Colles série 7 : Intervalles de confiance

pour les sujets 1-2-3 : données `apnee`

Sujet 1:

1. Extraire du data.frame `apnee` l'échantillon des poids des femmes et l'affecter à `poidsf`. Extraire celui du poids des hommes et le nommer `poidsh`.

2. Calculer l'intervalle de confiance au niveau 90% et l'estimation du poids moyen d'une femme. De même pour le poids moyen d'un homme. Comparer les deux intervalles obtenus. Qu'en conclure ? (on utilisera `t.test`)

Sujet 2:

1. Extraire du data.frame `apnee` l'échantillon des poids des femmes et l'affecter à `poidsf`. Extraire celui du poids des hommes et le nommer `poidsh`.

2. Calculer l'estimation sans biais de la variance du poids d'une femme, celle de son écart-type, et calculer un intervalle de confiance de niveau 95% pour la variance puis en déduire l'intervalle de confiance pour l'écart-type. Commentaire.

Sujet 3:

1. Soit p la probabilité inconnue qu'une personne souffre d'apnée du sommeil. Calculer l'estimation de p obtenue à l'aide de l'échantillon observé.

2. Calculer l'intervalle de confiance de niveau approximatif 95% pour p, à la main (avec la formule du cours) et à l'aide de la fonction `prop.test`. Comparer.

pour les sujets 4-5-6 : modèle uniforme

Dans les exercices suivants, on étudie les propriétés de biais de trois estimateurs de a obtenus avec un échantillon de taille n du modèle uniforme sur $[0, a]$. On etudie aussi deux intervalles de confiance asymptotiques basés sur deux des estimateurs proposés.

Rappelons que si X suit une uniforme sur $[0, a]$, $f(x) = 1/a$ si $x \in [0, a]$ et 0 sinon. De plus (exo de TD) on sait montrer que $E(X) = a/2$ et $V(X) = a^2/12$. On pose $T_1 = 2\bar{X}_n$ et $T_2 = maxX_i = X_{(n)}$.

Sujet 4: Biais de T_2 et T_3

1. Choisir trois valeurs pour a, N et n. Tirer N réplications d'échantillons de taille n de X (uniforme sur $[0, a]$) qui seront stockées dans une matrice à N lignes et n colonnes. Calculer les N réalisations de T_2 obtenues et les affecter à un vecteur de taille N qu'on pourra nommer `echT2`. Faire de même pour $T_3 = (n + 1)T_2/n$.

2. Proposer une expérience numérique qui montre que T_2 est biaisé (i.e. $E(T_2) - a \neq 0$) alors que T_3 est sans biais (i.e. $E(T_3) - a = 0$). Lequel des deux retiendra t-on ?

Sujet 5: Intervalle de confiance construit avec T_2 ou T_3

1. Choisir trois valeurs pour a, N et n. Tirer N réplications d'échantillons de taille n de X (uniforme sur $[0, a]$) qui seront stockées dans une matrice à N lignes et n colonnes. Calculer les N réalisations de T_2 obtenues et les affecter à un vecteur de taille N qu'on pourra nommer `echT2`.

2. On définit à présent T_3=$(n+1)T_2/n$. Calculer les N réalisations de T_3 ainsi que les N réalisations de l'intervalle $[T_3 n/(n+1); T_3 n\alpha^{(-1/n)}/(n+1)]$. Montrer ensuite numériquement que cet intervalle est de niveau de confiance $1-\alpha$ pour tout n même petit.

Sujet 6: Intervalle de confiance construit avec T_1

1. Choisir trois valeurs pour a, N et n. Tirer N réplications d'échantillons de taille n de X (uniforme sur $[0, a]$) qui seront stockées dans une matrice à N lignes et n colonnes. Calculer les N réalisations de T_1 obtenues et les affecter à un vecteur de taille N qu'on pourra nommer `echT1`.

2. Montrer numériquement que $I(a, \alpha) = [\sqrt{3n}T_1/(\sqrt{3n} + u_{1-\alpha/2}); \sqrt{3n}T_1/(\sqrt{3n} - u_{1-\alpha/2})]$ pour n grand est un intervalle de confiance pour a de niveau de confiance approximatif $1-\alpha$.