Classe de Première S

Programme

BO HS n°7 du 31 août 2000

La partie du programme consacrée aux probabilités et à la statistique est centrée :

- sur la mise en place d'éléments de base indispensables pour comprendre ou pratiquer la statistique partout où elle est présente,

- sur l'acquisition de concepts de probabilité permettant de comprendre et d'expliquer certains faits simples observés expérimentalement ou par simulation.

Le programme de la classe de première introduit quelques outils descriptifs nouveaux :

¤ les diagrammes en boîtes qui permettent d'appréhender aisément certaines caractéristiques des répartitions des caractères étudiés et qui complètent la panoplie des outils graphiques les plus classiquement utilisés ;

¤ deux mesures de dispersion : l'écart-type et l'intervalle interquartile.

Ces éléments de statistique pourront notamment être travaillés pour des séries construites à partir de séries simulées; on rencontre ainsi des répartitions variées et on prépare la notion d'estimateur. Cette partie descriptive ne doit pas faire l'objet de longs développements numériques, ni être déconnectée du reste du programme de probabilité et statistique.

Contenus	Capacités attendues	Commentaires
Statistique Variance et écart-type. Diagramme en boîte; intervalle interquartile. Influence sur l'écart type et l'intervalle interquartile d'une transformation affine des données.	On cherchera des résumés pertinents et on commentera les diagrammes en boîtes de quantités numériques associées à des séries simulées ou non. On observera l'influence des valeurs extrêmes d'une série sur l'écart type ainsi que la fluctuation de l'écart type entre séries de même taille. L'usage d'un tableur ou d'une calculatrice permettent d'observer 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.
Probabilités Définition d'une loi de probabilité sur un ensemble fini. Espérance, variance, écart-type d'une loi de probabilité. Probabilité d'un événement, de la réunion et de l'intersection d'événements. Cas de l'équiprobabilité. Variable aléatoire, loi d'une variable aléatoire, espérance, variance, écart type. Modélisation d'expériences aléatoires de référence (lancers d'un ou plusieurs dés ou pièces discernables ou non, tirage au hasard dans une urne, choix de chiffres au hasard, etc.).	Le lien entre loi de probabilité et distributions de fréquences sera éclairé par un énoncé vulgarisé de la loi des grands nombres. On expliquera ainsi la convergence des moyennes vers l'espérance et des variances empiriques vers les variances théoriques; on illustrera ceci par des simulations dans des cas simples. On pourra aussi illustrer cette loi avec les diagrammes en boîtes obtenus en simulant par exemple 100 sondages de taille n, pour n = 10; 100; 1000. On simulera des lois de probabilités simples obtenues comme images d'une loi équirépartie par une variable aléatoire (sondage, somme des faces de deux dés, etc.).	On pourra par exemple choisir comme énoncé vulgarisé de la loi des grands nombres la proposition suivante : *Pour une expérience donnée, dans le modèle défini par une loi de probabilité P, les distributions des fréquences calculées 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. La modélisation avec des lois ne découlant pas d'une loi équirépartie est hors programme. On évitera le calcul systématique et sans but précis de l'espérance et de la variance de lois de probabilité.

Annexe 1 : statistique et probabilités

Le programme de *probabilité-statistique* de la section S introduit les notions de loi de probabilité et de variable aléatoire, notions indispensables pour comprendre l'esprit de la statistique et aborder la problématique propre de cette discipline ; quelques éléments de statistique descriptive sont introduits, mais la statistique descriptive a une part modeste dans cette section.

La terminologie en usage pour la statistique et les probabilités sera réduite au minimum ; on s'efforcera de garder le langage ensembliste et de ne pas développer deux terminologies parallèles : on introduira seulement le terme événement ; l'ensemble des éventualités liées à une expérience gardera le nom d'ensemble, on parlera d'événements complémentaires plutôt que d'événements contraires ou incompatibles ; on pourra , pour éclairer certains exemples mais sans systématiser ou rendre obligatoire un tel langage, parler d'univers, d'événements contraires ou incompatibles ; l'important est de savoir résoudre des problèmes, de relier les résultats de calculs faits sur des observations aux questions posées ; on trouvera en fin de cette annexe un lexique des principaux termes liés à la statistique et que l'élève doit savoir manipuler.

