Le problème des pesées

1. Description du problème

Toute expérimentation est faite pour répondre à une ou plusieurs questions dans un domaine donné. Le but de la planification expérimentale est de choisir astucieusement des essais qui permettront de répondre aux questions posées avec une précision maximum.

Par un exemple simple, nous allons illustrer l'importance de la planification expérimentale et son implication sur la précision des résultats qu'on peut retirer de l'expérimentation.

Le problème consiste à déterminer la masse de trois objets A, B et C au moyen d'une balance à deux plateaux.

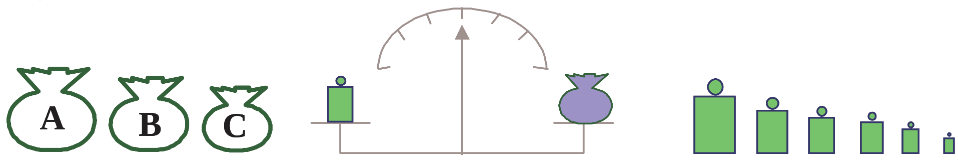

Figure 1 : Problème des pesées : le matériel

Un essai (ou pesée) consiste à poser aucun, un ou plusieurs objets sur un ou les deux plateaux de la balance puis à établir l'équilibre en ramenant l'aiguille sur le "zéro" à l'aide de masses marquées. Nous supposerons que l'expérimentateur dispose d'un budget de 400 € pour faire cette étude et que chaque pesée coûte 100 €. Les masses des 3 objets doivent donc être déterminés en 4 pesées.

Nous adopterons les notations et hypothèses suivantes:

- Nous noterons **Y_i, le résultat de la ième pesée**. Ce résultat représente la valeur totale des masses qu'il a fallu mettre sur le plateau de droite de la balance pour l'équilibrer. Si les masses sont mises sur le plateau de gauche, Y_i sera affecté d'un signe négatif.
- Nous admettrons que l'expérimentateur fait des **erreurs de lecture** ; c'est-à-dire que l'aiguille n'est pas toujours exactement ramenée sur zéro. Chaque pesée Y_i sera donc entachée d'une erreur notée ε_i
- Nous supposerons que l'ordre de grandeur de l'erreur est constante quelque soit l'objet à peser :Variance(ε_i) = σ^2 et que les résultats ne sont pas liés d'un essai à l'autre : Covariance(Y_i,Y_j)=Cov(ε_i, ε_j)=0
- Finalement, nous admettrons qu'en l'absence d'objet sur la balance, l'aiguille ne pointe pas forcément sur "zéro" : une erreur systématique constante inconnue est introduite à chaque pesée.

Ces hypothèses peuvent-être réécrites sous la forme suivante :

$$Y_i = \pi_0 + \pi_i + \varepsilon_i \qquad E(\varepsilon_i) = 0, V(\varepsilon_i) = \sigma^2 \text{ et } \mathrm{cov}(\varepsilon_i, \varepsilon_j) = 0 \qquad \textbf{(1)}$$

où π_0 est l'erreur systématique et π_i est la valeur théorique de la masse qu'il faudrait mettre sur la balance pour l'équilibrer parfaitement à zéro.

Nous allons proposer 4 stratégies expérimentales et montrer que selon la stratégie choisie, les résultats peuvent être très différents.

2. Stratégie 1

La première stratégie consiste à travailler comme la plupart des expérimentateurs ont l'habitude de le faire c'est-à-dire en pesant un objet à la fois, la première pesée étant utilisée pour déterminer l'erreur systématique (voir figure 2). Notons Y_1, Y_2, Y_3 et Y_4, les résultats de ces 4 pesées.

