
ATTENTION ! UN MODÈLE PEUT EN CACHER UN AUTRE

Les avatars de certains sujets de Bac habillés par des situations "pseudo-concrètes"

Hubert RAYMONDAUD, Lycée Agricole de Carpentras
Michel HENRY, IREM de Franche-Comté (1)

I. LA MODÉLISATION EN QUESTION

Selon le Petit Robert, un modèle est "une représentation simplifiée d'un processus, d'un système". Cette définition nous paraît insuffisante, un modèle pour être opératoire ne peut se limiter à une représentation : il permet d'interpréter, et d'inférer. Ainsi, un modèle que nous appellerons "pseudo-concret" apparaît comme un intermédiaire nécessaire dans le processus de mathématisation. Pour ce qui concerne les phénomènes aléatoires, "...Pour approfondir (la notion d'expérience aléatoire) on

est inmanquablement amené à séparer la description de situations réelles, des modèles simplifiés qui permettent de les mathématiser..." (2).

La modélisation, bien que le mot n'y soit pas mentionné, est au cœur du chapitre "Probabilité" des programmes de Première et Terminale, toutes sections confondues. Les programmes précisent : "...Pour introduire la notion de probabilité, on s'appuiera sur l'étude de séries statistiques obtenues par répétition d'une expérience aléatoire...". Il semble donc bien qu'il faille

(1) Cet article, imaginé et rédigé par Hubert RAYMONDAUD, a fait l'objet de nombreux échanges avec Michel HENRY, responsable en outre dans le dernier exemple (un problème de pêche miraculeuse), d'une échappée stratosphérique.

(2) Selon Michel HENRY dans le livre de la commission inter-IREM "Statistique et Probabilités" : *Enseigner les probabilités au lycée*, Ed. IREM de Reims, octobre 1997 (que l'on peut se procurer dans tous les IREM), introduction à la partie B : *modélisation en probabilités*, p. 55.

ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE

partir d'une expérience réelle, en décrire et en analyser les résultats, et s'appuyer sur ce travail pour introduire le modèle mathématique dans lequel prend place la notion de probabilité.

C'est encore de modélisation qu'il s'agit quand le programme de Première annonce : "...L'objectif est d'entraîner les élèves à décrire quelques expériences aléatoires simples et à calculer des probabilités..." parce que la réalité ne se réduit pas à des expériences aléatoires simples. Elles ne peuvent donc être que l'invention d'un professeur de mathématiques, qui présente alors des modèles. Et si ce professeur se *hasarde* à décrire une expérience réelle, il va bien falloir qu'il la modélise (la simplifie et l'idéalise) pour pouvoir y appliquer les mathématiques du programme.

Dans les autres chapitres des programmes de mathématiques, il est implicitement question de modélisation : "...On exploitera largement des situations issues de l'algèbre, de la géométrie, des sciences et techniques et de la vie économique et sociale...". La réalité économique et technique est mise sur le même plan que l'algèbre et la géométrie! Il n'est pas précisé que la réalité en question doit être interprétée pour se soumettre à un traitement mathématique. Encore moins comment on peut contrôler et exploiter les résultats mathématiques obtenus.

De même, la modélisation n'apparaît pas clairement dans les propositions méthodologiques suivantes : "...On dégagera les différentes phases du traitement d'un problème : mise en équation, résolution, contrôle et exploitation des résultats..." et "...Formuler un problème, conjecturer un résultat, expérimenter sur

des exemples, bâtir une démonstration, mettre en œuvre des outils théoriques, mettre en forme une solution, contrôler les résultats obtenus, évaluer leur pertinence en fonction du problème posé, ne sont que des moments différents d'une même activité mathématique...".

C'est pourtant bien de la modélisation que l'on fait (peut-être souvent sans le savoir) quand on formule un problème et qu'on le "met en équation". De même le contrôle de la pertinence des résultats et leur exploitation ne peut se faire qu'en liaison avec les hypothèses faites dans le modèle et en tenant compte de la précision exigée pour la réponse finale. Par exemple, un modèle polynomial d'ajustement d'une loi probabiliste peut s'avérer pertinent, mais il lui sera préféré un modèle exponentiel dans lequel les paramètres auront une signification concrète ; un modèle probabiliste hypergéométrique peut être préféré à un modèle de type binomial car il sera plus précis...

Par contre la modélisation figure en bonne place dans : "...On marquera les différentes phases : modélisation, traitement mathématique, contrôle et exploitation des résultats..." où elle est bien séparée du traitement mathématique. Elle fait bien partie intégrante du programme et devrait donc faire l'objet d'un travail en classe de mathématiques.

Malheureusement, en probabilités, les indications du programme sur ce qu'il convient de faire dans le vaste domaine de la modélisation sont plutôt maigres, voire inexistantes. Les questions d'approximation n'y sont pas abordées, ce qui pose problème. La pratique des exercices et des sujets d'examen montre que l'on en fait grand usage sans aucune rigueur dans leur

justification, ni la moindre pertinence en fonction du problème posé (3).

Nous avons donc regardé quelques exercices de baccalauréat sous l'angle de la modélisation. Il apparaît que la plupart présentent une situation artificielle, habillage "pseudo-concret" d'une situation probabiliste simple qu'il s'agit de décrypter pour y reconnaître un modèle accessible. La confusion entre réalité et modèle fait alors partie du jeu contractuel. Elle débouche souvent sur des ambiguïtés ingérables. Nous vous présentons seulement quatre exemples typiques pour les commenter de ce point de vue.

II. ANALYSES DE SUJETS DE BAC

II.1. Principes d'analyse

Afin de conduire nos analyses des sujets de bac, nous reprenons les propositions de M. HENRY (4) qui présente la modélisation en trois étapes :

- 1.- observer la situation réelle et la décrire en termes courants,
- 2.- traduire cette description en un système simplifié et structuré : le modèle pseudo-concret,
- 3.- formaliser ce modèle pseudo-concret pour en faire un modèle mathématique.

(3) Certains libellés de programmes ne sont pas exempts de contradiction. Ainsi, le programme de probabilité de Terminale S.T.A.E. "se limite aux univers finis", ce qui ne l'empêche pas trois paragraphes plus bas de faire figurer sous cette même rubrique l'utilisation de la loi de GAUSS. On n'en est pas à une correction de continuité près.

(4) Dans *Enseigner les probabilités au lycée (op. cit.)*, article : *Notion de modèle et modélisation dans l'enseignement*, p. 77.

Jean-Claude GIRARD (5) propose une définition de l'expérience aléatoire mathématique :

"Une expérience aléatoire mathématique (celle que l'on trouve dans les énoncés) est un modèle d'expérience aléatoire réelle. Il suppose que soient vérifiés les trois points suivants :

exp 1 - La description des conditions de l'expérience réelle (protocole expérimental) détermine son modèle de façon précise et suffisante pour en admettre l'unicité.

exp 2.- L'ensemble des issues possibles que l'on a choisi de considérer est parfaitement déterminé.

exp 3.- On ne peut prévoir quel résultat l'expérience va produire".

Nos analyses critiques s'attachent à mettre en évidence les insuffisances des énoncés ne permettant pas une modélisation simple et unique par les modèles figurant au programme, et ceux ne permettant pas la détermination unique de l'univers des issues possibles. Nous soulevons les problèmes de pertinence de certains énoncés pseudo-concrets et des approximations qu'ils supposent implicitement.

