Veillons au grain

Bernard ANGUENOT
Professeur à l'I.N.S.A. de Lyon

I. Préambule

Professeur dans une école d'ingénieur, j'ai eu l'occasion, depuis trois ans, de m'occuper de formation permanente d'ingénieurs.
Qui voit-on venir ?

1) Des utilisateurs qui, ayant appliqué pendant un certain temps des recettes de cuisine, se trouvent un jour dans la situation de Monsieur-Tout-le-monde qui a cassé sa montre... ils ne comprennent plus pourquoi "cela ne marche pas".

2) Des curieux qui, ayant entendu parlé des méthodes statistiques, viennent "voir"...

3) Des scientifiques qui, confrontés avec des problèmes nouveaux pour eux, viennent en chercher la solution.

Les stagiaires ne disent généralement pas grand chose pendant les cours mais dans les couloirs... inévitablement nous nous voyons poser la question: "Dans ce cas-là, que dois-je faire ?"

Il y a dans cette petite phrase deux mots qui attirent l'attention:

1°-"Faire" : On attend du statisticien ou de la statistique une règle de conduite... le but de mon petit exposé est justement de montrer que le statisticien n'est pas là pour fournir une telle règle: il est là pour mesurer les risques encourus lorsqu'on décide de suivre telle ou telle règle d'action.
Néanmoins, ce mot "faire" semble bien réconfortant, il consacre l'utilité des mathématiques dans un domaine d'activités très vaste; à l'heure où est prônée l'ouverture de l'école sur la vie, la statistique me semble être un merveilleux pont pour relier le monde abstrait du professeur de mathématique à celui de l'univers aléatoire dans lequel nous vivons.

2°-"Dois" : c'est la trace indélébile de notre enseignement dogmatique et purement déterministe.
Un théorème de mathématique, c'est toujours: "si A alors B" ; (parfois, il y a aussi: "si B alors A") mais le "p't-être bien qu'oui, p't-être bien qu'non" a rarement droit de cité dans nos exposés...
Même notre enseignement de la physique est trop dirigé vers l'énoncé des lois déterministes...; d'où, en grande partie, le hiatus qui se fait sentir entre les mathématiciens et les physiciens: les premiers prétendant que les seconds font de la cuisine, les seconds que les élèves font du rigorisme qui dépasse souvent les besoins immédiats.

La statistique est l'endroit privilégié où on peut aborder des situations simples, y développer l'art subtil du calcul, des développements mathématiques sans jamais être l'esclave, le prisonnier, de sa méthode, où la réponse est très rarement <u>vrai</u> ou <u>faux</u>, rarement <u>oui</u> ou <u>non</u>...

Je vais donc traiter ici un problème banal que nous aborderons par un certain nombre de méthodes exposées avec un matériel mathématique de plus en plus élaboré, où chacun peut trouver matière à faire réfléchir des élèves de niveaux très divers.

II Le problème

Nous avons imaginé la situation suivante:

> Un marchand de graines a l'habitude de fournir à un ingénieur agronome des mélanges de 6 types de graines qui entrent toutes dans le mélange avec les mêmes proportions (1/6 chacune).

Question d'un auditeur: qu'entend-on par proportion ? Est-ce en poids ? en volume ?
Réponse: Aucune importance: admettons que les graines aient sensiblement le même poids et le même volume et qu'elles soient différenciables à l'oeil (par la forme, par exemple).

Les 6 types sont désignés par A, B, C, D, E, F.

> Un jour, l'ingénieur lui demande de préparer un mélange spécial dans lequel les graines A entreront pour $\frac{1}{4}$, B pour $\frac{1}{12}$, les autres en proportions égales entre elles.
> Le jour de la livraison, catastrophe ! Un employé trop zélé a déplacé tous les sacs, si bien que lorsque l'ingénieur arrive, il trouve le grainetier fort perplexe... il croit bien que le <u>premier sac</u> est le bon, mais il n'en est pas certain.
> Comme ce dernier est honnête, il avoue sa mésaventure au client.
> Celui-ci, compréhensif, étudie avec le vendeur comment faire.

1ère solution.
On trie toutes les graines...
Solution jugée trop fastidieuse par les deux parties.

2ème solution.
On va prélever une partie des graines de ce sac.
A partir de l'information qu'on va recueillir sur cet échantillon, il nous va falloir nous décider entre deux hypothèses: "ce sac est le bon" et "c'est un sac ordinaire".
Se décider, c'est se fixer une règle de conduite: c'est-à-dire indiquer quelle expérience on va faire pour recueillir de l'information et préciser quelle sera notre conduite suivant les résultats.

Il y a donc quatre cas :

	C'est le bon sac	C'est un sac ordinaire
On le vend	Tout va bien	Risque de l'acheteur β
On ne le vend pas	Risque du vendeur α	Tout va bien

Les risques sont les probabilités pour qu'on se trouve dans cette situation.

