Cours 9 : Test du χ^2

Nous abordons une nouvelle démarche statistique inférentielle : celle de test d'hypothèse. Il s'agit d'une démarche de réfutation : on envisage de réfuter une hypothèse, comme par exemple qu'une pièce de monnaie donnée à les mêmes chances de tomber sur pile que sur face. Réfuter cette hypothèse conduit à conclure que cette pièce est pipée. L'hypothèse envisagée est notée H_0 et est appelée l'*hypothèse nulle*. Dans notre exemple l'hypothèse nulle est donc que la probabilité de pile est $p = 0.5$. Si on procède à n tirages, on n'est pas surpris de ne pas obtenir $n/2$ piles et autant de face, mais la question est de savoir à partir de quel écart à cette valeur théorique il convient de rejeter l'hypothèse nulle. Il y aura bien sûr un seuil de confiance α qui fixera la probabilité de rejeter à tort cette hypothèse. La méthode envisagée ici est une alternative facile à mettre en oeuvre à la simple vérification que les résultats observés ne sont pas dans un intervalle de confiance de la valeur théorique sous l'hypothèse H_0.

1 Mise en place de la statistique

Il convient de formuler l'hypothèse nulle H_0 de façon à ce qu'elle permette clairement, devant un échantillon, de déterminer les quantité suivantes :

- les classes numérotées généralement par $i = 1, \ldots, k$, où k désigne le nombre de classes,
- les nombres O_i d'individus de l'échantillon qui sont dans la classe i ; on les appelle les *effectifs observés*,
- les nombres T_i d'individus qui seraient idéalement dans la classe i sous l'hypothèse H_0. Souvent $T_i = np_i$ si H_0 fixe la probabilité d'être dans la classe i, où n est la taille de l'échantillon. On les appelle les *effectifs théoriques*.

Avec ces nombres on forme alors la *statistique*

$$S = \sum_{i=1}^{k} \frac{(O_i - T_i)^2}{T_i}.$$

On comprend facilement qu'on est d'autant plus enclin à rejeter l'hypothèse nulle que les effectifs observés s'écartent des effectifs théoriques, situation qui correspond à une statistique S élevée. Il convient alors de voir, en fonction de la valeur du seuil de confiance α, à partir de quelle valeur de la statistique S on peut rejeter l'hypothèse nulle.

2 La loi du χ^2 suivie par S

Dès que $k-1$ effectifs observés sont connus, l'effectif de la dernière classe est déterminé, par soustraction à l'effectif total n de l'échantillon. On dit que S a $k-1$ degrés de liberté. Une preuve probabiliste montre que, pour n assez grand, la loi de S peut être assimilée à une *loi du* $\chi^2(k-1)$, dite à $k-1$ degrés de liberté. Dans `R`, pour la loi du χ^2, les instructions usuelles pour une loi sont notées en utilisant le radical `chisq`, pour "chi squared". Ainsi la probabilité d'être inférieure à x pour une variable aléatoire S suivant une loi $\chi^2(k-1)$ est donnée par `pchisq(x,k-1)`. De façon analogue, le quantile q tel que la soit égale à p la probabilité que S soit inférieurs à q est donnée par l'instruction `qchisq(p,k-1)`. La densité de la loi du $\chi^2(k-1)$ au point x vaut `dchisq(x,k-1)`.

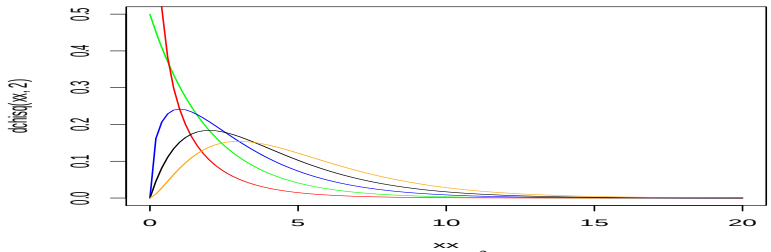


FIGURE 1 – Les densités de la loi du χ^2 à 1,...,5 degrés de liberté.