Nos propositions sont guidées par l'idée qu'il est préférable de réfléchir à la modélisation, comme nous y invitent implicitement les programmes, plutôt qu'apprendre à appliquer un ou deux modèles, par simple effet de contrat didactique.

(5) Dans *Enseigner les probabilités au lycée (op. cit.)*, article : "Qu'est-ce qu'une expérience aléatoire ?", p. 60.

ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE

II.2. Un cas fréquent : une situation hypergéométrique traitée par un modèle binomial

Dans beaucoup de sujets d'exercices de Bac, l'expérience aléatoire décrite (quand elle l'est effectivement) présente une situation de tirages aléatoires simultanés ou successifs sans remises dans une population (tirages "exhaustifs"). Ce type de situations relève d'un modèle dit hypergéométrique ⁽⁶⁾ (nous le désignerons par "**modèle de type H**"), dont l'étude n'est pas au programme. Par contre les situations de tirages aléatoires successifs avec remises (schéma de Bernoulli) font appel au modèle binomial (que nous appellerons "**modèle de type B**"), le seul modèle inscrit au programme des séries ES, STAE et STPA. Il n'est pas explicitement au programme de la série S mais beaucoup de problèmes de Bac s'y ramènent très directement.

Les élèves sont donc conduits, par effet de contrat, à traiter une situation de type H avec un modèle de type B. Ce qui théoriquement est erroné, les formules générales obtenues pour le calcul d'une probabilité dans l'un et l'autre modèle sont différentes. Mais lorsque la population est assez vaste devant le nombre limité de tirages effectués (dans un rapport de 1 à 100 par exemple), le modèle de type B est une bonne approximation du modèle de type H. On explique cela qualitativement aux élèves en leur disant que le tirage sans remise d'un élément de la population ne

modifie pas sensiblement sa composition et le tirage suivant peut-être valablement représenté par la même épreuve de Bernoulli.

Les problèmes quantitatifs d'approximation du modèle de type H par le modèle binomial ne sont pas au programme et il ne nous semble pas souhaitable de les rendre exigibles à ce niveau, alors qu'ils font partie intégrante de programmes de BTS.

Il est cependant regrettable que les seules possibilités de justification offertes aux élèves, placés devant un énoncé de type H, soient des "options par défaut" comme : "*puisque l'on ne peut rien faire d'autre, appliquons ce que nous savons sur le schéma binomial*" ou bien de façon plus diplomatique (ménageons la susceptibilité des correcteurs) "*supposons que les hypothèses d'approximation (de quoi par quoi ?) soient vérifiées*".

Dans le cas où une justification serait envisageable, il est dommage que l'on ne puisse en général baser les hypothèses de l'approximation sur aucune donnée numérique de l'énoncé. Ne serait-il pas souhaitable de donner aux élèves tous les éléments permettant de prendre une décision, selon une démarche scientifique, basée sur un seuil de précision (par exemple : donner un résultat à 10^{-3} près) ou sur un rapport d'effectifs acceptable (le rapport de l'effectif prélevé à la taille de la population serait, par exemple, inférieur à 0,01) ?

Plutôt que de faire croire (confusion modèle-réalité) à un problème concret, dont il manque malheureusement les données essentielles, l'énoncé pourrait dire clairement que l'on se situe au niveau d'un modèle (pseudo-concret), en suppo-

(6) Pour une présentation comparative et synthétique des modèles des lois discrètes, on pourra se référer au tableau synoptique p. 252 du livre de la commission inter-IREM "Statistique et Probabilités" : *Enseigner les probabilités au lycée*.

sant vérifiées toutes les hypothèses nécessaires aux approximations par le modèle de type B (ce qui n'empêcherait pas de demander de les préciser). Pour illustrer

cette ambiguïté, voici un sujet de bac (7), choisi "au hasard" parmi une riche collection. Nous soumettons à votre appréciation l'analyse qui le suit.

SUJET I : Bac. Technologique, séries STAE et STPA, sujet national, session normale, 1996.

Exercice 1 (7 points) . Les questions 1), 2), 3) sont indépendantes.

Un fleuriste cherche des fleurs à acheter auprès de ses fournisseurs. Il dispose de dix adresses de fournisseurs et veut en choisir quatre.

1) De combien de façons peut-il effectuer ce choix ?

2) Un des fournisseurs choisis fournit des fleurs bon marché mais dont une sur dix présente un défaut indiscernable (manque de résistance dans le temps). Le fleuriste constitue, avec ces fleurs, des bouquets de cinq fleurs. L'expérience consiste à choisir au hasard l'un de ces bouquets.

Déterminer (on pourra utiliser – en justifiant – une table de la loi binomiale) :

a) La probabilité pour que ce bouquet soit constitué de cinq fleurs présentant le défaut.

b) La probabilité pour que ce bouquet soit entièrement sain.

c) La probabilité pour que ce bouquet comporte 3 fleurs saines et 2 fleurs présentant le défaut.

3) Avec les fleurs d'un autre fournisseur, le fleuriste constitue des bouquets qu'il compte tous vendre au même prix. Ce prix suit une loi normale (Laplace-Gauss) de moyenne 50 francs et d'écart-type 10 francs. On prend un bouquet au hasard.

a) Déterminer la probabilité pour que le prix du bouquet soit inférieur à 70 francs.

b) Déterminer la probabilité pour que le prix de ce bouquet soit compris entre 40 et 60 francs.

Commentaires :

A) Le 1°) n'a aucun rapport avec le reste du problème, c'est dommage.

B) Dans le 2°) on est sans ambiguïté dans le cas d'un **tirage sans remise** donc, en toute rigueur, pas dans le cas d'un modèle binomial de type B, mais dans celui d'un **modèle de type H**. Celui-ci nécessite de connaître la composition exacte de la

population de fleurs dans laquelle sont choisies les 5 fleurs des bouquets : le nombre total m de fleurs présentes chez le fleuriste et le nombre A de fleurs présen-

(7) On trouvera un autre exemple, celui du boulanger (Bac. général, série ES, sujet national, juin 1995) analysé en détail par Jean-Claude GIRARD dans l'article : *Un exemple de confusion modèle-réalité* paru p. 70 dans le livre *Enseigner les probabilités au lycée (op. cit.)*.

**ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE**

tant le défaut. Voyons quelques détails sur la démarche de résolution :

* on reconnaît un tirage simultané, il n'y a donc ni remise, ni ordre.

* L'univers observable Ω (celui que l'on noterait si l'on réalisait pratiquement l'expérience) est différent de l'univers des possibles Ω envisageable, car on ne distingue pas les fleurs saines (S) entre elles, ni les défectueuses (D).

* Ω' peut être représenté par l'ensemble des groupes de 5 lettres, avec répétition, que l'on peut former avec des S et des D. Ω peut être représenté par l'ensemble des groupes de 5 lettres que l'on peut former avec m lettres dont A lettres D distinctes (D_1, D_2, \dots, D_A) et $m - A$ lettres S distinctes (S_1, S_2, \dots, S_{m-A}). C'est sur Ω que l'on pourra faire l'hypothèse d'équiprobabilité.

* On a alors :

$$a) P(5 \text{ fleurs défectueuses}) = \frac{C_A^5}{C_m^5};$$

$$b) P(5 \text{ fleurs saines}) = \frac{C_{m-A}^5}{C_m^5};$$

$$c) P(3 \text{ fleurs saines et } 2 \text{ défectueuses}) = P(3 \text{ saines}) = P(2 \text{ défectueuses}) = \frac{C_A^2 \times C_{m-A}^3}{C_m^5}.$$

Afin de montrer l'influence de la composition de la population de fleurs sur la précision de l'approximation par le modèle de type binomial, nous proposons de comparer les calculs et les résultats de la question c), dans les deux modèles (cf. *tableau ci-dessous*).