Question du même auditeur: "Mais la notion de risque est liée à celle du coût de l'erreur, que vous ne faites pas intervenir ici".

Réponse: "Aucune importance: j'ai posé, par définition,

risque = probabilité ...

il sera temps par la suite de lier risque et coût, mais laissez-moi, au moins, maître de mes choix !"

TEST 1

Idée : Si c'est le bon sac, A est le plus probable.

Règle de décision : On tire 1 graine.

$$\begin{cases} \text{Si elle est du type A :} & \text{On vend} \\ \text{Si elle n'est pas du type A:} & \text{On ne vend pas} \end{cases}$$

Calcul des risques

$$\alpha = 0{,}750 = \text{prob [non A / c'est le bon sac]}$$

$$\beta = \frac{1}{6} = 0{,}167 = \text{prob [A / sac ordinaire]}$$

Critique

. L'acheteur, à l'extrême rigueur, peut être satisfait de ce test.

. Le vendeur, certainement pas.

TEST 2

Idée : On ne va pas se livrer pieds et poings liés au résultat d'une expérimentation portant sur un seul essai.

On va élever la taille de l'échantillon.

Règle de décision : On prélève deux graines :

$\begin{cases} \text{Si l'une au moins est du type A,} & \text{on vend} \\ \text{Si aucune n'est du type A,} & \text{on ne vend pas.} \end{cases}$

Calcul des risques :

α = prob [aucune du type A / c'est le bon sac]
$= (0,75)^2 = 0,563$

β = prob [une au moins de type A / c'est un sac ordinaire]

$$= 1 - \left(\frac{5}{6}\right)^2 = 0,306$$

Critique :

- Le vendeur peut se considérer comme un peu moins insatisfait.
- L'acheteur est en droit de considérer qu'on l'entraîne sur une pente savonneuse.

De toutes façons ce n'est pas équitable, le vendeur court un risque presque double de celui de l'acheteur.

Question du même auditeur: "Vous parlez de tirer deux graines, vous ne dites pas comment: avec ou sans remise ?"
Réponse: "A ce niveau-là, c'est sans importance: si le sac est suffisamment important, tirer une graine ne change pas sensiblement la composition du sac et les probabilités ne changent pas assez d'un tirage à l'autre".
Néanmoins, ce lièvre méritait d'être soulevé et je me proposais de le faire plus loin...; disons que nous ferons des tirages avec remise...

TEST 3

Idée : Continuer sur la même voie pour essayer d'équilibrer les risques, c'est-à-dire rendre le marché équitable.

Règle de décision : On prélève trois graines :

$\begin{cases} \text{Si l'une au moins est du type A,} & \text{on vend} \\ \text{Si aucune n'est du type A,} & \text{on ne vend pas.} \end{cases}$

Calcul des risques

$$\alpha = \text{prob [aucune de type A / bon sac]} = (0,75)^3 = 0,422$$

$$\beta = \text{prob [au moins une de type A / sac ordinaire]} = 1 - \left(\frac{5}{6}\right)^3 = 0,421$$

Critique

Sur le plan de l'équité, c'est une assez belle réussite, mais, si les risques sont partagés, ils le sont à un niveau tel qu'aucun n'est en droit de se montrer satisfait.

Possibilités :

Continuer; le vendeur qui voit son risque diminuer sera certainement d'accord; l'acheteur qui voit le sien augmenter, certainement pas.
Plus précisément, on prélève n graines

. Si l'une au moins est du type A, on vend
. Si aucune n'est du type A, on ne vend pas.

$$\alpha = (0,75)^n$$

$$\beta = 1 - \left(\frac{5}{6}\right)^n$$

Il est bien évident que ce test devient ridicule dès que n dépasse 4 ; étant donné que dans les deux cas la proportion de graines du type A ne dépasse par 1/4 , il est stupide de prendre une règle de décision fondée sur le fait qu'il y ait moins du quart des graines de type A dans un échantillon.

TEST 4

Idée : On peut évidemment augmenter la taille de l'échantillon, mais il faudra aussi être plus exigeant sur le nombre de graines du type A.

Règle de décision : On prélève deux graines :

$$\begin{cases}\text{Si elles sont toutes deux du type A,} & \text{on vend.}\\ \text{Si elles ne sont pas toutes deux du type A,} & \text{on ne vend pas.}\end{cases}$$

Calcul des risques

$$\alpha = \text{prob [elles ne sont pas toutes deux de type A / c'est le bon sac]}$$
$$= 1 - \left(\frac{1}{4}\right)^2 = 0,938$$

β = prob [elles sont toutes deux de type A / c'est le sac ordinaire]

$= \left(\frac{1}{6}\right)^2 = 0{,}028$

Critique

De plus en plus inacceptable pour le vendeur.

Une très bonne affaire pour l'acheteur, mais l'injustice est des plus flagrantes.

TEST 5

Idée : persévérer dans cette voie, mais augmenter la taille n de l'échantillon.

