Statistiques TP 2
Intervalles de confiance avec Maple**7**

1 Loi de Student

La loi de Student à n degrés de liberté est connue par maple sous le nom de *studentst[n]*.

Remarque : pour connaître les différentes lois connues par maple, aller dans l'aide du package *stats* (par exemple en tapant *?stats;*) puis dans *distributions* ou taper directement *?distribution;*.

Générer des graphes $G[i]$ de la densité de la loi de Student pour des degrés de liberté $5i$ allant de 5 à 50 par pas de 5, (faire une boucle) et les représenter sur le même graphe que la densité de la loi normale centrée réduite (faites celle-ci de couleur différente en utilisant l'option *color=...* de *plot*; on peut prendre *color=blue* par exemple).
Rappel : pour tracer plusieurs graphes G_1, G_2, G_3 on doit utiliser la commande $plots[display](G_1, G_2, G_3)$;
Lorsqu'il y en a n, on peut fabriquer une suite à l'aide de *seq* puis utiliser
$plots[display](seq(G[i], i = 1..n))$;

Qu'est-ce qui est ici illustré ?

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:= statevalf[icdf,normald[0,1]](0.375);
par exemple, fournit le nombre t tel que $P(X \leq t) = 0.375$.

1. 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 $[a, b]$. Remarques éventuellement importantes... :
   la commande $describe[count](data)$ fournit le nombre d'observations de l'échantillon *data*.
   La commande $describe[standarddeviation[1]](data)$ fournit l'écart-type de l'échantillon basé sur l'estimateur non biaisé de la variance. Si on ne met pas le [1], il est basé sur l'estimateur biaisé de la variance.
   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 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 $[a, b]$.

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 tableaux de taille 100 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. Obtenir deux autres tableaux pour les intervalles de confiance avec écart-type inconnu.

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. 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).

   Faire une boucle qui génère 200 300-échantillons de loi $\Gamma(2,1)$ (elle s'appelle *gamma[2,1]*), fournit les 200 intervalles de confiance successifs au risque 0.1 pour l'espérance de cette loi (Remarque : on n'aura plus besoin des échantillons après, donc ce n'est pas la peine de les conserver, on peut les réeffacer au fur et à mesure), compte le nombre de fois N où l'intervalle en question contient l'espérance (qui est 2). Combien vaut ici $\frac{N}{200}$.

   Question subsidiaire : quelle est la loi de N ?

8. Tracer les 20 premiers intervalles de confiance sous forme de barres verticales à l'abscisse correspondant au numéro de l'échantillon, ainsi qu'une droite horizontale correspondant à la vraie espérance (qui est 2).

   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 avec les 20 intervalles et la droite horizontale, puis les représenter sur le même graphique.

9. Tracer les deux courbes donnant les deux valeurs extrémales de l'intervalle de confiance en fonction du numéro de l'échantillon (pour les 200 échantillons) ainsi qu'une droite horizontale correspondant à l'espérance.

10. Lorsqu'on cherche à estimer une probabilité, on cherche en fait à estimer l'espérance d'une loi de Bernoulli. Donc on peut aussi obtenir un intervalle de confiance en utilisant les procédures précédentes. Cependant, dans le cas d'une loi de Bernoulli de paramètre p, l'écart-type est $\sqrt{p(1-p)}$. Donc plutôt que de l'estimer par l'estimateur $\hat{sigma}$, on préfère souvent l'estimer par $\sqrt{\hat{p}(1-\hat{p})}$, plus simple à obtenir.

    Écrire une procédure qui prend en paramètres la taille n de l'échantillon de la loi de Bernoulli, le nombre de 1 de l'échantillon, le niveau α, et fournit un intervalle de confiance au risque α pour l'espérance de la loi de Bernoulli. Faire l'exercice 5 de la feuille de TD, ainsi que les premières questions de l'exercice 8.