* Pour utiliser le modèle de type binomial, on fait comme si l'on avait un tirage avec remise. On néglige donc le changement de composition de la population dû au tirage des fleurs. Ce changement apparaît bien lorsque l'on compare terme à terme les produits donnant les résultats.

* On remarque bien que les valeurs calculées par l'approximation binomiale se

m : taille de la population de fleurs chez le fleuriste	20	200	2000
A : nombre de fleurs défectueuses et proportion p	2 0,1	20 0,1	200 0,1
n : taille du bouquet (l'échantillon) et rapport n/m	5 0,25	5 0,025	5 0,0025
Réponse à c) : valeurs calculées avec le modèle de type H (hypergéométrique)	$\frac{C_2^2 \times C_{18}^3}{C_{20}^5} \approx 0,0526$	$\frac{C_{20}^2 \times C_{180}^3}{C_{200}^5} \approx 0,0716$	$\frac{C_{200}^2 \times C_{1800}^3}{C_{2000}^5} \approx 0,0728$
Réponse à c) : approximation du modèle de type H par le modèle de type B (binomial)	$C_5^2 \times (0,1)^2 \times (1 - 0,1)^3 \approx 0,0729$	$C_5^2 \times (0,1)^2 \times (1 - 0,1)^3 \approx 0,0729$	$C_5^2 \times (0,1)^2 \times (1 - 0,1)^3 \approx 0,0729$

rapprochent de plus en plus des valeurs obtenues avec le modèle de type H, lorsque la proportion n/m diminue : dans ce cas particulier un rapport n/m de 0,025 assure une précision de l'approximation à 10^{-3} près, un rapport de 0,0025 donne une précision à 10^{-4} près.

* Rappelons que le modèle de type H est aussi une approximation de la réalité, dans la mesure où l'équiprobabilité n'est jamais exactement vérifiée. L'équiprobabilité peut être contrôlée par un protocole expérimental qui assure seulement que l'on ne s'en écarte pas trop.

Dans la mesure où il n'existe pas de relation simple entre la précision de l'approximation binomiale et les paramètres du modèle de type H, une démarche rigoureuse consiste à comparer les résultats dans les intervalles de valeur des variables et paramètres fixés par le problème. A précision équivalente, on choisira le modèle le plus simple à mettre en œuvre. Pourquoi ne pas imaginer des exercices basés sur cette démarche ? (8)

C) L'énoncé du 3°) n'est pas clair : le prix est-il fixé comme c'est dit dans la première phrase, ou bien aléatoire comme le propose la deuxième ? On ne voit d'ailleurs pas du tout d'où vient le caractère aléatoire annoncé. Sur quel univers travaille-t-on ? La confusion modèle-réalité bat son plein, le prix réel ne suit pas une loi Gaussienne, qui de plus est un mauvais modèle pour ce

(8) Par exemple, dans l'exercice du sujet national du Bac général, série S en 1996 (on demandait à trois inscrits dans un club sportif, choisis "au hasard", de remplir un questionnaire), les consignes de correction étaient de refuser l'application de la loi binomiale alors qu'elle donnait une approximation acceptable quant à la précision demandée dans les calculs.

genre de variable. Ainsi le passage au modèle pseudo-concret est mal fait et complètement artificiel.

II.3. Exemple de situations alambiquées où l'on ne sait plus comment interpréter les "tirages au hasard" indiqués dans l'énoncé

SUJET II : Bac. Technologique, séries STAE et STPA, sujet national, session normale, 1997

Exercice 1 (4 points). L'utilisation des calculatrices et du formulaire est autorisée.

On considère que 20 % des pièces de 10 F qui sont en circulation dans une ville donnée sont fausses.

Dans une laverie de cette ville, les machines à laver fonctionnent avec des pièces de 10 F.

Si une pièce est fautive, la machine la refuse 4 fois sur 5.

Si une pièce est vraie, la machine l'accepte à coup sûr.

1) Si un client met une pièce de 10 F prise au hasard, quelle est la probabilité pour que cette pièce soit refusée par la machine ?

2) L'une des machines de la laverie nécessite que l'on y mette successivement exactement 5 pièces de 10 F.

Soit X la variable aléatoire présentant le nombre de pièces (sur les 5) rejetées par la machine.

**ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE**

a) Justifier que X suit la loi binomiale de paramètres $n = 5$ et $p = 0,16$.

On exprimera les résultats suivants à 10^{-3} près :

b) Calculer la probabilité pour qu'aucune pièce ne soit rejetée.

c) Calculer la probabilité pour qu'une seule pièce soit rejetée.

d) Calculer la probabilité pour qu'au moins deux pièces soient rejetées.

Il faut donc que l'élève imagine comment pratiquement (parce que les élèves cherchent souvent à se raccrocher à une image concrète) il pourrait "prendre une pièce de 10 F au hasard" parmi toutes les pièces en circulation dans une ville. Ça n'a rien d'évident. N'y aurait-il pas, dans l'esprit du rédacteur de l'énoncé, une confusion entre le hasard de la contingence qui a amené ces pièces dans le porte monnaie en question, et le hasard dont on s'occupe en probabilités à ce niveau, celui du prélèvement aléatoire dans une population ? Passons sur cette remarque philosophique.

Dans cet exercice, tel qu'il est habillé, il n'est donc pas facile d'imaginer concrètement l'univers des possibles dont on a besoin.

Commentaires

a) Afin d'éviter la confusion modèle-réalité nous proposons de préciser simplement que les valeurs $4/5$ et $1/5$, données pour le fonctionnement de la machine, sont des hypothèses de travail.

b) Dès le 1°) il y a ambiguïté sur l'univers à prendre dans le modèle. L'élève peut penser, avec le bon sens pratique qu'il a souvent dans ces sections, que le client prend la pièce de 10 F dans son porte monnaie. Malheureusement tout va l'en dissuader, dans le 1°) et dans le 2°) !

Le premier écueil est que l'on ne connaît pas la composition en pièces de 10 F vraies et fausses de ce fameux porte monnaie. Rien dans l'énoncé n'empêcherait un élève espiègle d'écrire : *"je fais l'hypothèse que le client est particulièrement vigilant (ce que tout le monde a intérêt à être puisque la seule détention de fausse monnaie est illégale) et qu'il ne laisse passer aucune fausse pièce de 10 F"*. L'exercice en serait fort simplifié !

Si on décide malgré tout que le client va piocher dans son porte monnaie, il faut bien supposer que sa composition est exactement la même que celle de l'ensemble des pièces qui circulent en ville, ce qui ne va pas de soi (mais qui peut-être une hypothèse raisonnable, en vertu du principe du maximum de vraisemblance que nous présenterons dans le dernier exemple). Alors pourquoi ne pas dire simplement dans l'énoncé que l'on supposera que le client possède 20% de pièces fausses de 10 F dans son porte monnaie ?

c) Dans le 2°) surgit le second écueil. Si le client pioche dans son porte monnaie, le bon sens nous fait penser à un tirage sans remise. Et voilà que deux lignes plus loin, on nous demande de justifier, à tort, une loi binomiale, archétype de la modélisation d'un tirage avec remise ! Il n'en faut pas tant pour déstabiliser une tête d'élève même "bien faite". De plus l'expérience se

fait en deux temps, le tirage des pièces puis leur insertion dans la machine, deux épreuves dont on se rend bien compte qu'elles ne sont pas indépendantes puisque pièces vraies et fausses ne sont pas traitées de la même manière !

