Estimation ponctuelle et par intervalle de confiance

Ce TP est inspiré d'un TP de Bruno Portier, enseignant de statistique inférentielle en GM3 à l'INSA de Rouen.

Estimation par la méthode du maximum de vraisemblance

L'objet de cette partie est d'étudier par simulations le comportement de l'estimateur du maximum de vraisemblance du paramètre d'une loi géométrique.
Soit X une variable aléatoire de loi géométrique de paramètre p avec $p \in]0,1[$. On s'intéresse à l'estimation du paramètre p. On sait que $\mathbb{E}(X) = 1/p$. Il est donc facile de proposer un estimateur du paramètre p. En effet, si on dispose d'un échantillon de variables aléatoires $X_1, X_2, \ldots, X_n$, indépendantes et de même loi géométrique de paramètre p, alors $\overline{X}_n$ est un estimateur sans biais de $1/p$. Ainsi, pour estimer p, on peut proposer l'estimateur $\widehat{p}_n = 1/\overline{X}_n$. Il se trouve que cet estimateur est aussi l'estimateur du maximum de vraisemblance. L'objet de cette première partie est d'étudier par simulations le comportement à distance finie de l'estimateur du paramètre p.

1 Simuler 400 échantillon de 1000 réalisations d'une variable aléatoire géométrique de paramètre $p = 2/5$ et stocker la dans la matrice `data`.

2 Pour chaque colonne de données, construisez la suite des valeurs **successives** de l'estimateur du paramètre p et stocker les résultats dans une matrice `M`.

3 Tracer, sur le même graphique, les boites à moustaches des 400 estimations pour les tailles d'échantillon $n = 50, 100, 200, 500$ et 1000 .

4 Tracer, sur ce même graphique, la valeur de p et son estimation pour chaque taille d'échantillon.

5 Analyser et commenter les résultats obtenus. Que sommes-nous amener à illustrer avec un tel graphique ?

Estimation par intervalle de confiance

L'objet de cette partie est de construire des estimateurs par intervalle de confiance pour des paramètres de la loi gaussienne.
Dans un centre avicole, des études antérieures ont montré que le poids d'un oeuf choisi au hasard peut être considérée comme la réalisation d'une variable aléatoire gaussienne X, d'espérance μ et de variance σ^2. On admet que les poids des oeufs sont indépendants les uns des autres. On prend un échantillon de $n = 36$ oeufs que l'on pèse. Les mesures obtenues (exprimées en g) sont dans (par ordre croissant) le fichier `oeufs.txt` sur disponible Moodle.

1 Stocker les données dans le vecteur `x`. (utiliser la fonction `np.genfromtxt(,)`) et tracer l'histogramme des fréquence et superposer sur ce dernier la densité de la loi normal. Commenter

2 Proposer des estimateurs sans biais de la moyenne μ et de la variance σ^2 et donner une estimation de ces deux paramètres. On notera, respectivement, ces estimateur $\bar{X}, S^2$ et les estimations $\bar{x}$ et s^2

3 Quelle est la loi de la variable $T = \sqrt{n}(\bar{X} - \mu)/S$.

4 En vous aidant de la loi de la variable aléatoire T, construisez un intervalle de confiance bilatérale au niveau 95% pour le poids moyen des oeufs.

5 En vous aidant de cet intervalle, un oeuf ayant un poids égal à 60 g est-il typique ou atypique de la population ?.

> On sait que la demi-longueur de l'intervalle de confiance pour μ est égale à $t_{35,1-\alpha/2} \times s/\sqrt{36}$ ($t_{35,1-\alpha/2}$ étant le quantile d'ordre $1 - \alpha/2$ de la variable T).

6 A quel niveau de confiance correspondrait un intervalle centré en $\bar{x}$ et de demi-longueur 0,76 ? (`Utiliser la fonction t.cdf() dans scipy.stats`)

> On souhaite maintenant construire un intervalle de confiance bilatérale au niveau 95% pour la variance σ^2 de la variable aléatoire X

7 Quelle est la loi de la variable aléatoire $Z = (n-1)S^2/\sigma^2$.

8 Calculez les quantiles d'ordre 0.975 et 0.025 de la variable aléatoire Z.

9 Construisez un intervalle de confiance au niveau 95% pour la variance et commenter.