M.A. Haddouche
A. Rogozan Année 2019-2020

ASI3

Estimation ponctuelle et par intervalle de confiance

Partie 1

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. On rappelle que la loi géométrique de paramètre p est la loi du nombre X d'épreuves de Bernoulli indépendantes de même paramètre p, nécessaires à l'obtention d'un succès. On note alors $X \sim G(p)$. Ainsi, X prend les valeurs 1, 2, 3, ... avec pour tout $x \in \mathbb{N}^*$, $\mathbb{P}[X = x] = p(1-p)^{x-1}$. En utilisant la fonction génératrice, on montre que $\mathbb{E}(X) = 1/p$.

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 dont on 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 $n = 50, 100,200, 500, 1000$ et tracer, sur ce même graphique, la valeur de p.

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

Partie 2

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 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 $\sqrt{n}(\bar{X} - \mu)/S$.

4 Construisez un intervalle de confiance bilatérale au niveau 95% pour le poids moyen des oeufs et commenter.

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

5 Quelle est la loi de la variable aléatoire $Z = (n-1)S^2/\sigma^2$, où S^2 est un estimateurs sans biais de σ^2

6 Calculez les quantiles d'ordre 0.975 et 0.025 de Z.

7 Construire un intervalle de confiance au niveau 95% pour la variance et commenter.

7 On sait que la demi-longueur de l'intervalle de confiance pour μ est égale à $t_{35,1-\alpha/2} \times s/\sqrt{36}$. A quel niveau de confiance correspondrait un intervalle centré en m et de demi-longueur 0,76 ? (`Utiliser la fonction chi2.cdf() dans scipy.stats`)