Il aurait été plus judicieux de parler d'approximations, mais comme celles-ci ne sont pas explicitement au programme, on en fait sans le dire ! Nous allons montrer que l'approximation binomiale est très mauvaise lorsque l'on fait le tirage dans le porte monnaie.

Le calcul des probabilités avec le modèle le plus approprié (que nous appellerons modèle "correct", par abus de langage), fait intervenir un modèle de type hypergéométrique pour le tirage des pièces puis un modèle de type binomial pour l'épreuve de la machine, les deux modèles étant liés par le nombre de fausses pièces mises dans la machine. Nous n'avons montré que le calcul de $P(X = 0)$, qui nous semble déjà

être d'un bon niveau de difficulté pour les sections concernées. Le tableau suivant permet de faire les comparaisons avec le modèle de type binomial (cf. tableau ci-dessous).

Pour le porte monnaie de Monsieur Toutlemonde (5 pièces de 10 F dont 1 fausse), le modèle de type binomial et le modèle "correct" donnent respectivement 0,4182 et 0,2. Pour des bourses plus garnies, 20 pièces de 10 F dont 4 fausses et 100 pièces de 10 F dont 20 fausses, le modèle "correct" donne respectivement 0,3846 et 0,4120. Même dans le dernier des cas, on est loin de la précision à 10^{-3} demandée par l'énoncé.

L'élève qui veut réfléchir à une modélisation plutôt qu'appliquer le modèle imposé va au devant de difficultés certaines. Aussi n'est-il pas étonnant de le voir sauter allègrement la justification pour se réfugier dans une ribambelle de calculs rassurants et sans doute payants !

Hypothèses de composition du porte monnaie	$m = 5$ pièces dont $A = 1$ fausse (proportion de fausses = 0,2)	$m = 100$ pièces dont $A = 20$ fausses (proportion de fausses = 0,2)
$P(X = 0)$, calcul par l'approximation binomiale $B(n = 5 ; p = 0,16)$ du modèle "correct"	$(0,16)^0 \times (1 - 0,16)^{5-0} \times C_5^0 = 0,4182$	$(0,16)^0 \times (1 - 0,16)^{5-0} \times C_5^0 = 0,4182$
$P(X = 0)$: calcul avec le modèle "correct", de type $H(m ; n = 5; A)$ (tirages sans remise)	$\frac{C_4^4 \times C_1^1}{C_5^5} \times \left(\frac{4}{5}\right)^0 \times \left(1 - \frac{4}{5}\right)^{1-0} \times C_1^0$ tirage des pièces ; passage dans la machine = 0,2000	$\sum_{k=0}^5 \frac{C_{80}^{5-k} \times C_{20}^k}{C_{100}^5} \times \left(\frac{4}{5}\right)^0 \times \left(1 - \frac{4}{5}\right)^{k-0} \times C_k^0$ tirage des pièces ; passage dans la machine $\approx 0,4120$

ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE

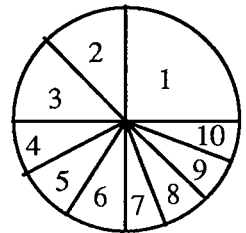
II.4. Comment gérer une situation multinomiale ?

SUJET III : Bac. général, série ES, sujet national, juin 1996, spécialité

Exercice 2

Dans une fête foraine, une loterie utilise une roue circulaire tournant autour d'un axe et une flèche fixe déterminant la position d'arrêt de la roue. Cette roue est partagée en 10 secteurs tels que :

- le secteur 1 occupe le premier quart de la roue ;
- les secteurs 2 et 3 se partagent également le deuxième quart ;
- les secteurs 4, 5 et 6 se partagent également le troisième quart ;
- les secteurs 7, 8, 9 et 10 se partagent également le dernier quart.



Quand la roue est lancée, elle s'arrête de façon aléatoire, et la flèche ne peut indiquer qu'un seul secteur.

1) Le nombre n étant un entier de $[1, 10]$, la probabilité pour que la flèche indique le secteur n est notée p_n . On suppose qu'elle est proportionnelle à l'angle au centre de ce secteur.

Calculer p_1 , p_2 , p_4 , p_7 . (Les résultats seront donnés sous forme de fractions irréductibles).

2) Le jeu proposé est le suivant :

Le joueur mise une certaine somme.

Il perd sa mise si la flèche indique les secteurs 1, 2, 4 ou 7.

Sa mise lui est remboursée si la flèche indique 3, 5 ou 8.

Il gagne le double de sa mise si la flèche indique un autre secteur.

a) Montrer que la probabilité p'_1 pour que le joueur perde est égale à $25/48$, et que la probabilité p'_2 pour qu'il soit remboursé vaut $13/48$.

b) Calculer la probabilité p'_3 pour que le joueur gagne et celle p'_4 pour qu'il ne perde pas.

3) Un joueur joue 5 parties.

(Dans les questions suivantes les résultats seront arrondis à 0,001 près).

a) Calculer la probabilité p'_5 pour qu'il gagne au moins 4 fois.

b) Calculer la probabilité p'_6 pour qu'il perde deux fois et qu'il ne perde pas trois fois.

c) Calculer la probabilité p'_7 pour qu'il gagne deux fois et qu'il ne perde pas trois fois.

Commentaires

Dans le 3), l'expérience consiste à jouer cinq parties.

Il n'est pas précisé si ces parties sont indépendantes, mais le bon sens et les exercices analogues nous conduisent à cette hypothèse, que rien dans l'énoncé n'interdit.

L'élève va sans doute penser au modèle de type binomial et en chercher les justifications. Malheureusement il n'y pas deux possibilités opposées à chaque tirage mais trois (si on peut dire) ! Il va croire que le modèle de type binomial ne convient pas. Et il n'a pas tort puisque l'on est en présence d'un modèle multinomial ("trinomial"). Ainsi, il faut comprendre dans cet énoncé que le contraire de gagner n'est pas de perdre, on peut aussi être remboursé !

Dans le a) et le b) on se ramène, en conformité avec le programme, à une loi binomiale en ne considérant que les alternatives gagner - ne pas gagner, ou bien perdre - ne pas perdre, mentionnées dans l'énoncé.

Mais pour le c), il faut abandonner le modèle de type binomial pour faire un dénombrement détaillé des combinaisons de 5 des lettres G, R, P (qui représentent les issues "gagner", "être remboursé" et "perdre"), réalisant l'événement "il gagne 2 fois et il ne perd pas 3 fois". Admettons que cet événement soit la conjonction de l'événement "il gagne exactement 2 fois" et de l'événement "dans l'ensemble il ne perd pas 3 fois". Le joueur peut donc perdre deux fois, une fois ou zéro fois. L'événement dont on cherche la probabilité est alors constitué des issues {GGPPR}, {GGPRR}, {GRRRR}. Mais on peut aussi valablement

interpréter l'indication "il ne perd pas" par "il gagne ou il est remboursé" et l'événement cherché doit alors être interprété par "à trois reprises exactement, il gagne ou il est remboursé". Cet événement est représenté par la seule partie {GGPPR}, car GG réalise déjà deux fois "ne pas perdre", R réalisant la troisième fois.

