Quelques pistes sur la statistique et les probabilités avec des extraits de la formation de S Ducay du 28/1/11

Statistique et probabilité : Description des observations et modèle théorique.

1. Statistique

La statistique consiste à étudier un ensemble d'objets (on parle de population, composée d'individus ou unités statistiques) sur lesquels on observe des caractéristiques, appelées variables statistiques.

Dans le programme : Effets de structure lors du calcul de moyennes

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.

Exemples d'exercices :

Exercice 2. Le service qualité d'une entreprise de conserverie a pesé un échantillon de 100 boîtes de conserve en fin de chaîne de production et il a obtenu le tableau statistique suivant :

Masse (en g)	[915 ; 919]	]919 ; 923]	]923 ; 927]	]927 ; 931]	]931 ; 935]
Nombre de boîtes	12	25	33	28	?

1) Combien y a-t-il de boîtes dont la masse est comprise entre 931 et 935 g ?
2) a) Construire l'histogramme représentant cette série statistique.
   b) Le tableau statistique et l'histogramme donnent-ils les mêmes informations ?
3) a) Etablir un tableau faisant apparaître les effectifs et les fréquences cumulées croissants.
   b) Représenter graphiquement ces résultats.
   c) En déduire graphiquement la médiane, le premier et le troisième quartile.
   d) Représenter ces résultats par une boîte à moustaches.

Exercice 3. On a demandé à cent élèves d'une classe de seconde le nombre de SMS qu'ils envoyaient au cours d'une semaine :

Nombre de SMS	[0 ; 15]	]15 ; 25]	]25 ; 35]	]35 ; 45]	]45 ; 70]
Nombre d'élèves	15	18	22	16	29

1) Représenter cette série avec un histogramme.
2) Calculer la moyenne et l'écart-type de cette série. Interpréter les résultats obtenus.
3) Déterminer la médiane, le premier et le troisième quartile.

2. Probabilités

Le calcul de probabilités permet de proposer un modèle théorique d'une situation concrète afin de quantifier la fiabilité des affirmations.

Variable aléatoire

Exemple d'exercice:

Exercice 9. Dans une urne contenant 5 boules numérotées de 1 à 5, on tire deux boules avec remise. On effectue le produit des deux nombres obtenus et on note le chiffre unité du nombre produit.

1) a) Déterminer l'univers associé à cette expérience à l'aide d'un tableau à double entrée.
   b) Déterminer les probabilités de tous les événements élémentaires.
   c) Comment présenter ces résultats avec la variable aléatoire X qui à chaque tirage de deux boules avec remise.associe le chiffre unité du nombre produit des deux nombres obtenus.
2) a) Effectuer à l'aide d'un tableur la simulation de 1000 tirages
   b) Calculer la fréquence de chaque résultat possible.
   c) Comparer avec les probabilités du 1).

La loi de Bernoulli et la loi binomiale

La répétition de n expériences identiques et indépendantes à deux ou trois issues peut être représentée par un arbre pondéré. La probabilité d'une liste de résultats (un chemin sur l'arbre) est égale au produit des probabilités de chaque résultat.

Introduction possible

a) 3 lancers d'une pièce équilibrée.

Il y a 2^3 =8 chemins possibles, chaque chemin ayant pour probabilité $\frac{1}{2}\times\frac{1}{2}\times\frac{1}{2}=\left(\frac{1}{2}\right)^3=\frac{1}{8}$.
On peut faire remarquer la cohérence avec l'équiprobabilité « naturelle » P sur l'univers. Ω = {résultats de l'expérience aléatoire} = {chemins}. L'ensemble Ω peut être décrit complètement, et on peut faire le lien entre chaque élément de l'ensemble et le chemin correspondant.

On peut s'intéresser à la probabilité de l'évènement A_k : « obtenir k fois Pile en 3 lancers », pour k = 0, 1, 2, 3.
La probabilité de l'évènement A_k étant égale à la somme des probabilités des évènements élémentaires qui le constituent, on obtiendra en additionnant les probabilités de tous les chemins réalisant k Pile en 3 lancers.

