac{\sigma^2}{N}$.

La condition nécessaire et suffisante pour que $\operatorname{var}(\hat{p}_i) = \frac{\sigma^2}{N}$ est que l'on ait $\underline{P}'\underline{P} = N I_{K+1}$

N = nombre de pesées $(N > K+1)$

K = nombre d'objets

$\underline{P}$ = matrice des coefficients des variables $\hat{p}_i$

$\underline{P}'$ = matrice transposée de $\underline{P}$

I_{K+1} = matrice unité de rang $K+1$

En quatre pesées, le dernier expérimentateur a obtenu la variance minimale $\frac{\sigma^2}{4}$ et on vérifie facilement la relation :

$P_4 P_4' = 4 I_4$

Ce problème des pesées peut servir de modèle à d'autres situations expérimentales mais nous arrêterons là cette introduction.

Nous retiendrons simplement que lorsqu'un problème est clairement posé, il y a interet à prévoir les expériences qui permettront de le résoudre au mieux. Ceci exclut les expériences faites au hasard et introduit la notion de "plan"

Nous retiendrons également que l'"intuition" n'est pas un guide suffisant (voir le passage expérimentateur 3 $\Rightarrow$ expérimentateur 4).

Nous avons également précisé ce qu'est une matrice d'expérience, procédé systématique pour représenter les expériences à effectuer. [matrice D (N×K) N expériences, K facteurs pouvant varier]

Nous retiendrons enfin la propriété des matrices carrées telles que P_4 : $P_4 P_4' = 4 I_4$ qui se généralise en :

$$\boxed{X_N X_N' = N I_N}$$

Les matrices X_N vérifiant cette relation sont appelées matrices de Hadamard et nous les retrouverons dans la suite