Règle de décision : Prélever 4 graines

$\begin{cases} \text{Si deux au moins sont du type A,} & \text{on vend;} \\ \text{Si moins de deux sont du type A,} & \text{on ne vend pas.} \end{cases}$

Calcul des risques

α = prob [0 ou 1 graine de type A / bon sac]

$= \underbrace{(0{,}75)^4}_{\text{il n'y en a pas}} + \underbrace{(0{,}25)\ (0{,}75)^3 \times 4}_{\text{il y en a 1}} = 0{,}738$

il n'y en a pas — il y en a 1
(la première, ou la deuxième, ou la troisième ou la quatrième)

Remarque : Autre raisonnement

Il y a 3 graines au moins qui ne sont pas du type A $\longrightarrow (0{,}75)^3$

Sachant cela, pour la quatrième :

elle n'est pas du type A	0,75	
c'est la première	0,25	
c'est la seconde	0,25	1,75
c'est la troisième	0,25	
c'est la quatrième	0,25	

soit $(0{,}75)^3 \times 1{,}75$

β = prob [2, 3, 4 du type A / sac ordinaire]

= prob [(il y en a 2) ou (il y en a 3) ou (il y en a 4)]

= prob (2A) + prob (3A) + prob (4A)

$= 6 \left(\frac{1}{6}\right)^2 \times \left(\frac{5}{6}\right)^2 + 4 \left(\frac{5}{6}\right) \left(\frac{1}{6}\right)^3 + \left(\frac{1}{6}\right)^4 = 0{,}132$

Critique

Le risque du vendeur diminue, celui de l'acheteur augmente.

Par rapport au précédent, l'injustice est moins criante, mais le résultat est encore fort mauvais.

TEST 6

Idée : Continuer à augmenter la taille de l'échantillon : $n = 8$

Si deux graines au moins sont de type A : on vend;

Si moins de deux graines sont de type A: on ne vend pas.

Calcul des risques

$$\alpha = \text{prob}\ [\,0 \text{ ou } 1 \ / \text{ bon sac}\,]$$
$$= (0{,}75)^8 + 8\ (0{,}25)\ (0{,}75)^7 = 0{,}367$$

$$\beta = \text{prob}\ [\,2,\ 3,\ \ldots,\ 8 \ / \text{ sac ordinaire}\,]$$
$$= 1 - \text{prob}\ [\,0,\ 1 \ / \text{ sac ordinaire}\,]$$
$$= 1 - \left[\left(\frac{5}{6}\right)^8 + 8\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^7\right] = 0{,}395$$

Remarque

$$n = 7 \longrightarrow \alpha = (0{,}75)^7 + 7\ (0{,}25)(0{,}75)^6 = 0{,}445$$
$$\beta = 1 - \left[\left(\frac{5}{6}\right)^7 + 7\left(\frac{1}{6}\right)\left(\frac{5}{6}\right)^6\right] = 0{,}330$$

Commentaire :

Les risques arrivent à peu près à s'équilibrer vers 0,38 ; c'est mieux que dans le test 3, mais ils sont encore bien trop élevés.

TEST 7

Une idée : Augmenter simultanément la taille n de l'échantillon et le nombre de graines du type A nécessaires pour conclure.

Règle de décision :

On se fixe deux naturels n et r $(r < n)$

On prélève n graines; parmi celles-ci, k sont du type A.

$$\begin{cases} \text{Si } k \geqslant r, & \text{on vend} \\ \text{Si } k < r, & \text{on ne vend pas} \end{cases}$$

Calcul des risques

$$\alpha = \text{prob } (k < r \text{ / bon sac}) = \sum_{k=0}^{r-1} C_n^k \, (0{,}25)^k \, (0{,}75)^{n-k}$$

$$\beta = \text{prob } (k \geqslant r \text{ / sac ordinaire}) = \sum_{k=r}^{n} C_n^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

Commentaires

- Si on cherche la justice, c'est-à-dire $\alpha = \beta$, on est conduit à une équation à deux inconnues n et r, d'où un certain nombre de solutions (qui dépendent de la valeur commune aux deux risques).

  Comme n et k doivent être entiers, il se peut qu'il n'y ait aucune solution au sens mathématique exact.

  Au sens statistique, ce sera plutôt $|\alpha - \beta| < \varepsilon$ (ε donné), d'où un plus grand nombre de solutions.

- On peut être plus exigeant (se fixer la valeur des risques): $\alpha = \alpha_0$ et $\beta = \beta_0$, d'où un système de deux équations à deux inconnues.

  De toutes façons ces calculs nécessitent l'usage d'un ordinateur ou d'une table.

TEST 8

Idée : Jusqu'ici on n'a opéré qu'avec A; mais on pourrait opérer avec B. On pourrait recommencer la suite des raisonnements.