Comme tous les chemins ont la même probabilité $\frac{1}{8}$, on aura $P(A_k)=\frac{1}{8}\times \textit{nombre de chemins réalisant } k \textit{ Pile en 3 lancers}$. Désignant par $\binom{3}{k}$ ce nombre, on aura : $P(A_k)=\binom{3}{k}\frac{1}{8}=\binom{3}{k}\left(\frac{1}{2}\right)^3$

Sur cet exemple, on peut compter les chemins et obtenir : $\binom{3}{0}=1$, $\binom{3}{1}=3$, $\binom{3}{2}=3$ et $\binom{3}{3}=1$.

b) 3 lancers d'une pièce truquée

Par exemple avec la probabilité de Pile $\frac{1}{4}$ et pour face $\frac{3}{4}=1-\frac{1}{4}$
Tous les chemins n'ont plus la même probabilité. On remarque cependant que, suivant le principe de multiplication, les probabilités de tous les chemins réalisant k « Pile » en 3 lancers ont la même probabilité $\left(\frac{1}{4}\right)^k\left(1-\frac{1}{4}\right)^{3-k}$, ce qui donne $P(A_k)=\binom{3}{k}\left(\frac{1}{4}\right)^k\left(1-\frac{1}{4}\right)^{3-k}$

c) Généralisation

En remplaçant $\frac{1}{4}$ par $p\in]0;1[$, puis 3 par n, ce qui donne $P(A_k)=\binom{n}{k}(p)^k(1-p)^{n-k}$

d) Le passage à une variable aléatoire X suivant la loi binomiale se fait « naturellement » : on considère la variable aléatoire X qui à chaque résultat de l'expérience (les 3 lancers) associe le nombre k de Pile obtenus. On observe alors que k peut prendre les valeurs 0, 1, 2, 3, et que X « prend la valeur k » correspondant à l'évènement A_k, ce que l'on note (X = k) = A_k. On a alors $P(X=k)=\binom{n}{k}(p)^k(1-p)^{n-k}$

e) Détermination générale des coefficients binomiaux $\binom{n}{k}$

Considérons un entier naturel non nul *n*, un entier naturel *k* tel que $0\le k\le n$. Considérons la répétition de n expériences identiques et indépendantes à deux issues (succès ou échec) représentée par un arbre pondéré.

Le coefficient binomial $\binom{n}{k}$ correspond au nombre de chemins réalisant *k* succès en *n* expériences.

Exemple : le QCM

Un élève répond au hasard aux 4 questions d'un QCM. Pour chaque question, 3 réponses sont proposées dont une seule est exacte. X la variable aléatoire égale au nombre de bonnes réponses.

Montrer que la loi de probabilité de X est une loi binomiale
Calculer la probabilité d'avoir au moins 3 bonnes réponses
Calculer l'espérance du nombre de bonnes réponses.

Fichier Excel Simulation de la loi binomiale (formation de S Ducay en ligne prochainement)

Exemple d'exercice (d'après BAC ES : Antilles juin 2010)

Un bijoutier propose des perles de culture pour fabriquer des bijoux. Il dispose dans son stock de 1000 perles de deux types de couleurs : les perles argentées et les perles noires.
Chacune de ces perles a :

- soit une forme dite sphérique ;
- soit une forme dite équilibrée ;
- soit une forme dite baroque.

On sait que dans son stock, 44 % des perles sont équilibrées, deux cinquièmes sont baroques et les autres sont sphériques. De plus, 60 % des perles sont argentées dont 15 % sont sphériques et la moitié sont baroques.

1. Recopier le tableau des effectifs ci-dessous et le compléter à l'aide des données de l'énoncé (on ne demande pas de justification).

	Sphérique	Équilibrée	Baroque	Total
Argentée				
Noire				
Total				1000

2. Le bijoutier choisit une perle du stock au hasard. On suppose que chaque perle a la même probabilité d'être choisie.
   On note :
   - *A* l'événement : « la perle est argentée » ;
   - *N* l'événement : « la perle est noire » ;
   - *S* l'événement : « la perle est de forme sphérique » ;
   - *E* l'événement : « la perle est de forme équilibrée » ;
   - *B* l'événement : « la perle est de forme baroque ».

   Toutes les probabilités seront données sous forme décimale exacte.

   a. Quelle est la probabilité que le bijoutier choisisse une perle de forme baroque ?
   b. Quelle est la probabilité que le bijoutier choisisse une perle noire de forme équilibrée ?
   c. Déterminer la probabilité de l'événement $A \cup B$ puis interpréter ce résultat.
   d. Le bijoutier a choisi une perle de forme baroque. Quelle est la probabilité qu'elle ne soit pas argentée ?