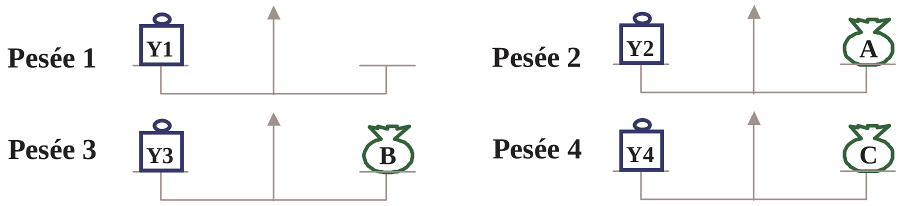

Figure 2 : Problème des pesées : stratégie 1

On peut présenter ces quatre pesées sous forme d'une **matrice d'expérience** formée de 4 lignes (4 essais) et 3 colonnes (3 objets à peser) :

Stratégie 1	Objet A	Objet B	Objet C
Pesée 1	0	0	0
Pesée 2	1	0	0
Pesée 3	0	1	0
Pesée 4	0	0	1

où 0 indique que l'objet n'est pas sur la balance
1 indique que l'objet est sur le plateau de gauche de la balance
-1 indique que l'objet est sur le plateau de droite de la balance

Le **modèle** sous-jacent à ces 4 pesées est le suivant :

$$\begin{aligned} Y_1 &= \pi_0 + \varepsilon_1 \\ Y_2 &= \pi_0 + \pi_A + \varepsilon_2 \\ Y_3 &= \pi_0 + \pi_B + \varepsilon_3 \\ Y_4 &= \pi_0 + \pi_C + \varepsilon_4 \end{aligned} \qquad \textbf{(2)}$$

dans lequel π_0 représente l'erreur systématique et π_A, π_B, π_C les masses des 3 objets.

Les **estimateurs** des masses peuvent être calculés en résolvant le système d'équation (2) dans lequel, ne connaissant pas les erreurs de mesure, on ne pourra que les considérer comme nulles. On obtient :

$$\begin{aligned} P_0 &= \hat{\pi}_0 = Y_1 \\ P_A &= \hat{\pi}_A = Y_2 - Y_1 \\ P_B &= \hat{\pi}_B = Y_3 - Y_1 \\ P_C &= \hat{\pi}_C = Y_4 - Y_1 \end{aligned} \qquad \textbf{(3)}$$

Nous pouvons maintenant nous demander quelle est la précision de ces estimateurs en tenant compte du fait que chaque pesée Y_i est entachée d'une erreur ε_i de variance σ^2. Pour calculer les variances des estimateurs P_i, il suffit d'appliquer la formule bien connue de la variance d'une somme :

$$V(aY_i \pm bY_j) = a^2 V(Y_i) + b^2 V(Y_j) \pm 2ab\,\mathrm{Cov}(Y_i, Y_j) \qquad \textbf{(4)}$$

Les pesées, étant indépendantes, on obtient :

$$V(P_0) = V(Y_1) = \sigma^2 \quad V(P_A) = V(Y_2 - Y_1) = 2\sigma^2 \quad V(P_B) = V(P_C) = 2\sigma^2 \qquad \textbf{(5)}$$

Le but de l'expérimentation est d'obtenir des estimations précises des masses des objets. Nous verrons que suivant la stratégie expérimentale choisie la précision des estimateurs P_i peut varier.

En gardant cette même stratégie, étudions ce qu'il faudrait faire pour obtenir une meilleure précision. La solution est simple, il suffit de répéter les pesées plusieurs fois. Si l'on répète par exemple les 4 pesées une seconde fois, on obtiendra les estimateurs suivants pour les masses des objets :

$$\begin{aligned} P_0 &= \hat{\pi}_0 = (Y_1 + Y_{1'})/2 \\ P_A &= \hat{\pi}_A = (Y_2 + Y_{2'})/2 - (Y_1 + Y_{1'})/2 \\ P_B &= \hat{\pi}_B = (Y_3 + Y_{3'})/2 - (Y_1 + Y_{1'})/2 \\ P_C &= \hat{\pi}_C = (Y_4 + Y_{4'})/2 - (Y_1 + Y_{1'})/2 \end{aligned} \qquad \textbf{(6)}$$

où Y_1', Y_2', Y_3'et Y_4' sont les résultats de la seconde série de pesées.