Règle de décision : On prélève 1 graine

$$\begin{cases} \text{Si elle n'est pas du type B ,} & \text{on vend;} \\ \text{Si elle est du type B,} & \text{on ne vend pas.} \end{cases}$$

Calcul des risques

$$\alpha = \text{prob } (B \text{ / bon sac}) = \frac{1}{12} = 0{,}083$$

$$\beta = \text{prob } (B \text{ / sac ordinaire}) = \frac{5}{6} = 0{,}833$$

Critique

Les rôles du vendeur et de l'acheteur sont inversés. Les réactions seront les mêmes.

TEST 9

Idée : Augmenter la taille de l'échantillon.

Règle de décision : On prélève 3 graines :

$$\begin{cases} \text{Si aucune n'est du type B,} & \text{on vend} \\ \text{Si au moins une est du type B,} & \text{on ne vend pas} \end{cases}$$

Calcul des risques

$$\alpha = \text{prob } [1 \text{ B au moins / bon sac}] = 1 - \left(\frac{11}{12}\right)^3 = 0{,}230$$

$$\beta = \text{prob } [0 \text{ B / sac ordinaire}] = \left(\frac{5}{6}\right)^3 = 0{,}579$$

TEST 10

Idée : Chercher à équilibrer les risques.

Règle de décision : Prélever n graines

$$\begin{cases} \text{Si aucune n'est du type B,} & \text{on vend;} \\ \text{Si au moins une est du type B,} & \text{on ne vend pas.} \end{cases}$$

Calcul des risques

$$\alpha = \text{prob } [1 \text{ B au moins / bon sac}] = 1 - \left(\frac{11}{12}\right)^n$$

α est une fonction croissante de n

$$\beta = \text{prob } [0 \text{ B / sac ordinaire}] = \left(\frac{5}{6}\right)^n$$

β est une fonction décroissante de n

Comment choisir n pour que:

$$1 - \left(\frac{11}{12}\right)^n < \left(\frac{5}{6}\right)^n \qquad \text{(par tâtonnement)}$$

	α	β
n = 4	0,294	0,482
n = 5	0,353	0,402
n = 6	0,407	0,355

Discussion : On ne pourra être vraiment équitable.

TEST 11

<u>Idée</u> : Continuer en augmentant la taille de l'échantillon et le nombre de graines du type B.

<u>Règle de décision</u> :

On prélève n graines; k sont du type B (r est donné)

$$\begin{cases} \text{Si } k \leq r \text{ , on vend} \\ \text{Si } k > r \text{ , on ne vend pas} \end{cases}$$

<u>Calcul des risques</u> :

$$\alpha = \text{prob}\,[\,k > r \;/\; \text{bon sac}\,]$$

$$= \sum_{k=r+1}^{n} C_n^k \left(\frac{1}{12}\right)^k \left(\frac{11}{12}\right)^{n-k}$$

$$\beta = \text{prob}\,[\,k \leqslant r \;/\; \text{sac ordinaire}\,]$$

$$= \sum_{k=0}^{r} C_n^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}$$

Les résultats sont consignés sur les courbes.

TEST 12

Idée : Se servir simultanément de A et B.

Règle de décision : On prélève n graines :

$$\begin{cases} \text{Si au moins 1 A ou pas B : on vend} \\ \text{Si pas de A et au moins 1 B : on ne vend pas.} \end{cases}$$

Calcul des risques

$$\alpha = \text{prob [pas de A et au moins 1 B / bon sac]}$$
$$= \text{prob [pas de A]} - \text{prob [ni A ni B]}$$

En effet:

[pas de A] = (pas de A et au moins 1 B) ou (pas de A et pas de B)

$$\alpha = (0{,}75)^n - \left(\frac{2}{3}\right)^n$$

$$\beta = \text{prob 1 A au moins ou pas de B / sac ordinaire}$$

$$= 1 - \text{prob 1 B au moins et pas de A / sac ordinaire}$$

$$= 1 - \left[\left(\frac{5}{6}\right)^n - \left(\frac{2}{3}\right)^n\right]$$

. Etude du comportement de α en fonction de n.

Soit f l'application dans R telle que :

$$f : x \longmapsto (0{,}75)^x - \left(\frac{2}{3}\right)^x$$

$$f'(x) = (\ln 0{,}75) \,.\, (0{,}75)^x - (\ln \tfrac{2}{3}) \,.\, \left(\frac{2}{3}\right)^x$$

$$f'(x) < 0 \longmapsto (\ln 0{,}75).(0{,}75)^x < (\ln \tfrac{2}{3}) \,.\, \left(\frac{2}{3}\right)^x$$

$$0{,}709\ldots = \frac{\ln 0{,}75}{\ln \frac{2}{3}} > \left(\frac{2}{3}\right)^x : (0{,}75)^x = \left(\frac{2}{3}\right)^x \times \left(\frac{4}{3}\right)^x = \left(\frac{8}{9}\right)^x$$

$$\ln \frac{\ln 0{,}75}{\ln \frac{2}{3}} > x \ \ln \frac{8}{9}$$

$$x > \frac{\ln \dfrac{\ln 0{,}75}{\ln \frac{2}{3}}}{\ln \dfrac{8}{9}} = 2{,}91$$

A partir de n = 3 , α décroît.

