Classe de Première ES

Programme

BO HS n°7 du 31 août 2000

Contenus	Modalités de mise en œuvre	Commentaires
Statistique Étude de séries de données: - nature des données (effectifs, données, moyennes, indices, pourcentages,…); - lissage par moyennes mobiles;	On s'intéressera en particulier aux séries chronologiques. On effectuera à l'aide d'un tableur le lissage par moyennes mobiles et on observera directement son effet sur la courbe représentant la série.	
- histogrammes à pas non constants; - diagrammes en boîte.	Les histogrammes à pas non constants ne seront pas développés pour eux mêmes, mais le regroupement en classes inégales s'imposera lors de l'étude d'exemples comme des pyramides des âges ou de salaires. On apprendra à interpréter diverses formes de diagrammes en boîtes à partir d'exemples. En liaison avec le paragraphe "probabilité", on étudiera plusieurs séries obtenues par simulation d'un modèle; on comparera les diagrammes en boîte. L'utilisation d'un logiciel informatique est indispensable pour accéder à une simulation sur un nombre important d'expériences.	Sans développer de technicité particulière à propos des histogrammes à pas non constants, on montrera l'intérêt d'une représentation pour laquelle l'aire est proportionnelle à l'effectif.
Effet de structure lors du calcul de moyennes. Mesures de dispersion: intervalle interquartile, écart-type.	On observera dynamiquement et en temps réel, les effets des modifications des données.	L'objectif est de résumer une série par un couple (mesure de tendance centrale; mesure de dispersion). Deux choix usuels sont couramment proposés: le couple (médiane ; intervalle interquartile), *robuste* par rapport aux valeurs extrêmes de la série, et le couple (moyenne ; écart-type). On démontrera que la moyenne est le réel qui minimise $\Sigma(xi - x)^2$ alors qu'elle ne minimise pas $\Sigma\|xi - x\|$. On notera s l'écart-type d'une série, plutôt que σ, réservé à l'écart-type d'une loi de probabilité.
Tableau à double entrée : étude fréquentielle lien entre arbre et tableau à double entrée; notion de fréquence de A sachant B.		La fréquence de A sachant B sera notée $f_B(A)$; elle prépare à la notion de probabilité conditionnelle qui sera traitée en terminale.

Contenus	Modalités de mise en œuvre	Commentaires
Probabilités Définition d'une loi de probabilité sur un ensemble fini. Probabilité d'un événement, de la réunion et de l'intersection d'événements. Modélisation d'expériences de référence menant à l'équiprobabilité; utilisation de modèles définis à partir de fréquences observées.	Le lien entre loi de probabilité et distribution de fréquences sera éclairé par un énoncé vulgarisé de la loi des grands nombres. On mènera de pair simulation et étude théorique de la somme de deux dés (en liaison avec le paragraphe précédent).	Un énoncé vulgarisé de la loi des grands nombres peut être par exemple: *Pour une expérience donnée, dans le modèle défini par une loi de probabilité P, les distributions des fréquences obtenues sur des séries de taille n se rapprochent de P quand n devient grand.* On indiquera que simuler une expérience consiste à simuler un modèle de cette expérience. On pourra ne pas se limiter à l'étude d'une seule situation et envisager d'autres expériences (produit de deux dés, somme de trois dés…). On pourra repérer les difficultés soulevées par le choix d'un modèle mais sans s'y attarder : on utilisera directement des modèles que la statistique a permis de choisir.

Document d'accompagnement programme de la classe de Première ES

Extrait : À propos du titre "Traitement des données et probabilités"

Le choix a été fait pour la section ES de donner un rôle important aux séries chronologiques, particulièrement fréquentes dans les cours d'économie de cette section ; mais on veillera à ce que les questions et exemples traités conduisent jusqu'à une réflexion conceptuelle ou axiomatisée, ce qui constitue une bonne préparation à d'éventuelles études ultérieures davantage centrées sur des pratiques professionnelles.

a) Pourcentages

Comme indiqué dans le programme, il ne s'agit pas d'aborder ici quelque connaissance technique nouvelle, mais d'entretenir un apprentissage de base indispensable pour lire correctement et de façon critique l'information chiffrée. Ce paragraphe recoupe de nombreux autres titres du programme : statistique (fréquences, données en pourcentage, évolution de séries chronologiques,…), suites géométriques (obtenues par augmentations successives), dérivation (approximation affine), … Les notions qu'il décrit devront donc être régulièrement mises en jeu, en particulier à partir de données issues des médias ou de l'environnement scolaire de la classe.

