Chapitre 10 : Adéquation à une loi équirépartie -25-05-12-
Terminale ES 2, 2011-2012, Y. Angeli

Objectif : on va observer la répétition indépendante d'une même expérience aléatoire, et se donner un critère pour rejeter ou pas l'hypothèse : « Cette expérience obéit à une loi équirépartie ». On précisera le risque de rejeter à tort cette hypothèse lorsqu'elle est vraie.

1. Un exemple simple : tester si une pièce est équilibrée

Étude théorique d'une pièce équilibrée

On considère une pièce équilibrée.
On répète successivement quatre tirages indépendants de cette pièce (chaque tirage a deux issues équiprobables : pile « P », et face « F »).
On note par exemple $P_1P_2F_3P_4$ l'évènement : « les résultats des tirages sont dans l'ordre : pile, pile, face et pile »
Soit X la variable aléatoire égale au nombre de résultats « P » obtenus lors de quatre tirages.

$P(P_1P_2P_3P_4) =$ ....

$P(X = 4) =$ .... $P(X = 0) =$ ....

$P(P_1P_2P_3F_4) =$ ....

$P(X = 3) =$ ....

$P(X = 1) =$ ....

$P(X = 2) =$ ....

Compléter la loi de probabilité de X (et donner les résultats exacts sous forme décimale) :

k					
$P(X = k)$					

$P(X = 0 \text{ ou } X = 4) =$

Test pour une pièce quelconque

On considère une pièce quelconque (dont on ignore si elle est équilibrée ou non).

On formule l'hypothèse suivante : « La pièce est équilibreé ».

On va rejeter ou non cette hypothèse. On risque donc de commettre deux types d'erreur :
⋆ Rejeter l'hypothèse alors qu'elle est vraie (risque de « première espèce »)
⋆ Accepter l'hypothèse alors qu'elle est fausse (risque de « seconde espèce »)
On lance quatre fois la pièce et on note X le nombre de pile.
⋆ Si $X = 0$ ou $X = 4$, on rejette l'hypothèse « la pièce est équilibrée » au risque de $12,5\%$.
⋆ sinon, on ne peut rejeter l'hypothèse « la pièce est équilibrée » .

Préciser ce que représente ce risque de $12,5\%$: ....

....

2. Méthode générale, exemple du dé

Problème

On considère une expérience aléatoire et on souhaite répondre à la question suivante : peut on rejeter l'hypothèse « l'expérience obéit à une loi équirépartie », au seuil de risque de $n\%$ (de rejeter à tort cette hypothèse lorsqu'elle est vraie) ?

Exemple. On choisit un dé et on veut savoir s'il est équilibré ou non. On réalise un échantillon de 1 000 lancers, et on va déterminer un critère qui nous permmettra de rejeter (ou pas) l'hypothèse : « le dé est équilibré », au seuil de risque de 10% (de rejeter cette hypothèse alors qu'elle est vraie) ? »

Mesurer la distance entre des fréquences observées et des probablités théoriques

On considère un échantillon statistique d'effectif total N, avec k modalités (résultats possibles). On note f_i la fréquence de la modalité i. Afin d'évaluer le degré d'équirépartition de l'échantillon, on considère d^2_{obs}, le carré de la distance entre les fréquences observées f_i et la probabilité théorique $\dfrac{1}{k}$:

$$d^2_{\text{obs}} = \sum_{i=1}^{N} \left(f_i - \frac{1}{k} \right)^2$$

Ce nombre étant souvent petit, on calculera pour davantage de lisibilité : $N\,d^2_{\text{obs}}$ ou $N\,k\,d^2_{\text{obs}}$,

Rappel. fréquence $= \dfrac{\text{effectif de la modalité}}{\text{effectif total}}$.

Exemple. Résultats de l'échantillon de 1 000 lancers du dé étudié :

Modalité i	1	2	3	4	5	6	Total
Effectif n_i	151	163	166	151	201	168	
Fréquence f_i							

$N =$ .... $k =$ .... $P(\text{« 2 »}) =$ ....

On va effectuer le calcul de d^2_{obs} à l'aide de la fonction tableur de la calculatrice : on entre les effectifs dans la liste 1 (L_1).

Commande pour obtenir la liste des fréquences dans L_2 : ....

Commande pour obtenir la liste des termes de d^2_{obs} dans L_3 : ....