Etude du comportement de β

$$g \left| \begin{array}{l} \mathbb{R} \longrightarrow \mathbb{R} \\ x \longmapsto \left(\frac{5}{6}\right)^x - \left(\frac{2}{3}\right)^x \end{array} \right.$$

$$g'(x) = \left(\ell n \frac{5}{6}\right) \cdot \left(\frac{5}{6}\right)^x - \left(\ell n \frac{2}{3}\right) \cdot \left(\frac{2}{3}\right)^x$$

$$g'(x) < 0 \longmapsto \left(\ell n \frac{5}{6}\right) \cdot \left(\frac{5}{6}\right)^x < \left(\ell n \frac{2}{3}\right) \cdot \left(\frac{2}{3}\right)^x \longmapsto$$

$$\frac{\ell n \frac{5}{6}}{\ell n \frac{2}{3}} > \left(\frac{2}{3}\right)^x \times \left(\frac{6}{5}\right)^x = \left(\frac{4}{5}\right)^x$$

$$x \, \ell n \frac{4}{5} < \ell n \frac{\ell n \frac{5}{6}}{\ell n \frac{2}{3}}$$

$$x > \frac{\ell n \dfrac{\ell n \frac{5}{6}}{\ell n \frac{2}{3}}}{\ell n \frac{4}{5}} = 3{,}58$$

A partir de $n = 4$, β croît.

$n = 4$	:	$\alpha = 0{,}118$	$\beta = 0{,}715$	inutile d'augmenter n
$n = 3$	:	$\alpha = 0{,}125$	$\beta = 0{,}717$	
$n = 2$	:	$\alpha = 0{,}118$	$\beta = 0{,}750$	

Ce test ne sera jamais bon.

Critique

Il peut sembler intéressant de dire que s'il n'y a pas de graine B dans l'échantillon, c'est parce qu'il y en a peu dans le sac, qui a donc plus de chance d'être le bon, ou que si n est petit la présence d'un A milite en faveur du bon sac; mais on voit bien que si n augmente, la présence d'un A n'a plus guère de signification.
Par contre, l'autre volet de notre règle de décision paraît absurde: si on décide de ne pas vendre, c'est qu'on pense que le sac est un sac ordinaire, donc que A et B sont dans les mêmes proportions; or, on les traite de façon complètement différente ...

Pause

Jusqu'ici, nous n'avons utilisé qu'un outil mathématique très élémentaire qui conduit à des résultats peu probants.

Les difficultés proviennent en grande partie de ce que nous avons bricolé des aléa discrets et utilisé (explicitement ou pas) la loi binomiale qui est très lourde ...
Il est temps de changer de registre et d'utiliser un outil mathématique plus élaboré: la tendance de cette loi vers la loi normale ...
Les programmes du secondaire vont jusque là... Il s'agit d'aller légèrement plus loin ...

TEST 13

Idée : Reprendre l'idée du test 7, mais le traiter en utilisant l'approximation normale de la loi binomiale.

Règle de décision : On tire n graines; il y en a K du type A (r est donné)

$$\begin{cases} \text{Si } K \geqslant r, & \text{on vend} \\ \text{Si } K < r, & \text{on ne vend pas.} \end{cases}$$

Calcul des risques : On sait que K est un aléa $\mathcal{B}(n, p)$

Donc que $\left| \begin{array}{l} E(K) = np \\ \text{var } K = npq \end{array} \right.$

et que $K \xrightarrow[\mathcal{L}]{n \to \infty} \mathcal{N}(np, \sqrt{npq})$

donc $\dfrac{K - np}{\sqrt{npq}} \xrightarrow[n \to \infty]{\mathcal{L}} \mathcal{N}(0, 1)$

$$\alpha = \text{prob}\left[K < r \;/\; \text{bon sac}\right] = \text{prob}\left[\frac{K - np}{\sqrt{npq}} < \frac{r - np}{\sqrt{npq}} \;/\; \text{bon sac}\right]$$

$$= \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{r-np}{\sqrt{npq}}} e^{-\frac{t^2}{2}} \, dt \qquad \text{avec} \quad p = 0,25$$

$$\beta = \text{prob}\left[K \geqslant r \;/\; \text{sac ordinaire}\right] = 1 - \text{prob}\left[K < r \;/\; \text{sac ordinaire}\right]$$

$$= 1 - \text{prob}\left[\frac{K - np'}{\sqrt{np'q'}} < \frac{r - np'}{\sqrt{np'q'}} \;/\; \text{sac ordinaire}\right]$$

$$= 1 - \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\frac{r-np'}{\sqrt{np'q'}}} e^{-\frac{t^2}{2}} \, dt \qquad \text{avec} \quad p' = \frac{1}{6}$$