Statistique descriptive

Schématiquement, la statistique se pratique dans deux contextes :

- pour apporter des éléments de réponse à une question ; cela nécessite le plus souvent de reformuler la question puis de définir un protocole de recueil des données suivies d'une analyse descriptive des données recueillies et une modélisation probabiliste (exemple : les alliages produits par un certain procédé industriel sont-ils conformes aux normes ? téléphoner en voiture est-il un facteur de risque d'accident ? quelle fraction de la population doit-on vacciner pour enrayer une épidémie rapidement ? etc.).

- pour observer ou surveiller régulièrement une population (statistiques de natalité, de mortalité, de la répartition des dépenses des ménages suivant certains secteurs - alimentation, loisir, logement, dépenses de santé, etc- surveillance des taux de chômage, répartition des cultures céréalières dans chaque région, réussite aux bac par filières, etc.).

Dans tous les cas, il y a des données expérimentales et il convient de savoir :

- comment elles sont définies,

- dans quel cadre elles sont recueillies, quelles sont les sources d'erreurs et de variabilité possibles.

Les problèmes afférents au type de données traitées et au mode de recueil dépendent des champs disciplinaires concernés (économie, sociologie, enquêtes de marketing, médecine, finances). Nous avons choisi de ne pas traiter ce qui relève d'une formation professionnelle technique et qui est abordé avec plus d'efficacité dans des études supérieures spécialisées ; certains points pourront par ailleurs être abordés en physique, en biologie, ou en géographie. En première et terminale, on travaillera essentiellement des questions conduisant à une réflexion conceptuelle ou axiomatisée, i.e des questions nécessitant le recours à des calculs probabilistes et on continuera les activités de simulation ; il s'agit là d'éléments qui sont au cœur de tous les champs d'application considérés. Les TPE permettront à certains de faire un réel travail de statistique.

Quelques éléments de statistique descriptive sont présentés pour les séries numériques : écart-type et diagramme en boîte. Il est important ici de démontrer que la moyenne minimise la fonction $x \mapsto \sum_i (x_i - x)^2$ qui mesure la dispersion autour d'un point, et que ce minimum est appelé la variance. Les diagrammes en boîte permettent une comparaison graphique de plusieurs séries de données (voir fiche « sondages » du document d'accompagnement de seconde) ; on remarquera à l'aide d'exemples que le résumé d'une série par le couple (médiane, écart interquartile) est insensible aux valeurs extrêmes, mais ne peut faire l'objet de calculs par paquets ; résumer une série de données par (moyenne, écart-type) a des propriétés théoriques qui rendent son usage fréquent, notamment en biologie et physique pour ce qui touche aux mesures expérimentales. On regardera comment se transforment les quantités introduites si on change les unités et/ou qu'on décale l'origine ; les résumés numériques pouvant être déterminés à partir de la distribution des fréquences sont aussi affectés par la fluctuation d'échantillonnage : on observera sur des simulations que la variance d'une moyenne est en $n^{-1/2}$, d'où l'intérêt de prendre une moyenne. On notera que la moyenne d'une série de données étant l'isobarycentre des termes et le barycentre des valeurs prises pondérées par les fréquences, elle bénéficie de la propriété d'associativité des barycentres et permet donc les calculs par sous-groupes. La variance permet aussi les calculs par sous groupes ; la formule de décomposition ci-dessous de la variance n'est pas au programme et figure ici à l'usage des enseignants :

