TP n°1
Intervalles de confiance et estimateurs

Préambule :

— Scilab est un logiciel libre, téléchargeable ici.

— Scilab possède une interface d'aide consultable en tapant `help` dans la console. De manière générale, pour obtenir de l'aide concernant une fonction `bidule`, tapez `help bidule`. Scialb possède aussi une aide en ligne.

— Pour vous familiariser avec les fonctions de base de Scilab, vous trouverez de nombreux tutoriels sur internet. Vous pouvez aussi vous reporter au TP n°1 de CMMA, disponible ici, ou à un script contenant les fonctions basiques ici.

Dans ce TP, on apprendra à se familiariser avec les intervalles de confiance et les estimateurs. C'est aussi l'occasion de réviser la création de fonctions et de scripts, la génération de variables aléatoires, etc.

1 Échauffement

On considèrera $X_1, \dots, X_n$ un échantillon de taille n d'une loi dont on souhaite déterminer l'espérance m.

1. Créer une fonction `estim(`X`)` qui retourne la moyenne empirique et la taille d'un vecteur X. Plus précisément, `estim(`$X_1, \dots, X_n$`)` doit retourner le couple $(\overline{X}_n, n)$.
2. a. Créer un vecteur colonne X retournant 100 réalisations de la loi de Bernoulli $\mathscr{B}(1/4)$ (on pourra utiliser la fonction `grand`).

   b. Quelle est la différence entre les commandes `estim(`X`)` et `[`$xbar, n$`]=estim(`X`)` ?

2 Un premier échantillon

On considère dans cet exercice que la loi de X est la loi gaussienne $\mathscr{N}(m, 2)$, avec m à estimer. On rappelle que la moyenne empirique $\overline{X}_n$ suit la loi $\mathscr{N}(m, 2/n)$, et donc, si $Z \sim \mathscr{N}(0, 2/n)$,

$$\mathbb{P}(\overline{X}_n - a \leq m \leq \overline{X}_n + a) = \mathbb{P}(-a \leq Z \leq a).$$

1. En utilisant la fonction Scilab `cdfnor('X',...)`, retrouver que le quantile d'ordre $0,975$ de la loi gaussienne centrée réduite est 1.96.
2. Un échantillon de la loi $\mathscr{N}(m, 2)$ a été enregistré ici.

   a. Télécharger ce fichier puis, à l'aide de la fonction `fscanfMat`, créer un vecteur X contenant cet échantillon.

   b. Quelle est la taille n de l'échantillon ?

3. a. Créer une fonction `quantile` (n, α) qui retourne le réel a tel que

$$\mathbb{P}(-a \leq Z \leq a) = \alpha.$$

   b. En déduire un intervalle de confiance exact au seuil 95% pour m (à l'aide de l'échantillon récupéré précédemment).

4. On souhaite désormais apprécier l'évolution de l'intervalle de confiance au fur et à mesure que l'on obtient de nouvelles données. Créer une fonction `interv` (k, X, α) qui retourne les bornes d'un intervalle de confiance exact au seuil α pour m en prenant en compte uniquement les k premières valeurs de X (étant un échantillon de taille n, on choisira donc $k \leq n$). Tracer l'évolution de ces bornes pour $k \in \{1, \dots, n\}$ et $\alpha = 90\%$.
5. Générer 1000 échantillons de taille $n = 100$ de la loi $\mathscr{N}(0, 2)$ et illustrer de manière numérique la précision de l'intervalle de confiance généré à l'aide de la fonction `interv`.

3 Intervalles de confiance approchés

Dans cet exercice, on supposera que la loi de X est la loi de Bernoulli $\mathscr{B}(p)$, où p est un paramètre à estimer.

1. Un échantillon de la loi $\mathscr{B}(p)$ a été enregistré ici. Créer un vecteur X contenant cet échantillon.
2. Créer une fonction `intervExc` (X, α) retournant un intervalle de confiance par excès au niveau α pour m, à l'aide de l'inégalité de Bienaymé-Tchebychev. Que dire de cet intervalle ?
3. Créer une fonction `intervAsy` (X, α) retournant un intervalle de confiance asymptotique au niveau α pour m, à l'aide du théorème central limite.
4. Comparer les deux intervalles obtenus.

4 Pour aller plus loin…

1. Générer 10 000 variables aléatoires de loi gaussienne centrée réduite, en utilisant **uniquement** la fonction `rand()` comme générateur d'aléa (i.e. toute autre fonction Scilab ne générant pas de variables aléatoires est autorisée). Vérifier à l'aide de la fonction Scilab `histplot`.
2. a. En ne générant **que** des variables aléatoires gaussiennes, donner une approximation de la densité de la loi du chi-deux à k degrés de liberté $\chi^2(k)$.

   b. Même question pour la densité de la loi t de Student à n degrés de liberté $\mathscr{T}(n)$.