Les variances de ces estimateurs sont données par :

$$V(P_0) = \sigma^2/2 \quad V(P_A) = V(P_B) = V(P_C) = \sigma^2 \qquad \textbf{(7)}$$

Ceci montre que doubler le nombre d'essais permet de diminuer la variance des estimateurs par deux mais en doublant le prix de l'expérimentation. Si l'on veut encore augmenter la précision des résultats, il suffira de continuer à répéter les mêmes essais. Ceci est résumé dans le tableau suivant :

Nombre de répétitions	1	2	4	8
Nombre d'essais	4	8	16	32
$V(P_A)=V(P_B)=V(P_C)$	$2\sigma^2$	σ^2	$\sigma^2/2$	$\sigma^2/4$
Coût	400 €	800 €	1600 €	3200 €

Nous allons maintenant montrer qu'en changeant de stratégie expérimentale, il est possible d'obtenir des résultats plus précis sans pour autant augmenter le nombre d'essais.

3. Stratégie 2

Une autre manière de procéder consiste à peser deux objets à la fois. Cette stratégie est moins simple car elle ne permet pas de visualiser directement le résultat désiré mais permet, nous allons le voir, de gagner en précision. Nous allons placer successivement sur le plateau de gauche de la balance toutes les combinaisons des 3 objets pris 2 à 2 après avoir déterminé l'erreur systématique par un simple équilibrage de la balance. (voir figure 3)

Figure 3 : Problème des pesées : stratégie 2

La matrice d'expérience correspondante est la suivante :

Stratégie 2	Objet A	Objet B	Objet C
Pesée 1	0	0	0
Pesée 2	1	1	0
Pesée 3	0	1	1
Pesée 4	1	0	1

Le système d'équations à résoudre n'est plus aussi simple que pour la première stratégie, il est donné par :

$$\begin{aligned} Y_1 &= P_0 \\ Y_2 &= P_0 + P_A + P_B \\ Y_3 &= P_0 + P_B + P_C \\ Y_4 &= P_0 + P_A + P_C \end{aligned} \tag{8}$$

Après résolution de ce système par rapport à P_0, P_A, P_B et P_C, les estimateurs obtenus sont les suivants :

$$\begin{aligned} P_0 &= \hat{\pi}_0 = Y_1 \\ P_A &= \hat{\pi}_A = (Y_2 + Y_4 - Y_1 - Y_3)/2 \\ P_B &= \hat{\pi}_B = (Y_2 + Y_3 - Y_1 - Y_4)/2 \\ P_C &= \hat{\pi}_C = (Y_3 + Y_4 - Y_1 - Y_2)/2 \end{aligned} \tag{9}$$

Si on calcule les variances de ces estimateurs, on obtient :

$$V(P_0) = \sigma^2 \quad V(P_A) = V(P_B) = V(P_C) = (\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2)/4 \tag{10}$$

On constate que cette stratégie est meilleure que la première car elle permet d'estimer les masses des objets avec plus de précision sans pour autant augmenter le prix de l'expérimentation.

4. Stratégie 3

La seconde stratégie montre que mettre plusieurs objets sur la balance permet d'augmenter la précision des résultats. On peut donc proposer une stratégie plus élaborée qui consisterait à mettre les trois objets sur la balance à chaque pesée en posant cette fois des objets sur les deux plateaux. Cette stratégie est illustrée dans la figure 4.