On a dans ce cas, par exemple,

$$P(\{GGPPR\}) = \frac{10}{48} \times \frac{10}{48} \times \frac{25}{48} \times \frac{13}{48} \times \frac{13}{48} \times \frac{5!}{2!2!1!} \approx 0,04974$$

↑ gagner 2 fois
↑ perdre 1 fois ;
↑ être remboursé 2 fois
↙ nombre de "chemins"

Le calcul du nombre de "chemins" ou nombre de permutations avec répétitions ne nous semble pas faire partie des cas simples préconisés par les programmes, le calcul précédent peut-il être attendu d'un élève de Terminale le jour du bac ?

Outre la logique floue dont elle procède, qui peut à juste raison déstabiliser un candidat qui hésiterait entre les deux interprétations, l'intérêt de cette dernière question ne nous apparaît pas clairement dans la mesure où sa principale différence avec les questions précédentes semble être une difficulté technique de dénombrement.

III. DES EXERCICES QUI MORDENT LE TRAIT

III.1. Un énoncé écologique

Voici l'énoncé d'un problème d'entraînement au Bac qui change un peu des loteries, urnes et cartes traditionnelles :

ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE

**SUJET IV. proposé dans le manuel *Transmath* de Terminale ES (p. 311) ⁽⁹⁾ :
"BAC" Thèmes : Loi binomiale. Étude de fonctions puissances.**

Un étang contient un nombre N , inconnu mais défini de poissons. L'objet du problème est de proposer une évaluation de N , basée sur des hypothèses bien définies.

1. On pêche dans différents endroits de l'étang ; on en sort 20 poissons que l'on marque et que l'on remet vivants dans l'étang après les avoir marqués. Quelques jours plus tard, on effectue une nouvelle pêche dans des endroits variés du même étang en rejetant à l'eau les poissons pêchés, après avoir noté s'ils sont marqués ou non. On prend ainsi 50 poissons dont 4 sont marqués.

On suppose qu'entre les deux pêches, la population de l'étang n'a pas varié et que lors de la seconde pêche, à chaque coup il y a équiprobabilité de sortie pour chacun des N poissons de l'étang.

Avant la seconde pêche, on pouvait se poser le problème : quelle est la probabilité de sortir les poissons marqués sur les 50 pêchés ?

Répondez à cette question en donnant l'expression générale de $P(X = k)$ où X désigne la variable aléatoire "nombre de poissons marqués que l'on peut sortir sur 50 pêchés".

2. f est la fonction qui, à un réel x supérieur à 20, associe : $f(x) = \left(\frac{20}{x}\right)^4 \cdot \left(1 - \frac{20}{x}\right)^{46}$

Montrez que f a un maximum. Pour quelle valeur a de x ce maximum est-il atteint ?

3.a. Si l'on admet que l'événement $(X = 4)$ qui est réalisé correspond à l'événement de probabilité maximale parmi tous les événements possibles $(X = k)$, $k = 0, \dots, 50$, quelle valeur doit-on attribuer à la population N de l'étang ?

b. Cette hypothèse conduit à la même conclusion qu'une autre hypothèse très simple que l'on aurait pu faire pour évaluer N . Laquelle ?

**III.2. Commentaires sur l'habillage
proposé par cet énoncé**

Voici encore un énoncé maladroitement

habillé pour "faire plus vrai" d'un modèle déjà là (le schéma de Bernoulli) ⁽¹⁰⁾ ! Le thème est séduisant, mais dans sa formulation il a l'inconvénient didactique de

(9) Communiqué par Geneviève LORIDON (IREM de Clermont Ferrand), que nous remercions chaleureusement pour sa contribution à cet article.

(10) Ce type d'estimation de la taille de la population d'un étang, dite "par capture et recapture" se pratiquait effectivement autrefois.

confondre modèle et réalité, ou plutôt de chercher à évacuer dans les implicites le statut de modèle. Par exemple, lors de la pêche des 50 poissons il n'est pas précisé si chaque poisson est rejeté juste après avoir été observé et avant de remettre la ligne à l'eau, ou bien si l'on attend d'avoir pêché et observé les 50 poissons avant de les rejeter ensemble dans l'étang. Dans le premier cas on aura un modèle de type binomial éloigné de la pratique usuelle, dans l'autre un modèle de type hypergéométrique, plus approprié mais hors programme !

Comme le rédacteur de l'énoncé veut éviter certaines critiques de non adéquation du modèle binomial, il introduit des précisions comme : *"on effectue une nouvelle pêche dans des endroits variés du même étang"* sans que l'on discerne bien quelle hypothèse de modèle il désire qu'elle induise chez les élèves, et il ajoute : *"On suppose qu'entre les deux pêches, la population de l'étang n'a pas varié"*, pas de naissances donc, et pas de prédateurs, autres poissons, oiseaux ou pêcheurs braconniers... (11)

Pour couper court aux hésitations, l'énoncé précise : *"lors de la seconde pêche, il y a équiprobabilité de sortie..."*, au moins cela évite de se demander si cette hypothèse va de soi, mais alors pourquoi préciser que l'on pêche *"dans des endroits variés" ?...*

III.3. Quel modèle appliquer ?

Le schéma binomial est donc mal distingué du modèle hypergéométrique. Ce dernier (50 poissons pêchés, observés ensemble puis remis) s'impose plus naturellement et les élèves qui sont entraînés, peuvent trouver sans trop de mal une expression générale pour $P(X = k)$, demandée à la fin du 1., soit $\frac{C_{20}^k \times C_{N-20}^{50-k}}{C_N^{50}}$. Mais alors le 2. achèvera de

les rendre perplexes, puisque $f(x)$ ne ressemble pas du tout à ce $P(X = k)$ (si tant est qu'ils s'y retrouvent entre la variable entière X , prenant des valeurs k entre 0 et 50 alors qu'il n'y a que 20 poissons marqués, et la variable réelle x , prenant des valeurs supérieures à 20).

Remarquons qu'avec les données numériques ($50/250$) les deux modèles, binomial et hypergéométrique, sont relativement proches ($P(X = 4)$ est donnée à trois centièmes près (cf. le tableau ci-dessous), mais ils sont aussi proches d'un modèle de Poisson (P majuscule) $P(4)$, plus pratique. [Pour les spécialistes, le paramètre de Poisson, espérance de la variable X , est égal au produit np , où $n = 50$ et $p = 0,08$; on a dans ce modèle :

$$P(X = 4) = \frac{4^4 e^{-4}}{4!} = 0,1954]$$

A titre indicatif, comparons les résultats finaux des trois modèles, N est la taille de la population (on propose deux situations : $N = 250$ et $N = 500$), n est la taille de l'échantillon pêché, A est le nombre de poissons marqués, avec $p = A/N = 0,08$:

(11) Dans la pratique, on réalisait les deux pêches suffisamment proches pour admettre que l'hypothèse que la population n'avait pas varié n'était pas trop fautive. De même l'hypothèse d'équiprobabilité des poissons lors de la seconde pêche était plus ou moins vraie selon les espèces pêchées, car certains poissons se méfient plus lorsqu'ils ont déjà été pêchés une première fois.

**ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE**

modèle de Poisson $P(4)$	modèle binomial $B(50; 0,08)$	modèle hypergéométrique $H(N = 250; n = 50; A = 20)$	modèle hypergéométrique $H(N = 500; n = 50; A = 40)$
$P(X = 4) \approx 0,1954$	$P(X = 4) \approx 0,2037$	$P(X = 4) \approx 0,2274$	$P(X = 4) \approx 0,2146$

Même en doublant la taille de la population, l'approximation binomiale ne donne pas encore le centième !