Or la fonction de répartition F de la loi normale est tabulée:

$$F : x \longmapsto \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{x} e^{-\frac{t^2}{2}} \, dt = \text{prob}\left(\mathcal{N}(0, 1) < x\right)$$

donc

$$\left\{\begin{aligned} \alpha &= F\left(\frac{r - 0{,}25\,n}{\sqrt{n \times 0{,}25 \times 0{,}75}}\right) \\ 1 - \beta &= F\left(\frac{r - \frac{n}{6}}{\sqrt{n \times \frac{5}{36}}}\right) \end{aligned}\right.$$

Calculs

Choisissons $\alpha = \beta = 5\ \%$

$$\left\{\begin{aligned} &F(u) = 0{,}05 \longrightarrow F(-u) = 0{,}95 \longrightarrow -u = 1{,}65 \longrightarrow u = -\,1{,}65 \\ &F(v) = 0{,}95 \longrightarrow v = 1{,}65 \end{aligned}\right.$$

$$\left.\begin{aligned} &\frac{r - 0{,}25\,n}{\sqrt{n \times 0{,}25 \times 0{,}75}} = -\,1{,}65 \\ &\frac{r - \frac{n}{6}}{\sqrt{n \times \frac{1}{6} \times \frac{5}{6}}} = 1{,}65 \end{aligned}\right\} \longrightarrow \left\{\begin{aligned} &\frac{r}{n} - 0{,}25 = -\,1{,}65\ \sqrt{0{,}25 \times 0{,}75}\ \frac{1}{\sqrt{n}} \\ &\frac{r}{n} - \frac{1}{6} = 1{,}65\ \sqrt{\frac{1}{6} \times \frac{5}{6}}\ \frac{1}{\sqrt{n}} \end{aligned}\right.$$

$$\begin{aligned} 0{,}25 - \frac{1}{6} &= 1{,}65 \left[\sqrt{0{,}25 \times 0{,}75} + \sqrt{\frac{1}{6} \times \frac{5}{6}}\right] \frac{1}{\sqrt{n}} \\ &= 1{,}65\ \left[0{,}433 + 0{,}373\right]\ \frac{1}{\sqrt{n}} \\ 0{,}083 &= 1{,}329\ \frac{1}{\sqrt{n}} \\ 0{,}063 &= \frac{1}{\sqrt{n}} \\ 15{,}953 &= \sqrt{n} \\ 254{,}488 &= n \qquad \boxed{n = 255} \end{aligned}$$

$$\frac{r}{n} = 0{,}25 - 1{,}65 \times 0{,}433 \times 0{,}063 = 0{,}205$$

$$r = 255 \times 0{,}205 = 52{,}272$$

$$\boxed{r = 53}$$

Règle de décision

On prélève un échantillon de taille 255 .

On trouve k graines du type A.

$$\begin{cases} \text{Si } k \geqslant 53, & \text{on vend} \\ \text{Si } k < 53, & \text{on ne vend pas.} \end{cases}$$

Remarque

On doit pouvoir améliorer la technique: si à la 202ème graine on n'a pas encore de graines du type A ou si, avant la 255ème graine, on a déjà 53 graines A, il est inutile d'aller plus loin.

TEST 14

Idée : Exploiter la remarque finale du test 13: utiliser l'information recueillie lors de n tirages pour décider si on peut conclure ou s'il convient plutôt de recueillir plus d'informations en procédant à un (n+1)ème tirage.

Soit P_0^n la probabilité pour que, dans le cas du bon sac, on ait obtenu exactement k graines A lors de n tirages;

et P_1^n la probabilité pour que, dans le cas d'un sac ordinaire, on ait obtenu exactement k graines A lors de n tirages.

- Si P_1^n est nettement supérieur à P_0^n , on conclut : il est plus probable que ce soit un sac ordinaire, donc on ne vend pas;
- Si P_1^n est nettement inférieur à P_0^n , on conclut : il est plus probable que ce soit le bon sac, donc on vend.
- Si P_1^n et P_0^n ne diffèrent pas assez, on conclut à l'opportunité de procéder à un (n+1)ème tirage.

On dira que P_1^n est nettement supérieur à P_0^n si on est capable de se fixer un nombre $\lambda > 1$ tel que

$$R_n = \frac{P_1^n}{P_0^n} \geqslant \lambda$$

On conclut: c'est un sac ordinaire si $\quad P_1^n \geqslant \lambda\, P_0^n$

ou $\quad 1 - \beta \geqslant \lambda\alpha \quad$ c'est-à-dire $\quad \lambda \leqslant \dfrac{1-\beta}{\alpha}$