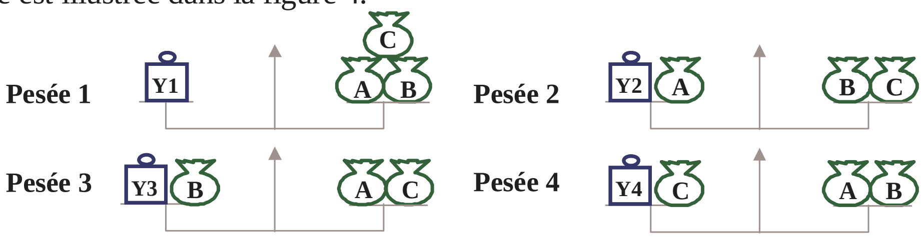

Figure 4 : Problème des pesées : stratégie 3

La matrice d'expérience correspondante est la suivante :

Stratégie 3	Objet A	Objet B	Objet C
Pesée 1	1	1	1
Pesée 2	-1	1	1
Pesée 3	1	-1	1
Pesée 4	1	1	-1

Le système d'équations correspondant à ces essais est :

$$\begin{aligned} Y_1 &= P_0 + P_A + P_B + P_C \\ Y_2 &= P_0 - P_A + P_B + P_C \\ Y_3 &= P_0 + P_A - P_B + P_C \\ Y_4 &= P_0 + P_A + P_B - P_C \end{aligned} \tag{11}$$

Il fournit les estimateurs suivants :

$$\begin{aligned} P_0 &= \hat{\pi}_0 = (Y_2 + Y_3 + Y_4 - Y_1)/2 \\ P_A &= \hat{\pi}_A = (Y_1 - Y_2)/2 \\ P_B &= \hat{\pi}_B = (Y_1 - Y_3)/2 \\ P_C &= \hat{\pi}_C = (Y_1 - Y_4)/2 \end{aligned} \tag{12}$$

et les variances :

$$V(P_0) = \sigma^2 \quad V(P_A) = V(P_B) = V(P_C) = (\sigma^2 + \sigma^2)/4 = \sigma^2/2 \tag{13}$$

Cette nouvelle stratégie permet donc d'encore améliorer la précision des estimateurs.

5. Stratégie 4

Les stratégies 2 et 3 ont montré qu'il est conseillé de peser plusieurs objets à la fois pour obtenir une meilleure précision. Etant donné que dans la dernière stratégie chaque objet a été mis sur la balance à chaque pesée, on peut se demander s'il est encore possible d'améliorer les résultats.

Si l'on examine la troisième stratégie, on constate une certaine asymétrie dans le nombre de fois qu'apparaissent les objets sur chacun des plateaux : chaque objet apparaît trois fois sur le plateau de gauche de la balance et une seule fois sur le plateau de droite. La quatrième stratégie proposée consiste à réétablir cette symétrie en plaçant tous les objets sur la plateau de droite à la première pesée plutôt que sur le plateau de gauche. Ceci est illustré dans la figure 5.

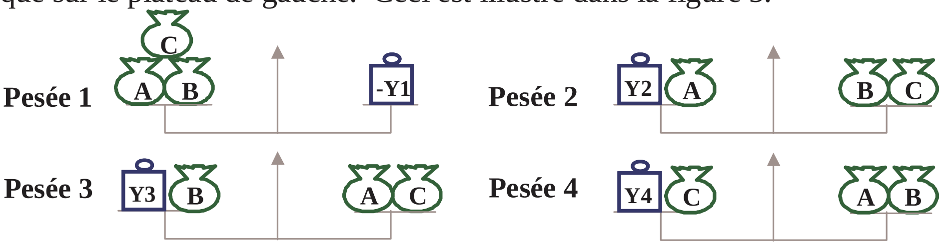

Figure 5 : Problème des pesées : Stratégie 4

La matrice d'expérience correspondante est la suivante :

Stratégie 4	Objet A	Objet B	Objet C
Pesée 1	-1	-1	-1
Pesée 2	-1	1	1
Pesée 3	1	-1	1
Pesée 4	1	1	-1

Le système d'équations correspondant à ces essais est maintenant :

$$\begin{aligned} Y_1 &= P_0 - P_A - P_B - P_C \\ Y_2 &= P_0 - P_A + P_B + P_C \\ Y_3 &= P_0 + P_A - P_B + P_C \\ Y_4 &= P_0 + P_A + P_B - P_C \end{aligned} \quad \textbf{(14)}$$