Heureusement, les élèves n'ont que le schéma binomial à leur programme, ils n'ont donc pas à avoir d'états d'âme sur le choix du meilleur modèle et cette phrase de l'énoncé doit être interprétée comme une hypothèse de remise des poissons un par un après chaque tirage.

Dans l'énoncé, la question est mathématisée à leur place : "donner l'expression générale de $P(X = k)$..." et, comme dans le formulaire (ou leur mémoire personnelle) il n'y a qu'une formule qui ressemble à cela, la seule réponse possible est formelle :

$$P(X = k) = C_{50}^k \left(\frac{20}{N}\right)^k \left(1 - \frac{20}{N}\right)^{50-k}$$

Doutons que dans ces conditions il y ait un seul élève pour chercher à justifier l'emploi de la loi binomiale à partir des données aussi alambiquées de l'énoncé.

Alors, à quoi sert tout l'habillage, si c'est pour faire copier une formule ? Ne valait-il pas mieux la demander clairement ? Ou mieux, demander les hypothèses nécessaires pour qu'un modèle, d'urne par exemple, conduise à l'étude d'une variable binomiale, dont on donnerait les probabilités élémentaires. Peut-on trouver satisfaisante une résolution qui doit plus aux effets de contrat qu'à une véritable maîtrise de la modélisation ?

III.4. Remarques sur la deuxième question, un peu d'analyse

La question 2. tombe ensuite comme un cheveu sur la soupe. Pourquoi a-t-on fait $k = 4$ et pourquoi cette recherche de maximum ?

Dans l'étude de la dérivée de la fonction f , quels sont les élèves, habiles en analyse, qui seront passés par l'étude de $\ln f(x)$?

La plupart devraient obtenir :

$$f'(x) = -6,4 \cdot 10^{-5} (x - 250)(x - 20)^{45} x^{-51}$$

pour $x > 20$ et donc un maximum pour $x = 250$!

Il faudra ensuite relier ce x réel à N entier, ce qui n'a rien d'évident. Les programmes précisent souvent de se limiter à des exercices ne comprenant aucune difficultés techniques de calcul ! Notre connaissance des classes de Terminale nous fait présager quelques difficultés de calcul face à des expressions comprenant des valeurs numériques trop compliquées comme celles de f et de f' .

III.5. Remarques sur le 3.a : de quel maximum s'agit-il ?

Dans 3.a, il est question de maximum dans un ensemble qui n'est pas le même que l'ensemble évoqué au 2. et auquel on semble faire référence. $P(X = 4)$ est maximale dans l'ensemble des $P(X = k)$, quand k varie, pour un N donné. Cela ne signifie pas pour autant que $P(X = 4)$ conserve cette propriété de maximum

lorsque cette fois l'on fait varier N .

Ainsi, l'hypothèse du maximum de $P(X = 4)$ est faite sur l'ensemble des entiers (k) de 0 à 50 qui est différent des réels supérieurs à 20 de la question 2. A la suite de quoi on demande sans plus attendre de faire le lien avec N pour trouver la taille de la population de l'étang. On s'y perdrait à moins ! On peut se demander si la démarche suggérée par l'énoncé n'est pas erronée. En fait il n'en est rien, mais cela n'est pas du tout évident. Nous allons voir pourquoi, ce qui va occuper la fin de cet article et nous amener à des développements passionnants, mais de plus haut niveau (un peu).

L'hypothèse qui est faite en 3.a. relève de ce que les spécialistes en inférence statistique appellent le "principe du maximum de vraisemblance" et fait intervenir une propriété de la distribution binomiale que l'on n'étudie pas en Terminale.

Le maximum de vraisemblance est introduit succinctement et intuitivement en BTS. On se demande bien comment un élève de Terminale pourrait trouver du sens aux questions 3.a. et 3.b. qui lui font allusion. Il ne nous semble donc pas réaliste de proposer un tel sujet, même correctement développé et rédigé en classes de Terminale.

III.6. Introduction du modèle binomial pour résoudre la question 3

Pour les lecteurs désirant en savoir un petit peu plus sur l'application du principe du maximum de vraisemblance dans ce problème, nous vous proposons les quelques compléments qui suivent.

La question 3.a. qui semble incohérente avec ce qui précède, procède en fait d'une démarche correcte mais tout à fait inacces-

sible, et au niveau du bac, et pour les correcteurs qui n'ont pas eu l'occasion de réfléchir sur les méthodes bayésiennes ou de décrypter *Ars Conjectandi* de Jacques Bernoulli, publié en 1713.

Cette question 3.a. repose donc sur une propriété de maximum des probabilités binomiales dans un schéma de Bernoulli, que l'on trouve dans *Ars Conjectandi* (cette propriété est utile pour la démonstration sophistiquée du *théorème d'or* de Bernoulli, forme particulière de la loi des grands nombres). Elle repose ensuite sur la compréhension de la méthode du maximum de vraisemblance en statistique inférentielle.

Mais auparavant, traduisons en termes de modèle d'urne le schéma des poissons pour préciser clairement les hypothèses.

Soit donc une urne de Bernoulli contenant N boules dont r blanches et $s = N - r$ noires. On pose $p = \frac{r}{N}$, la proportion des boules blanches dans l'urne. L'hypothèse de base dans un modèle d'urne est l'équiprobabilité des boules (*poissons*) dans un tirage "au hasard" de l'une d'entre elles (*pêche dans des lieux divers...*). L'obtention d'une boule blanche est alors un événement de probabilité p .

Soit le schéma de Bernoulli : n tirages successifs "avec remises" de boules de l'urne avec observation des couleurs obtenues. L'hypothèse de modèle dans ce schéma de Bernoulli est que la "remise" fournit la même urne de Bernoulli au tirage suivant (*i.e.* dotée des mêmes hypothèses), ce qui se traduit d'un point de vue probabiliste (axiomes) par le fait que les événements respectivement associés à deux tirages différents sont indépendants : la probabilité de leur conjonction est donc le produit des probabilités de chacun d'eux.

**ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE**

On s'intéresse au nombre X des boules blanches obtenues au cours de n tirages ($n = 50$ pour nous). On sait (le cours) que X suit une loi binomiale $B(n, p)$. Le fait d'obtenir k boules blanches pour un k fixé de 0 à n est donc un événement associé au schéma de Bernoulli dont la probabilité binomiale est donnée par :

$$P(X = k) = C_{50}^k p^k (1 - p)^{n-k}.$$

Dans la suite, il sera plus simple d'utiliser les notations introduites sous la forme :

$$P(X = k) = \frac{n(n-1) \dots (n-k+1) r^k s^{n-k}}{N^n}.$$

Ces probabilités $P(X = k)$, jouissent d'une propriété de maximum que nous allons présenter et qui sera démontrée ensuite.

III.7. Une propriété des probabilités binomiales due à Bernoulli

Propriété de Bernoulli

a) Dans les hypothèses précédentes, la valeur m de k qui rend maximale la probabilité binomiale $P(X = k)$ vérifie :

$$\frac{r}{N}(n+1) - 1 \leq m \leq \frac{r}{N}(n+1).$$

Remarques :

- Si $\frac{r}{N}(n+1)$ n'est pas entier, m est alors unique : $m = \text{Ent}[\frac{r}{N}(n+1)]$, et

$$P(X = m) = C_n^m \frac{r^m s^{n-m}}{N^n}$$

- Si $\frac{r}{N}(n+1)$ est entier, deux indices consécutifs donnent à cette probabilité une même valeur maximale égale à

$$P[X = \frac{r}{N}(n+1) - 1] = P[X = \frac{r}{N}(n+1)].$$

Propriété de Bernoulli (suite)

b) Si l'on ordonne les probabilités binomiales selon les $n+1$ valeurs de k , de 0 à n , ces probabilités vont d'abord en croissant jusqu'à l'indice m tel que $\frac{m}{n+1}$ soit la valeur

inférieure la plus voisine de $p = \frac{r}{N}$ puis décroissent.