Soit une série de n données, de moyenne m et de variance s^2, décomposée en k sous-séries de tailles respectives n_i, i=1..k. Soit m_i et s^2_i les moyennes et les variances dans chaque sous-série ; alors s^2 est la somme de la moyenne des variances s_i pondérées par les n_i /n et de la variance de la série des m_i affectés des poids n_i /n :

$$s^2 = \sum \frac{n_i}{n}(m_i - m)^2 + \sum \frac{n_i}{n}\ s_i^2$$

Pour de nombreuses questions abordables dans le champ scolaire, des données existent : elles sont réactualisées régulièrement, leur contenu est riche et elles sont accessibles dans des banques de données : on pourra les utiliser. Il est préférable d'aborder un seul exemple motivant, plutôt que des petits exemples dont le traitement se réduit à des calculs numériques sans objet. On pourra aussi traiter de grandes séries de données en fournissant aux élèves des résultats numériques complets ou partiels (par exemple somme des termes et somme des carrés des termes) ; l'important ici n'est pas de faire des calculs soi-même ou d'être virtuose de l'emploi des touches statistiques d'une calculatrice -celles ci ne sont en pratique utilisables que pour des séries de petites tailles- ; quand l'équipement est disponible, le tableur constitue un outil très approprié de traitement des données ; l'objectif reste toujours de comprendre les calculs, de savoir les interpréter.

On a souvent prôné, pour l'enseignement de la statistique, le recueil de données par les élèves eux-mêmes : ce recueil était considéré comme motivant et permettant de percevoir le champ de l'aléatoire. Mais la perception de l'aléatoire peut aussi s'acquérir aujourd'hui par la simulation : les paramètres de dispersion introduits pourront être calculés sur des séries simulées. Un recueil effectif de données par les élèves n'est donc à envisager que s'il ne prend pas beaucoup de temps et traite d'une question que les élèves ont fortement contribué à formuler.

Probabilité, modélisation, simulation

Le programme de probabilité des classes de première et terminale S concerne la modélisation d'expériences de référence, modélisation définie par la loi de probabilité équirépartie sur un ensemble fini convenablement choisi. Les lois de probabilité non équiréparties rencontrées en première et terminale sont le plus souvent l'image par une variable aléatoire d'une loi équirépartie, et on introduira donc presque simultanément la notion de loi de probabilité et celle de variable aléatoire.

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 $E=\{x_1,\ldots,x_r\}$	Loi de probabilité sur $E=\{x_1,\ldots,x_r\}$
$(f_1,\ldots,f_r)$ $f_i \geq 0$; $\sum f_i = 1$	$(p_1,\ldots, p_r)$ $p_i \geq 0$; $\sum p_i = 1$
$A \subset E$ Fréquence de A : $f(A) = \sum f_i$ Événement complémentaire : $f(\bar{A}) = 1 - f(A)$	$A \subset E$ probabilité de A : $P(A) = \sum p_i$ Événement complémentaire : $P(\bar{A}) = 1 - P(A)$
Cas numérique : Moyenne empirique : $\bar{x} = \sum f_i x_i$ Variance empirique : $s^2 = \sum f_i (x_i - \bar{x})^2$ Ecart-type empirique : $s = \sqrt{\sum f_i (x_i - \bar{x})^2}$	Cas numérique : Espérance d'une loi P : $\mu = \sum p_i x_i$ Variance d'une loi P : $\sigma^2 = \sum p_i (x_i - \mu)^2$ Ecart-type d'une loi P : $\sigma = \sqrt{\sum p_i (x_i - \mu)^2}$

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 » (on 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 en aucun cas d'avoir 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 notations 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 $(p_1,\ldots, p_r)$. Le fait de ne pouvoir simplement généraliser aux ensembles continus cette définition, et la nécessité d'une définition ensembliste seront abordés 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.

Variable aléatoire

Exemples d'activités préparatoires

- On s'intéresse au jeu suivant : on lance deux dés distincts et on gagne 1F si la somme des résultats est paire, on perd 1F sinon.

