Loi Normale

L'objectif de cette séance est d'étudier la loi normale ainsi que son utilité dans l'estimation statistique par l'utilisation d'intervalles de confiance.

Étude de la loi normale

Définition 1 loi normale $\mathcal{N}_{m,\sigma^2}$
On l'appelle aussi loi gaussienne ou loi de Laplace-Gauss. Si $m \in \mathbb{R}$ et si $\sigma > 0$, elle est définie par sa densité :

$$f(x) = \frac{1}{\sigma\sqrt{2\pi}} \exp -\frac{(x-m)^2}{2\sigma^2} \tag{1}$$

On considèrera dans tout le sujet la loi **normale centrée réduite** $\mathcal{N}(0,1)$, i.e. $m = 0$ et $\sigma = 1$. Sa densité est donc :

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp -\frac{x^2}{2}. \tag{2}$$

Exercice 1 Normalisation
Expliquez d'où vient le coefficient $\frac{1}{\sqrt{2\pi}}$ en facteur de la densité.

Exercice 2 Simulation par la méthode de Box-Muller (1958)

1. Soit X, Y des variables aléatoires indépendantes de loi $\mathcal{N}(0,1)$. Quelle est la loi du couple (X,Y) ? Par passage en coordonnées polaires, on note $\rho = \sqrt{X^2+Y^2}$ et $\theta = Arctan\frac{Y}{X}$. Faire un dessin.
2. Donner la loi de ρ et la loi de θ. Montrer que ρ et θ sont indépendants. En déduire un algorithme de génération de loi $\mathcal{N}(0,1)$.

Problème : Estimation statistique

On dispose d'une population de N individus et l'on souhaite préparer une politique publique en fonction du nombre d'individus concernés dans la population. Le problème qui se pose est l'*estimation* du nombre d'individus concernés.

Par exemple, dans la population française (N=65 millions) on cherche à savoir combien d'individus souhaitent se faire vacciner contre la grippe afin d'organiser la politique de vaccination et la commande de vaccins.

On suppose que la décision d'un individu de se faire vacciner est une variable aléatoire de Bernoulli de paramètre p que l'on cherche à déterminer le plus précisément possible.

Une manière classique de procéder est d'*échantillonner* la population de manière uniforme, et de poser la question aux n individus ainsi choisis (sondage). Intuitivement, la qualité de l'estimation dépendra de :
- la taille n de l'échantillon
- la représentativité de l'échantillon (choix uniforme sur l'ensemble de la population)

Question 1.1 : Estimation de p

Comment obtenir une estimation $\hat{p}$ de p à partir de la méthode précédente ? Justifier.

On suppose qu'un sondage de $n = 30$ individus donne les résultats suivants :

0 1 0 0 0 0 1 1 1 0 1 0 1 0 1 0 1 1 1 1 0 1 1 1 0 1 0 1 1 1 1 1 0

Question 1.2 : Calcul de moyenne

Donner une estimation numérique $\hat{p}$ de p.
En déduire une première estimation du nombre M de vaccins à prévoir.

On cherche maintenant à quantifier l'erreur commise sur p par cette estimation, afin de savoir quelle marge prendre sur le nombre de vaccins à préparer.

Question 1.3 : Erreur

Comment caractériser la loi de l'erreur commise sur p par cette méthode ?

Question 1.4 : Précision

Quelle est la probabilité que l'estimateur $\hat{p}$ soit à plus de 0.03 de p ?
Combien cela représente-t-il de vaccins potentiellement manquants ou inutiles ?

On souhaite finalement prendre une marge sur le nombre de vaccins afin que la probabilité d'en manquer soit inférieure à 1%.

Question 1.5 : Intervalle de confiance

Comment s'assurer d'un intervalle d'estimation de p qui soit exact avec une telle probabilité ?