$d^2_{\text{obs}} \approx$ .... $N\,k\,d^2_{obs} \approx$ ....

Remarque. De quelles formules la définition de d^2_{obs} est-elle proche ? ....

Lorsque $d^2_{\text{obs}} = 0$, ...

Plus d^2_{obs} est petit, ...

Comment savoir si d^2_{obs} est petit

Exemple. On a simulé informatiquement le tirage de 10 000 échantillons de 1 000 lancers d'un dé équilibré. Pour chacun des échantillons, on a calculé $6\,000d^2$ (suivant la même formule que précédemment) et consigné les résultats dans un tableau :

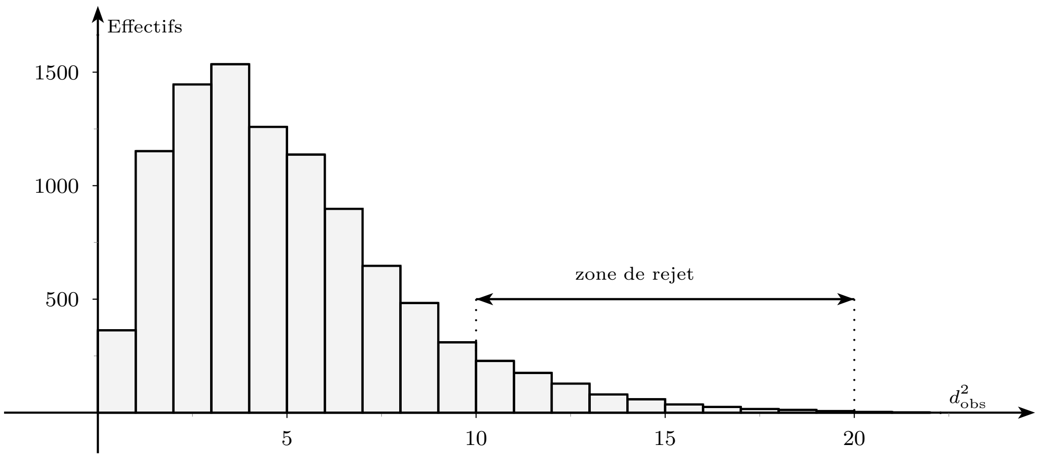
$6000d^2 \in$	$[0;1[$	$[1;2[$	$[2;3[$	$[3;4[$	$[4;5[$	$[5;6[$	$[6;7[$	$[7;8[$	$[8;9[$	$[9;10[$	$[10;11[$	$[11;12[$	$[12;13[$	$[13;+\infty[$
Effectif	363	1 152	1 446	1 535	1 259	1 137	898	647	483	310	228	175	128	239

Méthode. Afin de savoir si le d^2_{obs} calculé auparavant est petit ou grand, on simule informatiquement un grand nombre d'échantillons de même taille N suivant une loi équirépartie et on calcule dans chacun des cas d^2. Si le seuil de risque accepté est de 10%, on va comparer d^2_{obs} (de l'échantillon réel étudié) à D_9 (neuvième décile des échantillons équirépartis simulés) :

- $\star$ Si $k\,N\,d^2_{\text{obs}} > D_9$, on rejette l'hypothèse d'équirépartition avec un risque de 10% de la rejeter lorsqu'elle vraie.
- $\star$ Si $k\,N\,d^2_{\text{obs}} \leq D_9$, on ne peut rejeter l'hypothèse d'équirépartition, on choisira généralement de considérer que la loi recherchée est assez proche d'une la loi équirépartie.

Exemple. $D_9 \in$ .... Conclure : ....

....



Remarque. Le recours à une simulation informatique comme élément de comparaison n'est pas complètement satisfaisant, mais le moyen théorique de l'éviter est hors-programme. (test du Khi deux)

Remarque. Si le risque accepté est de 5%, on utilisera comme critère de décision le 19ème vingtile, s'il est de 1% le 99ème centile, etc...

Remarque. Comment choisir le seuil de risque ? Pas trop faible (sinon, on aurait tendance à accepter toutes les pièces) ni trop élevé (un test qui rejetterait 50% des cas d'équirépartition ne serait plus significatif)...

⚠ le risque dont il est question est le risque de rejeter à tort l'hypothèse lorsque elle est vraie.