b) Nature des données

On a parfois prôné, pour l'enseignement de la statistique, le recueil de données par les élèves eux-mêmes, cette pratique étant considérée comme motivante et permettant de percevoir le champ de l'aléatoire. Or la perception de l'aléatoire peut s'acquérir de manière plus profonde par la simulation. Par ailleurs, pour de nombreuses questions que l'on peut se poser dans le champ scolaire, des données existent : elles sont réactualisées chaque année, leur contenu est riche et elles sont accessibles dans des banques de données. Sans exclure complètement un recueil ponctuel de données par les élèves à condition qu'il ne prenne pas beaucoup de temps, on s'appuiera avant tout sur des données existantes. Certaines données sont des données brutes (exemple : série des hauteurs d'un fleuve mesurées en un point géographique précis tous les jours à la même heure) ; d'autres sont obtenues en prenant des moyennes de mesures brutes (séries des températures mensuelles en un point géographique précis, évolution sur une période de 10 ans des dépenses de logement suivant les catégories socioprofessionnelles) ; certaines encore sont des moyennes mobiles : par exemple, dans la présentation des mesures de température quotidienne à une heure donnée en un point donné, on remplace parfois les températures brutes $x_0,\ldots x_n$ par la série des moyennes mobiles d'ordre k, $y_k,\ldots y_{n-k}$, calculées ainsi : $y_i=\frac{1}{2k+1}\sum_{i-k}^{i+k} x_j$, pour i=k,…,n-k.

On pourra montrer sur des exemples, avec k=2 ou 3, que la courbe obtenue en remplaçant les données brutes par les moyennes mobiles est plus lisse.

c) Effets de structure

Exemple

Le revenu moyen global des individus actifs d'une population (par exemple la population parisienne) peut augmenter avec le temps alors que dans toutes les catégories socioprofessionnelles (CSP) le revenu baisse, l'augmentation globale étant liée à un changement de la répartition en CSP (à Paris, les CSP à faible revenu ont eu tendance à déménager en banlieue). La structure à une date donnée est ici la répartition des CSP à cette date.

d) Diagrammes en boîtes

Il ne s'agit pas là d'un élément à part ; il sera introduit à l'occasion du traitement de données expérimentales ou d'activités de simulation. On trouvera en annexe de ce document une note sur ce type de diagrammes à l'usage des enseignants.

e) Étude fréquentielle de tableaux à double entrée

Les commentaires sur les pourcentages des lignes (resp. des colonnes) se feront simplement à partir des distributions de fréquences associées aux marges horizontales (resp. verticales). On ne construira pas les « tableaux théoriques » (on n'introduira pas de façon formelle la notion de sur- et sous-représentation, celles-ci n'ayant vraiment de sens que si elles sont « significatives » au sens statistique).

Exemple

En 1979, le New-York Times a noté qu'entre 1973 et 1979, en Floride, il a été prononcé 131 peines capitales pour meurtre ; parmi ces condamnés, 55% étaient des Blancs alors que 48 % de ceux qui ont été jugés pour meurtre pendant la même période étaient des Blancs. Qu'en penser ?

Avant d'en penser quoique ce soit, il est opportun de s'intéresser aussi à la couleur de la victime.

1- Entre 1973 et 1979, il y a eu 2433 meurtres en Floride, dont la victime était de couleur blanche. Le tableau ci-dessous donne les détails suivant la couleur de la peau du suspect (blanche, B, ou non blanche, NB), les colonnes indiquant la sentence (PC : peine capitale ; AS : autre sentence)

	PC	AS	Totaux
NB	48	239	287
B	72	2074	2146
Totaux	120	2313	2433

- Dire comment est construit le tableau ci-dessous et ce que signifient les trois lignes :

	PC	AS	Totaux
NB	16,7	83,3	100,0
B	3,4	96,6	100,0
Totaux	4,9	95,1	100,0

Remarque : les différences de pourcentages de condamnation à la peine capitale suivant la couleur du suspect, lorsque la victime est de couleur blanche, est trop grande pour être imputée à la fluctuation d'échantillonnage.