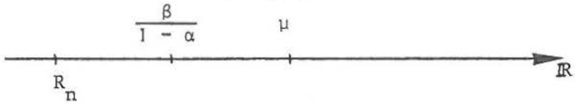
$$\left.\begin{array}{ll} R_n & \text{doit être supérieur à } \lambda \\ \lambda & \text{doit être inférieur à } \dfrac{1-\beta}{\alpha} \end{array}\right\} \quad \text{il suffit que} \quad R_n \geqslant \frac{1-\beta}{\alpha}$$

On dira: $\quad P_1^n$ est nettement inférieur à P_0^n si on est capable de se fixer $\quad \mu < 1 \quad$ tel que $\quad R_n \leqslant \mu$

On conclut: $\quad$ c'est le bon sac si $\quad P_1^n \leqslant \mu\, P_0^n$

ou si $\quad \beta \leqslant \mu\,(1-\alpha) \quad$ c'est-à-dire $\quad \mu \geqslant \dfrac{\beta}{1-\alpha}$

$$\left.\begin{array}{ll} R_n & \text{doit être inférieur à } \mu \\ \mu & \text{doit être supérieur à } \dfrac{\beta}{1-\alpha} \end{array}\right\} \quad \text{il suffit que} \quad R_n \leqslant \frac{\beta}{1-\alpha}$$



<u>Procédure pratique</u>

On procède au tirage de n graines.
Si on a obtenu k graines de type A, on forme R_n .

$$R_n = \frac{C_n^k \left(\frac{1}{6}\right)^k \left(\frac{5}{6}\right)^{n-k}}{C_n^k \left(\frac{1}{4}\right)^k \left(\frac{3}{4}\right)^{n-k}} = \left(\frac{2}{3}\right)^k \left(\frac{10}{9}\right)^{n-k}$$

- Si $R_n \geqslant \dfrac{1-\beta}{\alpha}$ $\quad$ on conclut: $\quad$ on ne vend pas
- Si $R_n \leqslant \dfrac{\beta}{1-\alpha}$ $\quad$ on conclut: $\quad$ on vend
- Si $\dfrac{\beta}{1-\alpha} < R_n < \dfrac{1-\beta}{\alpha}$ $\quad$ on forme R_{n+1} et on recommence.

Exécution pratique

. Si $\ell n\ R_n = k\ \ell n\ \frac{2}{3} + (n-k)\ \ell n\ \frac{10}{9} \geqslant \ell n\ \frac{1-\beta}{\alpha}$ on ne vend pas

$$k\left[\ell n\ \frac{2}{3} - \ell n\ \frac{10}{9}\right] \geqslant -n\ \ell n\ \frac{10}{9} + \ell n\ \frac{1-\beta}{\alpha}$$

$$k\ \ell n\left(\frac{2}{3}\times\frac{9}{10}\right) \geqslant -n\ \ell n\ \frac{10}{9} + \ell n\ \frac{1-\beta}{\alpha}$$

$$k\ \ell n\ \frac{3}{5} \geqslant -n\ \ell n\ \frac{10}{9} + \ell n\ \frac{1-\beta}{\alpha}$$

Si $k \leqslant -n\ \dfrac{\ell n\ \frac{10}{9}}{\ell n\ \frac{3}{5}} + \dfrac{\ell n\ \frac{1-\beta}{\alpha}}{\ell n\ \frac{3}{5}}$ on ne vend pas

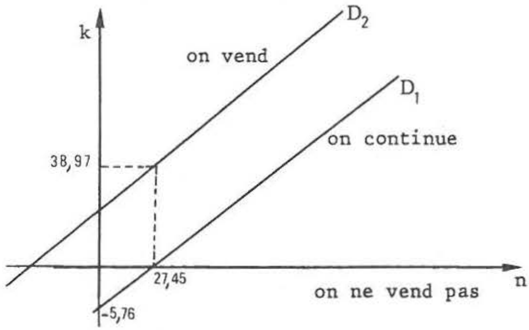
. Si $\ell n\ R_n = k\ \ell n\ \frac{2}{3} + (n-k)\ \ell n\ \frac{10}{9} \leqslant \ell n\ \frac{\beta}{1-\alpha}$ on vend

Si $k \geqslant -n\ \dfrac{\ell n\ \frac{10}{9}}{\ell n\ \frac{3}{5}} + \dfrac{\ell n\ \frac{\beta}{1-\alpha}}{\ell n\ \frac{3}{5}}$ on vend

Fixons $\alpha = \beta = 0,05$

D_1 : $k = 0,21\ n - 5,76$

D_2 : $k = 0,21\ n + 5,76$



Cela ne commence à devenir intéressant qu'à partir de $n = 28$.

TEST 15

Idée : Jusqu'ici on n'a utilisé que A ou B et jamais les autres graines... On va maintenant comparer la distribution obtenue à la distribution idéale obtenue si la fréquence d'apparition de chaque graine était rigoureusement sa probabilité.

On sait bien que le hasard introduit des fluctuations dans les résultats, ce que nous appellerons les hasards de l'échantillonnage...

Par rapport au modèle proposé, la variabilité peut être due à deux causes: le hasard de l'échantillonage si le modèle est bon, ou bien le choix d'un mauvais modèle.