On peut visualiser ce phénomène sur le schéma de la page suivante.

La démonstration est inspirée de celle de J. Bernoulli (prop. 2 et 3 de *Ars Conjectandi*). Elle repose sur l'étude du rapport $\frac{P(X = k)}{P(X = k - 1)}$ et de sa position par rapport à 1.

i - On a, pour $k > 0$:

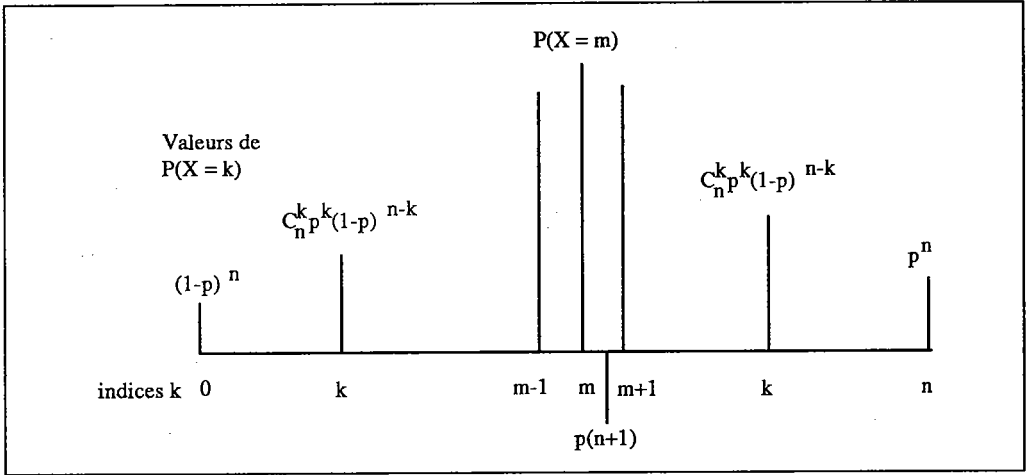
$$\frac{P(X = k)}{P(X = k - 1)} = \frac{n - k + 1}{k} \cdot \frac{r}{s},$$

$$\text{d'où } \frac{P(X = k)}{P(X = k - 1)} \geq 1,$$

ce qui équivaut à $\frac{r}{N}(n+1) \geq k$ (rappelons que $r + s = n$).

ii - Cela montre que si $0 < k \leq \frac{r}{N}(n+1)$, la probabilité binomiale $P(X = k)$ va en croissant.

ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE



iii - On a aussi :

$$\frac{P(X=k)}{P(X=k+1)} = \frac{k+1}{n-k} \cdot \frac{s}{r} \geq 1 \Leftrightarrow k \geq \frac{r}{N}(n+1) - 1,$$

ce qui montre que si

$$\frac{r}{N}(n+1) - 1 \leq k < n,$$

cette probabilité va en décroissant.

iv - $P(X = k)$ passe donc par un maximum pour la ou les valeurs entières de k comprises entre $\frac{r}{N}(n+1) - 1$ et $\frac{r}{N}(n+1)$. Cette valeur

est unique, notée m , si $\frac{r}{N}(n+1)$ n'est pas entier. On a alors $m = \text{Ent}[\frac{r}{N}(n+1)]$ et le maximum vaut : $P(X = m) = C_n^m p^m (1-p)^{n-m}$.

Si $\frac{r}{N}(n+1)$ est un entier m , la valeur commune $P(X = m - 1) = P(X = m)$ est le maximum de la probabilité $P(X = k)$. Cette égalité peut être vérifiée en écrivant le rapport :

$$\frac{P(X = m - 1)}{P(X = m)} = \frac{C_n^{m-1} r^{m-1} s^{n-m+1}}{C_n^m r^m s^{n-m}} = \frac{m}{n-m+1} \cdot \frac{s}{r},$$

ce qui donne effectivement :

$$\frac{P(X = m - 1)}{P(X = m)} = \frac{\frac{r}{N}(n+1)}{(n+1)(1 - \frac{r}{N})} \cdot \frac{s}{r} = \frac{r}{N-r} \cdot \frac{s}{r} = 1.$$

III.8. Le principe du maximum de vraisemblance

Revenons à l'exercice "Bac". On part du principe du maximum de vraisemblance :

Un événement observé à l'issue d'une expérience aléatoire est celui qui, parmi tous les événements comparables, avait la plus grande probabilité d'arriver.

On sait qu'en pratique cela n'est pas vrai, car à partir de la répétition de la même expérience, on observe différents résultats aléatoires qui ne sont donc pas tous de probabilité maximale. Mais, d'après la loi des grands nombres, en moyenne, dans un (très) grand nombre de

**ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE**

telles expériences, l'événement à probabilité maximale s'observe le plus souvent.

A partir de ce principe, à la base des méthodes bayésiennes, on infère que les paramètres inconnus de l'expérience aléatoire ont les valeurs qui rendent maximale la probabilité de l'événement observé. C'est la "méthode du maximum de vraisemblance".

En fait, ces valeurs hypothétiques minimisent (il faut le montrer) le risque de se tromper en les retenant à la place d'autres éventuelles : c'est une inférence statistique, c'est à dire une hypothèse faite sur la valeur du ou des paramètres inconnus. On obtient ces valeurs avec une certaine précision (ou imprécision) qui peut être déterminée, et un risque de se tromper en donnant cette valeur, risque qui peut être aussi calculé ou contrôlé. Une telle inférence, sans ces deux contrôles - précision et fiabilité - n'a pas beaucoup de sens (surtout si on ne sait pas évaluer le coût d'une erreur), car elle conduit à une affirmation qui ne peut être ni appréciée, ni contestée.

Le raisonnement par maximum de vraisemblance, appliqué malgré cela au problème des poissons, nous dit : pour N fixé (inconnu), la valeur $k = 4$ est celle qui rend maximale la probabilité $P(X = k)$ dans les conditions du schéma de Bernoulli. C'est à dire qu'avec les données numériques du problème, on pose $m = 4$, i.e.

$\text{Ent}[\frac{20}{N} \times 51] = 4$, où N est un entier tel que :

$$\frac{20}{N} \times 51 - 1 < 4 < \frac{20}{N} \times 51,$$

ce qui donne $204 < N < 255$.

Mais en appliquant à nouveau le même

principe du maximum de vraisemblance, la valeur de N est celle qui rend maximale la probabilité $P(X = m)$, atteinte pour $m = 4$: $P(X = 4) = C_{50}^4 (\frac{20}{N})^4 (1 - \frac{20}{N})^{46}$, c'est à dire $N = 250$, selon l'étude des variations de cette fonction faite en 2^e question.

III.9. De la question 3a à la question 3b

Ainsi, la question 3.a. est fondée. L'hypothèse de travail posée dans l'énoncé peut sembler surprenante a priori. Mais elle est pertinente du point de vue de l'application du principe du maximum de vraisemblance.

