STATISTIQUES INFÉRENTIELLES : TP1

1. Mesurer la vitesse de la lumière

Dans cette première partie on va illustrer les résultats fondamentaux de probabilité avec un modèle très simple. On suppose qu'on cherche à évaluer la vitesse de la lumière c qu'on arrondira à $c = 3 \times 10^5 m.s^{-1}$. On utilise pour ce faire un appareil de mesure qui commet une erreur indépendante de la valeur de c, et on fera l'hypothèse que l'erreur est normalement distribuée. On suppose donc qu'on fait des mesures indépendantes qui suivent le modèle :

$$X_i = c + \xi_i$$

où $\xi_i \sim \mathcal{N}(0, \sigma^2)$. Dans la suite on supposera qu'on effectue nos mesures en centaines de milliers de mètres par seconde (i.e. à cette échelle $c = 3$) et on prendra $\sigma^2 = 0.5$.

(1) Simuler 100 observations et les placer dans un vecteur **obs**.

(2) On estime c par la moyenne arithmétique des observations. Pour savoir si 100 observations fournissent un bon estimateur on peut procéder de la façon suivante :
- — on simule 100 observations et on calcule l'estimateur $\bar{X}_{100}$ associé
- — on répète cette opération 50 fois (calculer la moyenne empirique de 100 obervations), on conserve à chaque fois l'estimateur $\bar{X}_{100}$
- — quelle est la variance empirique de ces 50 estimateurs ?
- — répéter la procédure avec 1000 observations, que dire de la nouvelle variance empirique ? Quel principe est illustré ici ?

(3) A nouveau on simule 100 observations. Tracer un histogramme des observations.

(4) Calculer la variance empirique des observations, ainsi que l'écart-type. Quelle proportion des observations se situe à plus d'un écart-type de la moyenne ? A deux écart-types ?

(5) Comme dans la question 2 générer un vecteur de 100 estimateur de c basés sur 1000 observations. On s'intéresse à loi de l'écart entre ces estimateurs et c. Comment faut-il renormaliser la différence $\bar{X}_{1000} - c$? Tracer un histogramme de la distribution des estimateurs renormalisés, qu'observe-t'on ? Reprendre cette question avec $\xi_i \sim \mathcal{U}([-0.5, 0.5])$.

2. Un modèle de sondage

On veut estimer la proportion d'alcooliques dans la société française.

(1) On effectue un sondage, c'est-à-dire qu'on interroge une petite proportion de la population française. On suppose que chaque français répond et répond la vérité à la question "Etes vous alcoolique ?". Comment modéliser cette situation ? Donner un estiamteur sans biais de la proportion d'alcooliques. Suposons qu'on interroge 1000 personnes et que la proportion d'acooliques est de un dixième de la population totale, simuler cette situation et calculer la valeur de votre estimateur dans ce cas.

(2) Que dire de la variance de cet estimateur ? En utilisant le raisonnement de la première partie l'évaluer par simulation. Que se passe-t'il dans le cas où $p = 0.2$?

(3) Peut-on évaluer la variance de notre estimateur sans avoir à interroger plus de nos 1000 personnes initiales ? Si oui comment ?

(4) On suppose maintenant qu'une partie de la population va mentir en répondant à la question. On modèlise ce phénomène en supposant que les observations X_i suivent maintenant une Bernoulli de paramètre $\tilde{p} = 0.08$ au lieu de $p = 0.1$. Notre estiamteur est-il toujours non biaisé ? Que devient son risque quadratique ? Le simuler.

(5) Supposons que chaque individu ment avec une proportion de 0.6, modèliser cette situation. Peut on corriger notre estimateur ? Illustrer par la simulation.