3. Pour une création de bijou original, le bijoutier choisit dans son stock quatre perles au hasard et de manière indépendante. On admet que le nombre de perles est suffisamment grand pour que le choix d'une perle soit assimilé à un tirage avec remise.
   a. Calculer la probabilité qu'aucune des quatre perles choisies ne soit argentée.
   b. Calculer la probabilité qu'il y ait au moins une perle sphérique parmi les quatre perles choisies (donner une valeur approchée de ce résultat à 10^{-3} près).

Echantillonnage

Fluctuation d'échantillonnage

Lorsqu'on étudie un caractère sur plusieurs échantillons d'une même population, on peut observer que les résultats ne sont pas identiques selon les échantillons. Plus la taille de l'échantillon étudié est grande, plus les résultats obtenus seront fiables. Cela s'explique par la diminution de la variance, et aussi par la loi des grands nombres.

La fluctuation d'échantillonnage représente la fluctuation entre les différents résultats obtenus d'une même enquête sur différents échantillons d'une même population.

Ces différents résultats présentent une certaine régularité, ce qui se traduit par la notion d'intervalle de confiance.

Intervalle de fluctuation

Lorsqu'on connaît la proportion p d'individus vérifiant une propriété donnée.

On peut en déduire que 95 % des échantillons de taille n auront une fréquence f qui appartiendra à l'intervalle

$\left[p-\frac{1}{\sqrt{n}};p+\frac{1}{\sqrt{n}}\right]$.

En réalité cet intervalle est un intervalle de fluctuation de niveau supérieur ou égal à 0,95.

Intervalle de confiance

Lorsqu'on ne connaît pas la proportion p d'individus vérifiant une propriété donnée, dans une population, mais qu'on a observé une fréquence f de cette propriété sur un échantillon de taille n, on peut dire que p appartient à l'intervalle $\left[f-\frac{1}{\sqrt{n}};f+\frac{1}{\sqrt{n}}\right]$ (à condition que l'échantillon fasse bien partie des 95 % d'échantillons pour lesquels c'est vrai…) Chaque échantillon donne a priori un intervalle différent. Cela s'interprète en disant que 95 % des échantillons fournissent un intervalle qui contient p.

Remarque : Prendre de préférence $n \geq 30$ et $p \in [0{,}2 ; 0{,}8]$

Exemples

1) Le mildiou est une maladie qui attaque 60 % des plants de radis lorsque ceux-ci ne sont pas traités. Un institut spécialisé fait des tests avec quatre produits de traitement en vente libre, sur des échantillons de 105 plants.
Les résultats sont consignés dans le tableau suivant :

Produit de traitement	Pourcentage de plants malades
A	13,9
B	25,9
C	45,4
D	51,2

On veut déterminer les produits que l'on peut supposer efficaces au seuil de 95 %.

Rép : Ici on connaît la proportion théorique de plants malades, il s'agit donc d'un intervalle de fluctuation.
$I_F = [0{,}6-1/\sqrt{105} ; 0{,}6+1/\sqrt{105}] \subset [0{,}5 ; 0{,}7]$.
Ainsi 95% des échantillons ont une fréquence comprise entre 0,5 et 0,7.
On peut alors en déduire que les produits efficaces au seuil de 95 % sont les produits A , B et C.

2) Lors d'un référendum, un sondage aléatoire simple pratiqué sur 1000 personnes a donné 55 % pour le oui et 45 % pour le Non. Peut-on prévoir le résultat du référendum ?

Rép : Ici on connaît la fréquence du oui, 0,55 pour un échantillon de taille 1000. L'intervalle de confiance est $I_C = [0,55-1/\sqrt{1000} ; 0,55+1/\sqrt{1000}] \subset [0,51 ; 0,59]$. Ainsi la probabilité que l'intervalle contienne p est 0,95. On peut donc estimer au seuil de 95 % que le résultat du référendum sera Oui.

Intervalle de fluctuation et loi binomiale

On s'appuie ici sur le document d'accompagnement qui précise le contenu « Utilisation de la loi binomiale pour une prise de décision à partir d'une fréquence » et la capacité correspondante, « Exploiter l'intervalle de fluctuation à un seuil donné, déterminé à l'aide de la loi binomiale, pour rejeter ou non une hypothèse sur une proportion », des programmes du lycée.

On considère une population dans laquelle on suppose que la proportion d'un certain caractère est p. Pour juger de cette hypothèse, on prélève dans la population, au hasard et avec remise, un échantillon de taille n sur lequel on observe une fréquence f du caractère.