On sait, si le modèle proposé est bon, définir une variable qui mesure l'écart entre le modèle théorique proposé et l'échantillonnage observé: cette variable E est la somme des 6 termes écrits dans la colonne "Ecarts".

Prenons comme modèle: "c'est le bon sac".

On prélève n graines.

	Effectifs observés	Effectifs théoriques	Ecart
A	N_1	$0,25n$	$\dfrac{(N_1 - 0,25n)}{0,25n}$
B	N_2	$\frac{1}{12}\,n$	$\dfrac{\left(N_2 - \frac{n}{12}\right)^2}{\frac{n}{12}}$
C	N_3	$\frac{1}{6}\,n$	$\dfrac{\left(N_3 - \frac{n}{6}\right)^2}{\frac{n}{6}}$
D	N_4	$\frac{1}{6}\,n$	"
E	N_5	$\frac{1}{6}\,n$	"
F	N_6	$\frac{1}{6}\,n$	"
	N		E

On sait que E suit une loi du χ^2 à 5 degrés de liberté.

Si le modèle proposé est le bon, on sait calculer la probabilité pour que $E < \chi_0^2$ ou encore trouver χ_0^2 pour que prob $(E > \chi_0^2) = \alpha$

En effet, si E est supérieur à cette valeur χ_0^2, cela peut signifier deux choses:

- ou bien le modèle est mauvais, auquel cas on a raison de le rejeter;

- ou bien le modèle est bon, mais seul le hasard de l'échantillonnage nous a fait croire le contraire (et cette éventualité n'a qu'une probabilité α de se produire...; α apparaît bien alors comme la probabilité pour que nous rejetions une hypothèse qui est juste).

<u>Réalisation du test 15</u>

. Prélever une certaine quantité de graines. Il faut faire en sorte qu'il y ait au minimum 8 à 10 graines de chaque type (ce qui implique n de l'ordre de la soixantaine au moins); s'il n'en était pas ainsi, c'est que "vraisemblablement" le modèle serait mauvais.

. Se fixer α , risque du vendeur.

Trouver χ^2_0 tel que prob $(\chi^2(5) > \chi^2_0) = \alpha$ par lecture dans une table.

. Calculer l'écart e entre l'observation et le modèle théorique.

<u>Règle de décision</u>

Si $e > \chi^2_0$ ne pas accepter l'hypothèse suivant laquelle le sac est bon, c'est-à-dire ne pas le vendre.

Si $e \leqslant \chi^2_0$ rien ne s'oppose à ce que le sac soit bon, mais... on n'a pas calculé β.
Dans ce dernier cas, nous avons deux attitudes possibles :
ou bien:
"puisque je ne rejette pas mon hypothèse, je l'accepte" .
ou bien:
"je vais chercher de l'information ailleurs".

Voici donc proposé un certain nombre d'activités plus ou moins élaborées pour aborder le même problème: quelle décision prendre ?
Il n'est pas question ici d'expliquer en détail la philosophie des tests d'hypothèses, il s'agit principalement d'évoquer assez concrètement le problème dans l'espoir qu'un lecteur (au moins) sera intéressé par le problème... Le but de ce petit exposé est de montrer "à quoi ça sert". Beaucoup

d'enseignants connaissent de nombreux concepts et ne connaissent pas les applications qu'ils peuvent en faire... Or, notre époque ne pousse pas les jeunes vers les spéculations gratuites, ils souhaitent savoir vers quoi on les conduit, leur confiance en leurs maîtres n'est plus aussi complète (peut-être en apparence) qu'autrefois.... Il appartient donc aux enseignants de restaurer cette confiance en leur montrant qu'ils savent très bien où ils vont et quels sont les buts de leur enseignement.

N.B. - L'article ci-dessus n'est en réalité que la trace écrite (et à peine modifiée) de notes ayant servi à un exposé oral; que le lecteur veuille bien nous excuser s'il n'y retrouve pas le style habituel des articles conçus pour la lecture.

$$\alpha = \sum_{k=0}^{r-1} C_n^k \ (0{,}25)^k \ (0{,}75)^{n-k}$$

$$\beta = \sum_{k=r}^{n} C_n^k \ \left(\frac{1}{6}\right)^k \ \left(\frac{5}{6}\right)^{n-k}$$

Les courbes $\alpha(r)$, $\beta(r)$ ont été dessinées pour n = 10, 20, 30, 40, 50 et 60.

Les courbes vont donc par paires ; les points d'intersection (équilibre des risques α et β) ont été grasseyés ; ils correspondent de gauche à droite aux valeurs de n citées ci-dessus.

On constate que les risques s'équilibrent à un niveau d'autant plus bas que n est plus grand ; et que les abscisses de ces points d'intersection ne sont pas nécessairement des entiers (ce qui rend impossible dans la pratique l'équilibrage parfait des risques).