

Planche de TP 5 Test du χ^2 d'indépendance

Exercice 1 (Convergence vers une loi du χ^2)

Soit X une variable aléatoire dont l'ensemble des valeurs possibles est partagé en k classes de probabilités p_i . Pour n réalisations de cette variable X , on compte le nombre d'individus N_i appartenant à chaque classe i . L'objectif de cet exercice est d'illustrer le résultat suivant :

$$Z_n = \sum_{i=1}^k \frac{(N_i - np_i)^2}{np_i} \xrightarrow{\mathcal{L}} Z, \text{ quand } n \rightarrow +\infty \text{ pour } Z \sim \chi_{k-1}^2.$$

Cette propriété est fondamentale pour l'utilisation des tests du χ^2 . Dans un premier temps, on s'intéressera à des variables aléatoires suivant une loi de Bernoulli de paramètre 1/2. On verra ensuite comment automatiser le travail réalisé par l'intermédiaire d'une fonction.

1. Affecter respectivement les valeurs 50 et 100 aux variables n et M . Générer une matrice X de taille $M \times n$ dont les entrées sont des 0 ou des 1 tirés aléatoirement et suivant une loi de Bernoulli de paramètre 1/2.
2. Pour chaque ligne $i \in \{1, \dots, M\}$, calculer le nombre de 0 et de 1. Définir un vecteur Z de taille M contenant les différentes réalisations de la variable :

$$\frac{(N_0 - np_0)^2}{np_0} + \frac{(N_1 - np_1)^2}{np_1}$$

Pour a et b fixés, l'idée est d'approcher $P(a < Z_n < b)$ par :

$$\frac{\#\{i : a < Z[i] < b\}}{M}.$$

3. Créer un histogramme des fréquences du vecteur Z . Que produit l'option `proba=TRUE`. Superposer à l'historgramme la densité d'une loi du χ^2 à un degré de liberté (utiliser la commande `dchisq`). On pourra donner un titre au graphique et faire apparaître la taille de l'échantillon.
4. Calculer des approximations de $P(a < Z_n < b)$ pour quelques valeurs de a et b et comparer les résultats obtenus aux résultats théoriques.
5. Reprendre le travail précédant pour différentes valeurs de n et de M (créer une fonction utilisant l'ensemble des commandes précédentes).
6. Mêmes questions si X suit cette fois-ci une loi uniforme sur l'ensemble $\{0, \dots, 5\}$.

Exercice 2 (Efficacité de deux types de médicament)

Pour comparer l'efficacité de deux médicaments agissant sur la même maladie, mais aux prix très différents, la Sécurité Sociale a effectué une enquête sur les guérisons obtenues en suivant chacun des traitements. Les résultats sont consignés dans le tableau suivant :

	Médicament cher	Médicament bon marché
Guérison	44	156
Non-guérison	6	44

1. Effectuer "manuellement" un test du χ^2 pour vérifier l'indépendance de l'efficacité des médicaments en fonction de leur coût au niveau 5%.
2. Faire le même travail en utilisant à présent la fonction `chisq.test`. Afficher les effectifs théoriques et le tableau des résidus.

Exercice 3 (Poissons)

Dans la Durance vivent entre autres deux espèces de poissons : *Chondrostoma nasus* et *Chondrostoma toxostoma*. Ces deux espèces peuvent éventuellement s'hybrider donnant lieu à un troisième groupe que l'on nommera *hybride*. On a échantillonné 626 poissons sur quatre stations : Laragne, Manosque, Pertuis, Cavaillon. Le tableau de contingence correspondant est donné ci-dessous :

	Cavaillon	Pertuis	Manosque	Laragne
Hybrides	39	9	26	3
<i>Chondrostoma toxostoma</i>	84	136	141	15
<i>Chondrostoma nasus</i>	105	13	13	42

On souhaite décider si la répartition des espèces est indépendante de la station.

1. Préciser les hypothèses H_0 et H_1 .
2. Afficher les effectifs théoriques ainsi que les résidus.
3. Préciser le résultat du test du χ^2 respectivement aux niveaux 5% et 1% et commenter les résultats obtenus.