L'ensemble E des issues possibles est celui des couples (i,j) de nombres entiers entre 1 et 6. On considère l'application gain, notée G, qui à (i,j) fait correspondre 1 si i+j est pair, -1 sinon. On associe à une distribution de fréquences observée lors de n parties la distribution des fréquences des valeurs prises par G : c'est une distribution de fréquences sur {-1,1}.

- n lancers de deux dés discernables (n>30) .

On considère l'ensemble E des couples (i,j) de nombres entiers entre 1 et 6. Soit l'application T qui à (i,j) fait correspondre i+j. On associe, par le calcul, à une distribution de fréquences sur E obtenue en lançant n fois deux dés la distribution des fréquences des valeurs prises par T : c'est une distribution de fréquences sur {2, …,12}.

- n lancers de 5 pièces discernables (n>30) .

L'ensemble E considéré ici sera celui des listes (x_1, ...,x_5) de nombres 0 ou 1.

On considère l'application T qui à (x_1, ...,x_5) fait correspondre x_1+ ...+x_5. On associe à une distribution de fréquences sur E, obtenue en lançant n fois cinq pièces, la distribution des fréquences de T : c'est une distribution de fréquences sur {0,...,5}

Une application T de E dans E' (E et E' finis) permet de transporter la loi P définie sur E en une loi P' définie sur E' par le procédé suivant :

$$P'(x') = P(T=x')$$

où P(T=x') désigne la probabilité de l'ensemble des éléments de E dont l'image par T est x' (cet ensemble est souvent noté {T=x'}, on écrit en fait P(T=x') au lieu de P({T=x'})).

On notera sur des exemples que même si la loi P est équirépartie, la loi transportée P' par T, appelée *loi image de T*, n'est en général pas l'équiprobabilité (cf. activités préparatoires).

Dans ce cadre-là, ce qui nous intéresse dans l'application T, c'est ce qui se passe à l'arrivée : partant d'un élément x choisi dans E selon la loi P, on arrive à un élément x' choisi dans E' selon la loi P'. Le terme usuel pour nommer T est de parler d'une variable aléatoire.

Dans le cas numérique, l'espérance de la loi de P' est appelée plus simplement espérance de T et notée E(T).

Remarque :Les lois de probabilités manipulées dans les exercices pourront bien sûr aussi être définies sur des ensembles discrets non ordonnés et les variables aléatoires ne sont pas nécessairement numériques.

Exemples

Que gagne-t-on en moyenne si on lance deux dés distincts en gagnant 1F si la somme des résultats est paire, et en perdant 1F sinon.?

Le modèle associé est l'équiprobabilité sur l'ensemble des couples (i,j) de nombres entiers entre 1 et 6. On considère la variable aléatoire gain G qui à (i,j) fait correspondre 1 si i+j est pair, -1 sinon. On peut alors déterminer la loi de G et son espérance.

Il est important, à propos du concept de variable aléatoire, de comprendre qu'une fois fait le choix d'un modèle pour un expérience, d'autres lois de probabilité en découlent par le calcul probabiliste: on ne fait qu'une seule fois un choix de modèle pour traiter un problème.

Exercices

- Lancers de deux dés discernables

E= $\{1,\ldots,6\}^2$ et P est la probabilité équirépartie. Un tel choix de modèle pourra être confronté à l'expérience elle-même. On considère l'application T qui à (i,j) fait correspondre i+j. Calcul et simulation de la loi de T.

On considère l'application U de E dans {0,1} qui à (i,j) fait correspondre 0 si i et j ne sont pas premiers entre eux. Calcul et simulation de la loi de U.

- Lancers de 2 pièces de couleurs distinctes dont les faces sont codées 0 et 1.

On considère l'ensemble $\{0,1\}^2$ et la probabilité équirépartie.