2- Entre 1973 et 1979, il y a eu 4764 meurtres en Floride. Le tableau ci-dessous donne les détails suivant la couleur de peau du suspect (blanche, B, ou non blanche, NB), les colonnes indiquant la sentence (PC : peine capitale ; AS : autre sentence).

	PC	AS	Totaux
NB	59	2448	2507
B	72	2185	2257
Totaux	131	4633	4764

- Construire un tableau analogue au deuxième tableau de la question 1 et le commenter.

- Quel est le pourcentage de suspects de couleur blanche lorsque la victime est de couleur blanche ; quel est le pourcentage de suspects de couleur blanche lorsque la victime n'est pas de couleur blanche ?

- Faire l'étude lorsque la victime n'est pas de couleur blanche.

Remarque : les différences de pourcentages de condamnation à la peine capitale suivant la couleur du suspect, sans tenir compte de la couleur de la victime, ou lorsque la victime est non blanche, peuvent être imputées à la fluctuation d'échantillonnage, i.e. ne peuvent pas être interprétées comme une tendance des jugements en Floride.

Exemple

Que signifient les tableaux ci-dessous et comment sont-ils déduits du premier tableau ?

Les données concernent la répartition suivant le sexe et le poste (AG : agent spécial , P : personnel autre) des emplois au FBI au 31 janvier 1997.

	H	F	
AS	9199	1617	10816
P	4535	10165	14700
	13734	11782	25516

	H	F	
AS	36,1	6,3	42,4
P	17,8	39,8	57,6
	53,8	46,2	100,0

	H	F	
AS	85,0	15,0	100,0
P	30,9	69,1	100,0
	53,8	46,2	100,0

	H	F	
AS	67,0	13,7	42,4
P	33,0	86,3	57,6
	100,0	100,0	100,0

Remarque

On distinguera les calculs de l'interprétation, dans le contexte, des données d'observation.

f) Loi de probabilité

On recensera les propriétés mathématiques élémentaires de l'objet « distributions de fréquences » (cf. tableau ci-dessous) et on définira une loi de probabilité comme un objet mathématique ayant les mêmes propriétés.

Distribution de fréquences sur $E=\{x_1,\dots,x_r\}$	Loi de probabilité sur $E=\{x_1,\dots,x_r\}$
$(f_1,\dots,f_r)$ $f_i \geq 0$; $\sum f_i = 1$ $A\subset E$ Fréquence de A : $f(A)= \sum f_i$ Événement complémentaire : $f(\overline{A})=1-f(A)$ *Cas numérique :* *Moyenne empirique :* $\overline{x} = \sum f_i x_i$	$(p_1,\dots, p_r)$ $p_i \geq 0$; $\sum p_i = 1$ $A\subset E$ probabilité de A : $P(A)= \sum p_i$ Événement complémentaire : $P(\overline{A})=1-P(A)$ *Cas numérique :* *Espérance d'une loi P :* $\mu=\sum p_i x_i$

Modéliser une expérience aléatoire, c'est lui associer une loi de probabilité (qui est un objet du monde mathématique).

Une fréquence est empirique : elle est calculée à partir de données expérimentales, alors que la probabilité d'un événement est un « nombre théorique » (un objet du monde mathématique). Les distributions de fréquences issues de la répétition d'expériences identiques et indépendantes varient (fluctuent), la loi de probabilité est un invariant associé à l'expérience.

L'objectif est que les élèves comprennent à l'aide d'exemples (cf. paragraphe sur les expériences de référence) que modéliser, c'est ici choisir une loi de probabilité. Il ne s'agit cependant en aucun cas de tenir des discours généraux sur les modèles et la modélisation.

Les élèves devront bien distinguer ce qui est empirique (du domaine de l'expérience) de ce qui est théorique ; en particulier, on réservera la lettre grecque σ à l'écart-type d'une loi et on évitera de noter avec cette même lettre un écart-type empirique (il s'agit là de règles de notation internationales) .

En classe de première, une loi de probabilité P sur un ensemble fini est la liste des probabilités des éléments de E ; à partir de cette liste, on définit naturellement la probabilité d'événements, c'est-à-dire implicitement une application de $\mathcal{P}(E)$ dans $[0,1]$, application qui sera encore désignée par P .