On rejette l'hypothèse selon laquelle la proportion dans la population est p lorsque la fréquence f observée est trop éloignée de p, dans un sens ou dans l'autre. On choisit de fixer le seuil de décision de sorte que la probabilité de rejeter l'hypothèse, alors qu'elle est vraie, soit inférieure à 5 %.
Lorsque la proportion dans la population vaut p, la variable aléatoire X correspondant au nombre de fois où le caractère est observé dans un échantillon aléatoire de taille n, suit la loi binomiale de paramètres n et p.
On cherche à partager l'intervalle [0,n], où X prend ses valeurs, en trois intervalles $[0,a-1]$,$[a, b]$ et $[b+1,n]$ de sorte que X prenne ses valeurs dans chacun des intervalles extrêmes avec une probabilité proche de 0,025, sans dépasser cette valeur.

En tabulant les probabilités cumulées $P(X \leq k)$, pour k allant de 0 à n, il suffit de déterminer le plus petit entier a tel que $P(X \leq a) > 0,025$ et le plus petit entier b tel que $P(X \leq b) \geq 0,975$, c'est-à-dire $P(X > b) \leq 0,025$. Autrement dit, a est le plus grand entier tel que $P(X < a) \leq 0,025$. On observe aussi que $a < b$.

La règle de décision est la suivante :

Si la fréquence observée f appartient à l'intervalle de fluctuation à 95 % $\left[\frac{a}{n} ; \frac{b}{n}\right]$, on considère que l'hypothèse selon laquelle la proportion est p dans la population n'est pas remise en question et on l'accepte ; sinon, on rejette l'hypothèse selon laquelle cette proportion vaut p.

Pour $n \geq 30$, $np \geq 5$ et $n(1-p) \geq 5$, on observe que l'intervalle de fluctuation $\left[\frac{a}{n} ; \frac{b}{n}\right]$ est sensiblement le même que l'intervalle $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$ proposé dans le programme de seconde.

Exemple
Monsieur Z, chef du gouvernement d'un pays lointain, affirme que 52 % des électeurs lui font confiance.

On interroge 100 électeurs au hasard (la population est suffisamment grande pour considérer qu'il s'agit de tirages avec remise) et on souhaite savoir à partir de quelles fréquences, au seuil de 5 %, on peut mettre en doute le pourcentage annoncé par Monsieur Z, dans un sens, ou dans l'autre.

1. On fait l'hypothèse que Monsieur Z dit vrai et que la proportion des électeurs qui lui font confiance dans la population est $p = 0,52$. Montrer que la variable aléatoire X, correspondant au nombre d'électeurs lui faisant confiance dans un échantillon de 100 électeurs, suit la loi binomiale de paramètres $n = 100$ et $p = 0,52$.

2. On donne ci-contre un extrait de la table des probabilités cumulées $P(X \leq k)$ où X suit la loi binomiale de paramètres $n = 100$ et $p = 0,52$. Déterminer a et b tels que définis précédemment et comparer les intervalles de fluctuation à 95 % $\left[\frac{a}{n} ; \frac{b}{n}\right]$ et $\left[p - \frac{1}{\sqrt{n}} ; p + \frac{1}{\sqrt{n}}\right]$

k	$P(X \leq k)$
40	0,0106
41	0,0177
42	0,0286
43	0,0444
...	
61	0,9719
62	0,9827
63	0,9897
64	0,9941

3. Énoncer la règle décision permettant de rejeter ou non l'hypothèse $p = 0,52$, selon la valeur de la fréquence f des électeurs favorables à Monsieur Z obtenue sur l'échantillon.

4. Sur les 100 électeurs interrogés au hasard, 43 déclarent avoir confiance en Monsieur Z. Peut-on considérer, au seuil de 5 %, l'affirmation de Monsieur Z comme exacte ?

Remarque

- La recherche de l'intervalle de fluctuation peut-être illustrée par le diagramme en bâton de la loi binomiale de paramètres $n = 100$ et $p = 0,52$.
- *Utilisation du tableur Excel*
  Construire la table des probabilités et des probabilités cumulées de la loi Binomiale de paramètres $n = 100$ *et* $p = 0,52$. *Construire le diagramme en bâton de cette loi.*

Sources

S. Ducay (journée de formation statistiques et probabilités) – Transmath 2de

Ressources

Ressources pour faire la classe en mathématiques (*mise à jour, novembre 2010) pour le LP*