Un tel choix de modèle pourra être confronté à l'expérience elle-même dans le cas où les élèves ne l'ont pas déjà fait en seconde. On considère l'application T qui à (i,j) fait correspondre i+j. Calculer et simuler sa loi.

Supposons maintenant qu'un observateur ne voit pas les couleurs des pièces :

Pour lui, les issues observables de cette expérience sont « double 0 », « double 1 », « 1 et 0 ». Pour justifier le choix de la probabilité (0.25,0.25,0.5) pour ces trois issues, on pourra, outre la validation expérimentale en lançant des pièces, se ramener au cas précédent. Le résultat final ne doit pas dépendre du fait que l'observateur voit ou non les couleurs des pièces. On voit ainsi que pour certaines expériences, on est amené à choisir un modèle calculé à partir d'un modèle d'équiprobabilité sur un ensemble qui n'est pas celui des issues observables.

- Lancers de deux dés indiscernables .

Imaginons qu'on ne puisse observer que la somme des faces. L'ensemble des issues observables est donc {2,...,12}. On cherche à modéliser cette expérience. On pourra se faire une idée du résultat par expérimentation ou simulation et déterminer le modèle associé à cette expérience en partant d'un modèle équiréparti sur $\{1,\ldots,6\}^2$ et en introduisant la variable aléatoire somme.

- Urne dont les boules sont numérotées et de couleurs différentes

Lorsqu'on choisit au hasard une boule et qu'on s'intéresse à la loi de probabilité de la variable aléatoire couleur de la boule, la probabilité d'une couleur est égale à la proportion de boules de cette couleur dans l'urne.

- Urne à boules indiscernables

Le plus souvent la seule distinction entre les boules porte sur leur couleur et donc le seul résultat observable est une couleur ; mais la loi sur l'ensemble des couleurs ne doit pas dépendre de notre capacité à distinguer les boules individuellement. On conviendra implicitement (ou en l'explicitant si besoin est)que choisir au hasard des boules coloriées, c'est choisir une couleur selon une loi telle que la probabilité de chaque couleur bleue est égale à la proportion de boules de cette couleur dans l'urne (dans cette phrase le hasard signifie que la probabilité de chaque boule d'être tirée est la même).

Dans un casino, on peut jouer aux jeux suivants ; quel jeu choisir ?

Jeu A : On choisit un chiffre selon une loi de probabilité telle que la probabilité du chiffre i est proportionnelle à i+1. Le gain associé au chiffre i est 2520/(i+1).

Jeu B : On choisit un chiffre selon une loi de probabilité telle que la probabilité du chiffre i est proportionnelle à 1/(i+1). Le gain associé au chiffre i est 1000(i+1).

Modélisation d’expériences de référence

Modéliser une expérience aléatoire, c'est associer à cette expérience une loi de probabilité sur l’ensemble des issues possibles. Ce choix, c'est à dire la modélisation de l'expérience, 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. On se restreindra donc aux expériences de référence en évitant tout discours général sur ce qu’est ou n’est pas la modélisation.

La traduction mathématique de « une pièce a autant de chances de tomber sur pile que sur face » est ainsi « la probabilité de pile et de face sont égales ». On indiquera clairement que les termes *équilibré* et *choix au hasard* indiquent par convention un choix du modèle de l'expérience, où la probabilité est équirépartie.

En dehors de tels cas où des considérations quant à la nature des expériences permet de proposer un modèle, le choix d’un modèle à partir de données expérimentales est beaucoup plus délicat et ne sera pas abordé dans l'enseignement secondaire. On se contentera, si nécessaire, de fournir un modèle en indiquant que des techniques statistiques ont permis de déterminer et de valider un tel modèle.

La modélisation ne relève pas d’une logique du vrai et du faux : un modèle n'est ni vrai ni faux : il peut être validé ou rejeté au vu de données expérimentales. Une des premières fonctions de la statistique dite inférentielle est d'associer à une expérience aléatoire un modèle, ou une gamme de modèles compatibles en un certain sens à définir avec les données expérimentales dont on dispose, et aussi de définir des procédures de validation d’un modèle.