Cependant, elle aura sans doute un effet distracteur pour les bons élèves, celui de les lancer sur l'étude des variations de $P(X = k) = C_n^k p^k (1 - p)^{n-k}$ quand k varie de 0 à n, pour vérifier à quelles conditions (sur N), $P(X = m)$ est bien la valeur maximale de cette probabilité (objectif inaccessible sans indications à ce niveau), car on ne voit pas a priori pourquoi cela serait vrai pour $m = 4$ particulièrement. Par contre pour ceux qui ne se posent pas de question (est-ce le comportement attendu le jour du Bac ?) il devenait auto-mathique (pour reprendre le calembour de Stella Baruck) de renvoyer au maximum obtenu à la question précédente sans comprendre en quoi il concernait cette question 3.a.

On peut enfin se demander si le(s) rédacteur(s) de l'énoncé avai(en)t bien "digéré" les conditions de fonctionnement des méthodes bayésiennes et la validité du principe du maximum de vraisemblance. La réponse de bon sens attendue en 3.b vient confirmer cette remarque : nonobstant son aspect sibyllin, comment cette attente (formuler une hypothèse) peut-elle

être comprise par un élève de Terminale ? Le qualificatif de "très simple" n'apporte rien de plus et est au contraire un facteur déstabilisant et aggravant. En effet, après une hypothèse complexe que l'élève de Terminale ne peut comprendre (celle du principe du maximum de vraisemblance), faute d'avoir déjà réfléchi à sa signification, on lui demande d'en fournir une autre pour obtenir le même résultat, forcément pas évident, puisqu'issu de 3.a épaulé par l'étude des variations de f en 2. Comment peut-il s'attendre à ce que le simple "bon sens" suffise pour répondre à la question ?

Ce "bon sens" serait de faire l'hypothèse (qui relève aussi d'une certaine manière du principe du maximum de vraisemblance) que l'échantillon prélevé dans l'étang, dans les conditions d'un schéma de Bernoulli (ce qui dans l'habillage pseudo-concret de l'énoncé est moins que vraisemblable), est "représentatif" de la population de l'étang, c'est à dire composé proportionnellement à celle-ci : $\frac{4}{20} = \frac{20}{N}$, d'où $N = 250$.

Résultat nécessaire ou magie des données numériques ? On retrouve miraculeusement la même valeur par ce raisonnement de proportionnalité (qui n'est en rien probabiliste) que dans l'étude des variations de la fonction f qui n'a rien à voir, a priori, avec cette proportionnalité. On a vu que cette coïncidence n'est pas fortuite, car elle relève de la même démarche d'application du principe du maximum de vraisemblance !

IV. APRÈS CETTE ENVOLÉE, UNE CONCLUSION

Les principales remarques faites sur les sujets analysés portent sur plusieurs aspects.

A) Les modèles pseudo-concrets choisis présentent des incohérences avec la réalité. Les rédacteurs sont obligés de les "tordre" (par exemple par des approximations osées) pour qu'ils aboutissent aux modèles mathématiques du programme, sans que cette démarche d'approximation soit d'ailleurs explicitement au programme ! C'est en particulier le cas de l'approximation d'un modèle de type hypergéométrique (tirages exhaustifs), par un modèle de type binomial (tirages avec remise), dans les sujets I, II, III.

Les modèles de type hypergéométrique en tant que tels ne sont pas au programme, mais le calcul direct de probabilités hypergéométriques en fait partie, et rentre dans les cordes des élèves de Terminale. La plupart des exercices faisant intervenir des tirages simultanés aboutissent à des calculs de probabilités hypergéométriques.

Cette dérive des énoncés pseudo-concrets présente plusieurs inconvénients.

Les élèves sont confinés dans un modèle unique et peuvent finir par croire qu'il n'existe que celui-là. Ils l'appliquent, par simple effet de contrat, certains diraient par automatisme, ou par défaut. Ils sont dispensés de la démarche raisonnée du choix d'un modèle, qui est de loin la partie la plus formatrice et la plus intéressante.

Le choix entre plusieurs modèles ou modes de calcul est rendu impossible par le manque de données : taille de la population, mode de tirage, précision sur le résultat final...

Il est incohérent de mettre en avant, dans le choix d'un modèle, la facilité des calculs, alors que la programmation des

**ATTENTION ! UN MODELE
PEUT EN CACHER UN AUTRE**

calculettes, qui est fortement recommandée dans les programmes, rend secondaire cet aspect des choses !

Il est incohérent de ne jamais pouvoir baser le choix d'un modèle sur la comparaison de la précision des résultats finaux, alors que c'est ce que l'on doit faire dans la pratique, à une époque où, par exemple, les normes de qualité sont de plus en plus exigeantes et demandent des calculs de plus en plus précis ! Il est incohérent d'imposer une précision numérique contradictoire avec la réalité du problème, alors que les programmes insistent sur *"la vérification de la pertinence des résultats en fonction du problème posé"* !

B) Dans le sujet III (la loterie) question 3.c., le modèle (multinomial) sous-jacent est hors programme, et le calcul direct des probabilités présente des difficultés techniques de dénombrement qui semblent pourtant proscrites des programmes : *"...On s'attachera à étudier des situations permettant de bien saisir la démarche du calcul des probabilités et non des exemples comportant des difficultés techniques de dénombrement..."*.

C) Quant au sujet IV, (la pêche miraculeuse !) les auteurs ont sans doute profité de la liberté que leur laissait un sujet qui n'est pas de Bac, pour partir en orbite sur des notions du programme de BTS, avec des descriptions approximatives de protocoles, des calculs compliqués et sans respecter ce qui constitue le fond du programme de probabilité de Terminale : évaluer la démarche sur des exemples simples.

D) Un dernier point à soulever est l'incohérence des attentes des sujets d'une même année, d'une année à l'autre, d'une section à une autre. En 1996 en S on refusait (consignes de correction) le calcul

détaillé par l'approximation binomiale, alors qu'il est attendu dans la majorité des autres sujets, ES 1995, STAE 1996 et 1997. En ES spécialité en 1995 il faut forcer l'approximation binomiale alors que dans la même section en 1996, pour le 3.c., même en lui tordant complètement le cou, l'approximation binomiale ne marche plus, puisqu'on a un modèle multinomial ! A ce rythme l'enseignement des probabilités va finir par devenir un vrai casse tête.

**AU SECOURS !!!
AUTEURS DE SUJETS D'EXERCICES
SUIVEZ LE CONTENU ET L'ESPRIT
DES PROGRAMMES
PAR RESPECT POUR LES ÉLÈVES ET
PAR PITIÉ POUR LEURS
ENSEIGNANTS**

Pourtant quelques propositions simples pourraient suffire à améliorer la qualité des exercices de probabilité. La première serait d'adopter une démarche rigoureuse et méthodique de modélisation : une description précise, un modèle unique, un ensemble des issues bien déterminé. La seconde de respecter les indications des programmes : problèmes sans difficultés techniques de dénombrement, favorisant la démarche, incluant une phase de contrôle de la pertinence des résultats et leur exploitation. La troisième serait d'enrichir le vocabulaire des programmes, en donnant des précisions sur les modèles et la modélisation – on pourrait parler de modèles de type binomial, de type hypergéométrique (ce qui n'implique pas de devoir traiter les lois correspondantes) – des précisions sur le contrôle et l'exploitation des résultats, des précisions sur l'utilisation des approximations d'un modèle par un autre. On peut rêver ?