Les estimateurs qui s'en déduisent valent:

$$\begin{aligned} P_0 &= \hat{\pi}_0 = (Y_1 + Y_2 + Y_3 + Y_4)/4 \\ P_A &= \hat{\pi}_A = (Y_3 + Y_4 - Y_1 - Y_2)/4 \\ P_B &= \hat{\pi}_B = (Y_2 + Y_4 - Y_1 - Y_3)/4 \\ P_C &= \hat{\pi}_C = (Y_2 + Y_3 - Y_1 - Y_4)/4 \end{aligned} \quad \textbf{(15)}$$

Les variances sont en conséquence :

$$V(P_0) = V(P_A) = V(P_B) = V(P_C) = (\sigma^2 + \sigma^2 + \sigma^2 + \sigma^2)/16 = \sigma^2/4 \quad \textbf{(16)}$$

Cette nouvelle stratégie permet donc d'encore améliorer la précision des estimateurs et c'est la meilleure parmi les 4 stratégies proposées. Elle a permis de diviser les variances des estimateurs donnés dans la première stratégie par 8 sans augmenter le prix de l'expérimentation. Pour obtenir la même précision avec la première stratégie, on aurait dû faire 32 essais et payer 8 fois plus !

A ce jeu, on peut se demander s'il n'est pas possible d'encore faire mieux. Malheureusement, la réponse est non. Des résultats théoriques permettent de montrer que la dernière stratégie proposée est optimale dans le sens où il n'existe pas d'autre plan expérimental à 4 essais qui, utilisant le matériel fourni, permette d'obtenir des estimateurs plus précis pour les masses des objets.

6. Comparaison géométrique des 4 stratégies

Les quatre stratégies proposées pour le problème des pesées ont montré que le choix d'une matrice d'expérience particulière a une influence non négligeable sur la précision des résultats obtenus. Nous avons remarqué qu'il était plus intéressant de peser plusieurs objets à la fois si on veut obtenir une meilleure précision. Dans cette section, nous allons montrer qu'une représentation géométrique des 4 matrices d'expérience dans un espace à trois dimensions (une dimension par objet et un point par essai) permet également de comparer les quatre stratégies.

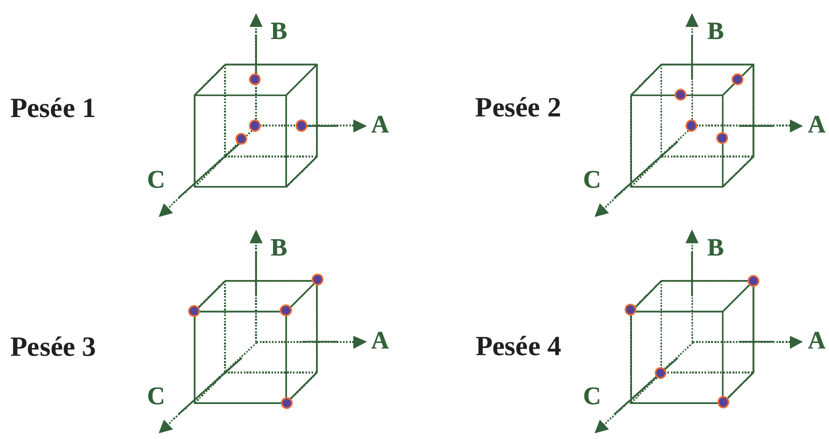

Figure 6 : Représentation géométrique des stratégies

Un examen des positions des essais dans l'espace amène les remarques suivantes :

- La proportion du cube (ou "domaine expérimental") couverte par chacun des plans est d'autant plus grande que la stratégie est bonne.
- Il paraît préférable de mettre les essais sur les sommets du cubes (stratégies 3 et 4) plutôt qu'au centre du cube ou sur les arrêtes. Ceci correspond à dire qu'il est préférable de peser 3 objets à chaque essai.
- Le seul plan qui soit distribué de façon tout à fait symétrique dans le cube est le plan 4.
- Le plan 4 est le seul plan pour lequel chaque "niveau des facteurs" est testé un même nombre de fois (chaque objet est mis un même nombre de fois sur chaque plateau).