Pour valider un modèle probabiliste, le premier outil dont on dispose est un théorème de mathématiques appelé loi (forte) des grands nombres, dont un énoncé intuitif est :

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

Ou encore

Si on choisit n éléments selon une loi de probabilité P, indépendamment les uns des autres, alors la distribution des fréquences « tend » vers P lorsque n tend vers l'infini.

Il s’agit là d’un énoncé vulgarisé. Pour être un peu plus précis, le théorème appelé loi forte des grands nombres dit que dans l’ensemble des suites infinies d’éléments choisis selon P, le sous-ensemble des suites pour lesquelles la distribution des fréquences ne converge pas vers P est « négligeable » . Par exemple, dans une suite finie de n lancers d’une pièce équilibrée, on peut n’obtenir que des faces (code 0) ou que des pile (code 1), ou 001001…001, etc. ; ces trois suites finies ont chacune une probabilité 2^{-n} ; si on « prolonge » ces suites, les fréquences de 1 et de 0 ne tendent pas vers 1/2, mais la probabilité de l’ensemble de ces trois suites tend vers 0. Le théorème ci-dessus indique que l’ensemble de toutes les suites imaginables pour lesquelles les fréquences ne tendent pas vers 0,5 est de « probabilité nulle » - pour une loi de probabilité construite sur l’ensemble des suites infinies de 0 et de 1 à partir de l’équiprobabilité sur {0,1} . Le mathématicien dira qu’il y a convergence presque sûre des fréquences des éléments vers leur probabilité. Le statisticien, s’il observe dans une longue série d’expériences des distributions de fréquences qui fluctuent de moins en moins choisira comme modèle, en vertu de cette loi des grands nombres, une loi de probabilité « proche » de la dernière distribution observée : il ne va pas choisir un modèle pour lequel ce qu’il a observé constitue un événement quasi-négligeable !

Une validation du modèle de la loi équirépartie pour le lancer d'un dé consistera à vérifier que la distribution des fréquences est « proche » de (1/6,…,1/6) sur {1,...,6} quand le nombre de lancers est grand (cf. l’annexe du projet de terminale S).

Exercice

1- Dans un jeu de pile ou face où on gagne le double de la mise sur pile et où on perd la mise sur face, un joueur qui dispose de 1000 F commence par miser un franc, double sa mise tant qu'il perd et ne s'arrête que s'il gagne où s'il ne peut plus miser. Simuler ce jeu, puis calculer l'espérance de gain du joueur (on admettra par généralisation que si on joue n fois le modèle est l'équirépartition sur les n-listes dont les éléments valent 0 ou 1).

2- Dans un jeu de pile ou face où on gagne le double de la mise sur pile et où on perd la mise sur face, un joueur qui dispose de 1000 F commence par miser un franc, triple sa mise tant qu'il perd et ne s'arrête que s'il gagne où s'il ne peut plus miser. Simuler ce jeu, puis calculer l'espérance de gain du joueur.

Simulation de chiffres au hasard

On clarifiera brièvement les positions respectives de la modélisation et de la simulation : modéliser consiste à associer un modèle à des données expérimentales, alors que simuler consiste à produire des données à partir d'un modèle prédéfini. On parlera de simulation d'une loi de probabilité P ; la simulation d'une telle loi avec des listes de chiffres au hasard ne peut se faire que si P peut être construite comme loi image d'une loi équirépartie. Pour simuler une expérience , on associe d'abord un modèle à l'expérience en cours, puis on simule la loi du modèle ; on pourra détailler ces étapes, sans cependant le faire systématiquement dans les cas simples des expériences de référence .

Exemple d'activités utilisant des simulations

1- On considère deux stratégies de choix de nombres parmi les nombres 1,3,5.