3. Amérique du Nord juin 2009

Un pépiniériste a planté trois variétés de fleurs dans une prairie de quelques hectares : des violettes, des primevères et des marguerites. Il se demande s'il peut considérer que sa prairie contient autant de fleurs de chaque variété. Il cueille au hasard 500 fleurs et obtient les résultats suivants :

Variétés	Violettes	Primevères	Marguerites
Effectifs	179	133	188

1. Calculer les fréquences f_V d'une fleur de variété Violette, f_P d'une fleur de variété Primevère et f_M d'une fleur de variété Marguerite. On donnera les valeurs décimales exactes.
2. On note $d^2_{\text{obs}} = \left(f_V - \dfrac{1}{3}\right)^2 + \left(f_P - \dfrac{1}{3}\right)^2 + \left(f_M - \dfrac{1}{3}\right)^2$.
   Calculer $500d^2_{\text{obs}}$. On donnera une valeur approchée arrondie au millième.
3. Le pépiniériste, ne voulant pas compter les quelques milliards de fleurs de sa prairie, opère sur ordinateur en simulant le comptage, au hasard, de 500 fleurs suivant la loi équirépartie. Il répète 2000 fois l'opération et calcule à chaque fois la valeur de $500d^2_{\text{obs}}$. Ses résultats sont regroupés dans le tableau suivant :

$500d^2_{\text{obs}} \in$	[0 ; 0,5[	[0,5 ; 1[	[1 ; 1,5[	[1,5 ; 2[	[2 ; 2,5[	[2,5 ; 3[	[3 ; 3,5[	[3,5 ; 4[	[4 ; 4,5[	[4,5 ; 5[
Effectif	163	439	458	350	231	161	80	47	37	34

   Par exemple : le nombre $500d^2_{\text{obs}}$ apparaît 163 fois dans l'intervalle [0 ; 0,5[.
   On note D_9 le neuvième décile de cette série statistique.
   Montrer que $D_9 \in [2,5\ ;\ 3[$.
4. En argumentant soigneusement la réponse, dire si pour la série observée au début, on peut affirmer avec un risque inférieur à 10 % que « la prairie est composée d'autant de fleurs de chaque variété ».

4. Métropole juin 2003

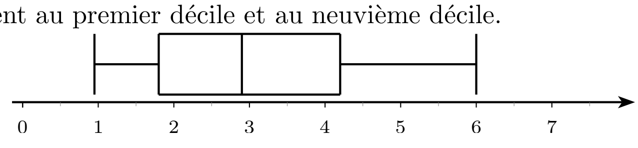
Les guichets d'une agence bancaire d'une petite ville sont ouverts au public cinq jours par semaine : les mardi, mercredi, jeudi, vendredi et samedi.
Le tableau ci-dessous donne la répartition journalière des 250 retraits d'argent liquide effectués aux guichets une certaine semaine.

Jour de la semaine	mardi	mercredi	jeudi	vendredi	samedi
Rang i du jour	1	2	3	4	5
Nombre de retraits	37	55	45	53	60

On veut tester l'hypothèse « le nombre de retraits est indépendant du jour de la semaine ». On suppose donc que le nombre des retraits journaliers est égal à $\dfrac{1}{5}$ du nombre des retraits de la semaine.

On pose $d^2_{\text{obs}} = \displaystyle\sum_{i=1}^{5}\left(f_i - \dfrac{1}{5}\right)^2$ où f_i est la fréquence des retraits du i-ème jour.

1. Calculer les fréquences des retraits pour chacun des cinq jours de la semaine.
2. Calculer alors la valeur de $1\,000d^2_{\text{obs}}$ (la multiplication par 1 000 permet d'obtenir un résultat plus lisible).
3. En supposant qu'il y a équiprobabilité des retraits journaliers, on a simulé 2 000 séries de 250 retraits hebdomadaires.
   Pour chaque série, on a calculé la valeur du $1\,000d^2_{\text{obs}}$ correspondant. On a obtenu ainsi 2 000 valeurs de $1\,000d^2_{\text{obs}}$.
   Ces valeurs ont permis de construire le diagramme en boîte ci-dessous où les extrémités des « pattes » correspondent respectivement au premier décile et au neuvième décile.



   Lire sur le diagramme une valeur approchée du neuvième décile.
4. En argumentant soigneusement la réponse, dire si pour la série observée au début, on peut affirmer, avec un risque d'erreur inférieur à 10 %, que « le nombre de retraits est indépendant du jour de la semaine » ?