Il est inutilement complexe, pour le cas des ensembles finis, de partir d'une application de $\mathcal{P}(E)$ dans $[0,1]$, vérifiant certains axiomes, puis de montrer ensuite que cette application est entièrement caractérisée par $(p_1,\dots, p_r)$. Le fait de ne pouvoir simplement généraliser cette définition aux ensembles continus, et la nécessité d'une définition ensembliste sera abordée en terminale.

Si tous les éléments d'un ensemble bien défini E ont même probabilité, celle-ci est dite équirépartie ; on parlera aussi d'équiprobabilité et on dira que les éléments de l'ensemble E sont choisis au hasard (i.e. si on ne spécifie rien de plus, le choix au hasard est un choix avec équiprobabilité). Si la loi P n'est pas équirépartie, on parlera de choix d'éléments selon la loi P (ou éventuellement de choix au hasard selon la loi P).

Dans le cas de choix au hasard, la probabilité d'un événement est le quotient de son nombre d'éléments par le nombre d'éléments de l'ensemble.

On évitera tout développement théorique sur le langage des événements et le calcul ensembliste qui en découle : ces notions et la pratique de la logique qu'ils impliquent (étude du complémentaire de l'événement "A ou B", ou de l'événement "A et B") s'acquièrent au fil d'exercices.

g) Modélisation d'expériences de référence

Modéliser une expérience aléatoire à valeurs dans un espace Ω, c'est choisir une loi de probabilité P définie sur Ω. Ce choix est en général délicat à faire, sauf dans certains cas où des considérations propres au protocole expérimental conduisent à proposer a priori un modèle. Il en est ainsi des lancers de pièces ou de dés pour lesquels des considérations de symétrie conduisent au choix d'un modèle où la loi de probabilité est équirépartie.

Sans faire une liste de conventions terminologiques, on indiquera clairement que les termes *équilibré* et *hasard* indiquent un choix du modèle de l'expérience, modèle où intervient quelque part une probabilité équirépartie.

h) Modélisation à partir de fréquences.

En dehors des cas évoqués ci-dessus où des considérations quant à la nature des expériences permettent de proposer un modèle, le choix d'un modèle à partir de données expérimentales est beaucoup plus délicat. On se contentera, pour certains exercices, de fournir un modèle en indiquant dans un premier temps que des techniques statistiques ont permis de déterminer et de valider un tel modèle à partir de nombreuses données expérimentales.

Pour déterminer et/ou valider un modèle probabiliste, le premier outil dont on dispose est un théorème de mathématiques appelé loi des grands nombres, dont un énoncé intuitif est :

Dans le monde théorique défini par une loi P sur un ensemble Ω, les fréquences des éléments de Ω dans une suite de n expériences identiques et indépendantes tendent vers leur probabilité quand n augmente indéfiniment.

On donnera des exemples où un modèle est déterminé à partir de fréquences ; les exemples les plus compliqués que l'on abordera consistent à associer une loi de probabilité à un tableau à double entrée.

i) Simulation

L'exemple ci-dessous montre comment mêler diverses composantes d'un travail mathématique : observation, premières conjectures, expérimentation à plus grande échelle, puis obtention et preuve de certains résultats :

Exemple

Somme de deux dés.

En répétant 100 fois de suite le lancer de deux dés, on **observe** que certains résultats s'obtiennent plus souvent que d'autres.

Á l'aide d'un tableur par exemple, il est possible d'**expérimenter** à plus grande échelle : simulation d'un plus grand nombre de lancers de deux dés et construction du tableau des effectifs : l'inégale répartition des fréquences de chaque résultat est flagrante

La recherche d'un modèle théorique adapté avec une loi de probabilité équirépartie permet ensuite calculs et démonstrations : on **prouve** que les résultats sont inégalement probables et on détermine précisément leur probabilité.

Résultats pour 5 séries de 100 lancers

2	0,02	0,02	0,01	0,03	0,01
3	0,03	0,02	0,09	0,06	0,04
4	0,08	0,08	0,08	0,08	0,1
5	0,11	0,11	0,12	0,09	0,11
6	0,17	0,13	0,1	0,11	0,11
7	0,15	0,15	0,12	0,14	0,2
8	0,12	0,2	0,1	0,13	0,22
9	0,15	0,1	0,12	0,13	0,06
10	0,09	0,1	0,14	0,1	0,08
11	0,06	0,05	0,06	0,09	0,06
12	0,02	0,04	0,06	0,04	0,01