(i)- Choix au hasard

(ii)- Choix selon la loi (0.1, 0.1,0.8)

Calculer l'espérance des deux lois ci-dessus. Simuler n choix selon l'une des deux méthodes ; regarder si au vu de la moyenne des séries observées, quelqu'un ne connaissant pas le modèle choisi pour simuler peut le deviner. On fera varier n.

2- Pour la grille suivante, on considère trois procédures de choix de cases blanches :

<table>
<tr>
<td>
<table>
<tr><td></td><td>N</td><td>N</td></tr>
<tr><td></td><td>N</td><td></td></tr>
<tr><td>N</td><td></td><td></td></tr>
</table>
</td>
<td>
(i)- On choisit une case blanche au hasard parmi les 5 cases blanches

(ii)- On choisit au hasard une ligne puis dans la ligne une case au hasard parmi les cases blanches de cette ligne

(iii)- On choisit au hasard une colonne puis dans la colonne une case au hasard parmi les cases blanches de cette colonne
</td>
</tr>
</table>

Déterminer dans chacun des trois cas la loi de probabilité mise en jeu sur l'ensemble des cases blanches.

Pour une des trois lois de probabilité, un élève simuler n choix de cases blanches ; au vu de la distribution des fréquences obtenu, un autre élève qui ignorerait quelle loi a été simulé peut-il la deviner ?

Dans les deux exemples ci-dessus, on insistera sur le fait que l'observation de résultats simulés ne permet pas de remonter au modèle à coup sûr : une des fonctions de la statistique est de calculer le probabilité que l'on a de se tromper en « remontant d'une distribution de fréquence à une loi de probabilité ».

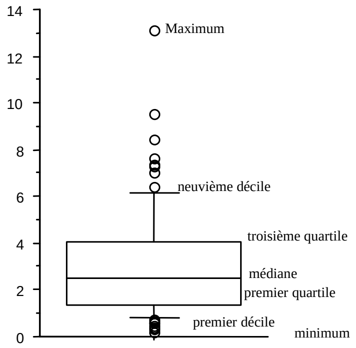
Exercice

Une expérience consiste à choisir successivement deux chiffres d et u au hasard. On considère l'application T qui à (d,u) fait correspondre 10d+u . Calculer la probabilité que 10d+u $\leq$ k, où k est un entier entre 0 et 9. On justifie ainsi que pour construire une liste de nombres au hasard entre 0 et 99, on lise les chiffres d'une table de chiffres au hasard deux par deux.

Lexique

Choix au hasard dans un ensemble E : la loi de probabilité en jeu sur E est l'équiprobabilité.

Diagramme en boîte (ou à pattes ou à moustaches ou diagramme de Tuckey) : on divise l'intervalle des valeurs de la série non plus en intervalles de même longueur comme pour de nombreux histogrammes mais en intervalles qui contiennent des pourcentages des données fixés à l'avance ; par exemple :



Écart interquartile : différence entre le troisième et le premier quartile

Ecart-type : racine carré de la variance ; l'unité de l'écart-type est celle des données.

Espérance d'une loi de probabilité sur un ensemble E de nombres : moyenne des valeurs des éléments de E pondérés par leur probabilité.

Espérance d'une variable aléatoire : espérance de la loi de cette variable aléatoire.

Étendue : différence entre la plus grande et la plus petite valeur de la série.

Événement A et B : si A,B sont deux événements, A et B est l'événement $A \cap B$

Événement Aou B : si A,B sont deux événements, A ou B est l'événement $A \cup B$.

Événement : dans le champ des probabilités, un événement est un sous-ensemble de l'ensemble des issues possibles d'une expérience aléatoire.

Expériences aléatoires identiques : signifie qu'on associe à chacune d'elles le même modèle. Les expériences de référence ont d'ailleurs aussi ceci de remarquable que par exemple le même modèle est pertinent pour des pièces (ou des dés) de fabrications différentes lancés par des individus différents : on parlera donc dans ce cas d'expériences identiques.