La représentation graphique des matrices de pesées montre clairement que la répartition des points expérimentaux dans l'espace a une influence directe sur la qualité des résultats. Les remarques intuitives faites ici se généraliseront dans des cas plus généraux et trouveront un fondement théorique plus rigoureux.

7. Reformulation de problème des pesées dans un contexte réel

Il est clair que l'exemple présenté ici est fictif et que peu d'expérimentateurs utilisent encore actuellement du matériel expérimental tel qu'une balance à deux plateaux. Nous allons essayer dans cette section de transposer le problème des pesées à une application plus réaliste.

Un expérimentateur désire étudier une réaction chimique et s'intéresse à l'influence de trois facteurs : la température de réaction, la quantité de catalyseur et le temps de réaction, sur une réponse le rendement de la réaction. Les domaines de variation de ces facteurs sont respectivement : 100 à 120 degrés, 1 à 5 % et 2 à 4 heures. Il désire répondre à sa question par 4 essais et croit que le phénomène peut être raisonnablement approximé par une équation du type :

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \beta_3 X_3 + \beta_4 X_4 + \varepsilon \qquad \textbf{(17)}$$

où Y est le rendement observé, β_i des paramètres inconnus, X_i les valeurs des différents facteurs et ε l'erreur expérimentale.

Ce problème peut-être exprimé comme une problème de pesées en disant que l'expérimentateur désire "peser" l'influence de trois "objets" (facteurs) sur la variable rendement de la réaction. Y représente le résultat d'une "pesée" (ou essai) pour laquelle on aura choisi un niveau pour chaque facteur explicatif X_i. Les paramètres β_i représentent les "poids" des différents facteurs. Si le niveau maximum est choisi pour un facteur, cela correspond à mettre l'objet sur le plateau de gauche de la balance. S'il est à sa valeur moyenne, cela correspond à ne pas mettre l'objet sur la balance et s'il est à son niveau minimum à le mettre sur le plateau de droite.

La 1ère stratégie expérimentale deviendrait donc :

Stratégie 1	Température	Catalyseur	Temps
Essai 1	110	3	3
Essai 2	120	3	3
Essai 3	110	5	3
Essai 4	110	3	4

Cette stratégie correspond à faire une première expérience au centre du domaine expérimental (110,3,3) et à faire varier ensuite un facteur à la fois autour de ce centre.

La stratégie 4 pourrait également être utilisée et, comme pour le problème des pesées on peut montrer qu'elle est la plus efficace. Elle est malheureusement plus complexe et moins intuitive pour l'expérimentateur puisqu'elle fait varier plusieurs facteurs à la fois comme le montre le tableau suivant :

Stratégie 1	Température	Catalyseur	Temps
Essai 1	100	1	2
Essai 2	100	5	4
Essai 3	120	1	4
Essai 4	120	5	2

Ce dernier plan expérimental est optimal pour le couple modèle, domaine donné dans le sens où il permet d'estimer les paramètres β_i du modèle (ou masses des facteurs) avec un maximum de précision. Nous verrons dans la suite que ce plan est de type *factoriel fractionnaire* et s'applique dans de nombreux domaines.

8. Conclusions

Nous avons monté sur un exemple simple que le choix d'une bonne stratégie expérimentale permet d'influencer fortement la qualité des résultats qu'on en retire.

Une stratégie expérimentale peut être réfléchie *a priori* en se basant sur un modèle théorique utilisé pour approximer le problème réel. La méthode consiste à comparer différentes stratégies qui paraissent raisonnables pour le problème donné et choisir, *sans faire aucune expérience*, celle qui est optimale par rapport à certains critères choisis (notamment le coût et la précision des résultats obtenus).