Statistiques TP 2
Intervalles de confiance

Pour déterminer des intervalles de confiance pour une espérance, on a besoin des nombres t_α tels que $P(|X| > t_\alpha) = \alpha$, où X suit soit une loi normale, soit une loi de Student. Pour les déterminer avec maple, on va utiliser le fait que les lois de Student et la loi normale sont symétriques, et donc que $P(|X| > t) = 2(1 - P(X \leq t)$, et le fait que maple donne accès à la fonction réciproque de la fonction de répartition : la commande
t:= Quantile(RandomVariable(Normal(0,1)),0.375);
par exemple, fournit le nombre t tel que $P(X \leq t) = 0.375$, si X suit une loi normale centrée réduite. D'autre part, la loi de Student à n degrés de liberté est connue par maple sous le nom de *StudentT(n)*.

1 Intervalles de confiance pour l'espérance d'une loi normale

1. Ecrire une procédure qui prendra en entrée d'une part un échantillon d'une loi normale d'espérance a priori inconnue, d'autre part l'écart-type de la loi (supposé connu donc) et le niveau α, et fournit en sortie un intervalle de confiance de niveau α pour l'espérance sous la forme d'une liste $[a, b]$. Remarques éventuellement importantes... :
   la commande Sample de maple fournit un échantillon qui est un vecteur. Pour connaître la taille d'un vecteur u , on peut utiliser la commande *op(1,u)*. Sinon, il faut mettre la taille de l'échantillon dans les paramètres d'entrée de la procédure.

   Pour que la procédure fournisse $[a, b]$ en sortie, il suffit de terminer la procédure par $[a, b]$;

2. Ecrire une procédure qui prendra en entrée un échantillon d'une loi normale d'espérance et de variance a priori inconnues, et un seuil α, et fournit en sortie un intervalle de confiance de niveau α pour l'espérance sous la forme d'une liste $[a, b]$. Rappel : La commande *StandardDeviation* fournit l'écart-type de l'échantillon basé sur l'estimateur non biaisé de la variance.

3. Utiliser ces procédures pour faire le premier exercice du TD sur les intervalles de confiance.

4. Générer un échantillon de taille 20 de la loi normale d'espérance 2.5 et d'écart-type 3. Comparer les deux intervalles de confiance pour l'espérance fournis par les deux procédures précédentes au même niveau $\alpha = 0.05$. Est-ce qu'ils contiennent l'espérance ?

5. Avec le même échantillon, obtenir deux listes I_{min} et I_{max} dont la ième valeur donne respectivement la valeur minimale et maximale de l'intervalle de confiance de niveau $\frac{i}{100}$, l'écart-type étant supposé connu.

6. Représenter sur le même graphe les valeurs extrêmales de l'intervalle de confiance en fonction de α, ainsi qu'une droite horizontale pour la vraie espérance. Jusqu'à quel seuil α l'intervalle de confiance contient-il l'espérance ? (Une réponse graphique suffit)

7. Faire une boucle qui génère successivement 200 300-échantillons de loi normale d'espérance 2.5 et d'écart-type 3, calcule les 200 intervalles de confiance successifs au risque 0.1 pour l'espérance de cette loi avec écart-type connu et compte le nombre de fois N où l'intervalle en question contient l'espérance. Combien vaut ici $\frac{N}{200}$?

   Question subsidiaire : quelle est la loi de N ?

8. *Question subsidiaire* Tracer les 20 premiers intervalles de confiance obtenus à la question précédente sous forme de barres verticales à l'abscisse correspondant au numéro de l'échantillon, ainsi qu'une droite horizontale correspondant à la vraie espérance.

Indications : *plot*$([[a,b],[c,d]])$ dessine un segment entre les points $[a,b]$ et $[c,d]$. On peut donc ici générer 21 plots différents $G[i]$ avec les 20 intervalles (faire une boucle !) et la droite horizontale, puis les représenter sur le même graphique, à l'aide de *plots[display](seq(G[i],i=1..21));*.

2 Intervalle de confiance pour l'espérance d'une loi quelconque : application à la méthode de Monte-Carlo

Lorsque le nombre d'observations est grand, on peut remplacer la loi de Student par la loi normale centrée réduite. De plus, lorsqu'on n'a pas un échantillon d'une loi normale, mais qu'on peut appliquer le théorème central limite, si on connaît l'écart-type, la deuxième procédure s'applique également, et si on ne connaît pas l'écart-type, on peut utiliser la première procédure, (éventuellement en remplaçant la loi de Student par une loi normale).

Soit $(U_1,\ldots,U_d)$ un vecteur de d variables aléatoires indépendantes toutes de loi uniforme sur $[0,1]$, et f une fonction de carré intégrable sur $[0,1]^d$. On considère la variable aléatoire $X = f(U_1,\ldots,U_d)$. D'après la loi forte des grands nombres, si X_n est une suite de variables aléatoires indépendantes toutes de même loi que X, $\lim_{n\to\infty} \frac{1}{n}\sum_{i=1}^n X_i = E(f(U_1,\ldots,U_d)) = \int_{[0,1]^d} f(u_1,\ldots,u_d)du_1\ldots du_d$. La méthode de Monte Carlo consiste à utiliser ceci pour calculer une approximation de $\int_{[0,1]^d} f(u_1,\ldots,u_d)du_1\ldots du_d$. Le thérorème central limite permet de plus d'avoir une idée de l'ordre de grandeur de l'erreur. La convergence est en $\frac{\sigma}{\sqrt{n}}$, indépendamment de d. L'intérêt par rapport à d'autres méthodes déterministes, qui approchent l'intégrale par exemple par des sommes de Riemann, se fait donc surtout sentir pour des intégrales multiples. En effet l'erreur avec ce genre de méthodes est en général en (Pas de la subdivision)$^{-1}$.

On va se servir de la méthode de Monte-Carlo pour calculer une approximation de π.

1. On va d'abord utiliser $\pi = \int_0^1 4\sqrt{1-x^2}dx$. Générer un 500 échantillon $[U_1,\ldots,U_{500}]$ de loi uniforme sur $[0,1]$. Obtenir l'échantillon $[X_1,\ldots,X_{500}]$ où $X_i = 4\sqrt{1-U_i^2}$. (On peut utiliser *map* : *map(x−> f(x), U)* applique la même fonction f à tous les éléments du vecteur U). Donner une estimation de $\pi = E[X]$ et de l'écart-type de X. Donner un intervalle de confiance pour π au risque 0.05.

2. On va déterminer un intervalle de confiance pour π en utilisant $\pi = 4\int_{[0,1]^2} 1_{x^2+y^2\leq 1}dxdy$. Si (U_1,U_2) suit une loi uniforme sur $[0,1]^2$, quelle est la loi de $X = 41_{U_1^2+U_2^2\leq 1}$? Donner son espérance et son écart-type en fonction de π.

   Utiliser votre 500 échantillon précédent $[U_1,\ldots,U_{500}]$ pour obtenir une estimation de π et de l'écart-type de X_i. Donner un intervalle de confiance pour π au risque 0.05.

3. Déterminer un intervalle de confiance pour π en utilisant $\pi = 6\int_{[0,1]^3} 1_{x^2+y^2+z^2\leq 1}dxdydz$.