Expériences de référence : lancers de dés, de pièces équilibrées ; choix de nombres au hasard, tirages au hasard de boules coloriées dans une urne, de cartes dans un jeu etc. Considérons une expérience aléatoire modélisée par une loi de probabilité P sur un ensemble fini E. On pourra le plus souvent trouver une expérience de référence qui à un recodage près des issues possibles est régie par le même modèle ; les questions relatives à l'expérience originelle pourront être traduites dans le cadre de l'expérience de référence. Associer l'expérience originelle à une expérience de référence n'est pas une activité à systématiser : l'élève aura recours à cette possibilité selon ses besoins.

Intervalle interdécile : intervalle dont les extrémités sont le premier et le neuvième décile.

Intervalle interquartile : intervalle dont les extrémités sont premier et le troisième quartile.

Loi de probabilité sur E={x_1,…,x_r} : c'est une liste (p_1,…, p_r) de nombres positifs et de somme 1 , associés aux éléments de E.

Loi des grands nombres : en langage imagé :

Si on choisit n éléments d'un ensemble fini E selon une loi de probabilité P, indépendamment les uns des autres, alors la distribution des fréquences « tend » vers la loi de probabilité P lorsque n tend vers l'infini.

Loi image : Soit T une variable aléatoire défini sur E à valeurs dans E'. Soit P une loi de probabilité définie sur E. La loi image de T est une loi de probabilité définie sur E' par :

Médiane empirique : on ordonne la série des observations par ordre croissant ; si la série est de taille 2n+1, la médiane est la valeur du terme de rang n+1 dans cette série ordonnée ; si la série est de taille 2n, la médiane est la demi-somme des valeurs des termes de rang n et n+1 dans cette série ordonnée. La définition de la médiane n'est pas figée : certains logiciels et certains ouvrages définissent la médiane comme étant le second quartile ou le cinquième décile : dans la pratique de la statistique, les différences entre ces deux définitions sont sans importance ; au lycée, on évitera tout développement la dessus qui ne serait pas une réponse individuelle à une question d'un élève.

Modèle d'une expérience aléatoire : c'est une loi de probabilité sur un ensemble, qui est souvent celui des issues observables de l'expérience.

Neuvième décile (empirique): : c'est le plus petit élément *d'* des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 90% des données soient inférieures ou égales à *d'*.

où P(T=x') désigne la probabilité de l'ensemble des éléments de E dont l'image par T est x' (cet ensemble est souvent noté {T=x'}, on écrit en fait P(T=x') au lieu de P({T=x'})).

P' (x') =P(T=x')

Pièce équilibrée : on choisit pour modéliser le lancer d'une telle pièce l'équiprobabilité de *pile* et de *face*.

Premier décile (empirique): c'est le plus petit élément *d* des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 10% des données soient inférieures ou égales à *d*.

Premier quartile (empirique): c'est le plus petit élément *q* des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 25% des données soient inférieures ou égales à *q*.

Probabilité d'un événement : c'est la somme des probabilités des éléments qui le composent.

Troisième quartile (empirique): c'est le plus petit élément q' des valeurs des termes de la série, ordonnées par ordre croissant, tel qu'au moins 75% des données soient inférieures ou égales à q'.

Variable aléatoire : application définie sur un ensemble muni d'une loi de probabilité P ; son rôle est de transporter P sur un autre ensemble (voir loi image).

Variance d'une loi de probabilité: espérance des carrés des écarts à l'espérance ; c'est aussi la différence entre l'espérance des carrés et le carré de l'espérance.

Variance d'une variable aléatoire : c'est celle de sa loi image.

Variance empirique : moyenne des carrés des écarts à la moyenne ; c'est aussi la différence entre la moyenne des carrés et le carré de la moyenne.