Statistiques TP 2
Probabilités et statistiques avec Maple**7**

A l'agrégation, seul maple7 étant pour l'instant disponible, on repart.... presque à 0. On commence par charger le package "stats", en tapant *with(stats):*. On aura aussi besoin du package "statplots", et du package "plots".

On va travailler au début sur une variable aléatoire supposée modéliser la taille d'une population, qui suit une loi normale $N(\mu, \sigma^2)$ avec $\mu = 175$ et $\sigma = 10$.

1 Manipulation de lois, utilisation de commandes de base

Différence avec la feuille précédente : les commandes RandomVariable, et les manipulations formelles comme par exemple ExpectedValue, PDF(X,x), CDF(X,x) n'existent plus.

Pour une loi normale d'espérance 175 et d'écart-type 10, on peut :

- obtenir la valeur de la densité en un point au moyen de *statevalf[pdf,normald[175,10]](180);*
- obtenir la valeur de la fonction de répartition en un point au moyen de *statevalf[cdf,normald[175,10]](180);*
  (Pour une variable aléatoire discrète, on peut obtenir la probabilité qu'elle vaille k à l'aide de *pf* : par exemple,) *statevalf[pf,binomiald[100,0.3]](35);*
- obtenir le graphe de la densité au moyen de *plot(statevalf[pdf,normald[175,10]],140..200);*
- obtenir une valeur approchée de l'espérance de X à l'aide de *f1:=x− >statevalf[pdf,normald[175.10]](x); evalf(int(x*f1(x),x=100..300));*
  obtenir un échantillon de taille 1000 de X (pour maple, l'objet obtenue est une liste) à l'aide de
  Ech:=[random[normald[175,10]](1000)];
  (On peut ne pas mettre les [] extérieurs, mais dans ce cas à chaque fois qu'on manipule Ech dans la suite, il faut écrire *[Ech]*)
- obtenir la moyenne empirique, l'écart-type de l'échantillon précédent à l'aide de
  describe[mean](Ech);
  describe[standarddeviation](Ech);
- obtenir un quantile à l'aide de
  describe[quantile[975/1000]](Ech)
  Attention 0.975 ne marche pas, il faut vraiment une fraction.
- dessiner un histogramme de l'échantillon à l'aide de *histogram(Ech);*. Si on veut des rectangles de largeur constante, utiliser *histogram(Ech,area=1)*.
- obtenir la courbe des quantiles à l'aide de *scatterplot(Ech, format=quantile);*
- obtenir la fonction de répartition empirique à l'aide de *xyexchange(scatterplot(Ech, format=quantile));*
- obtenir une "boî te à moustache" à l'aide de *boxplot(Ech)*.

2 Statistiques sur un échantillon de taille n

Ne refaites pas cette partie ! J'ai juste donné pour information comment répondre aux questions avec les commandes de maple7.

1. Générer un 100-échantillon *Ech* de loi normale $N(175, 100)$ à l'aide de la commande *random*
   Ech:=random[normald[175,10]](100);

2. Calculer la moyenne empirique $\bar{x} = \frac{\sum_{i=1}^{n} X_i}{n}$ de l'échantillon de trois manières : en faisant une boucle pour calculer la somme, à l'aide de la commande *add* (utilisée pour sommer un nombre fixé de termes), et à l'aide de la commande *mean*.
   add(Ech[i],i=1..100)/100;
   moy:=describe[mean]([Ech]);

3. Calculer l'écart-type empirique $\bar{\sigma} = \sqrt{\frac{\sum_{i=1}^{n}(X_i-\bar{x})^2}{n-1}}$ (attention au $n-1$, on verra sa justification plus tard en TD...) de deux manières : directement à l'aide de *add*, et à l'aide de *standarddeviation*.
   eca:=describe[standarddeviation]([Ech]);

4. Calculer le quantile correspondant à 0.975 à l'aide de *quantile*.
   describe[quantile[975/1000]]([Ech]);

5. Tracer un histogramme correspondant à l'échantillon *Ech* à l'aide de la commande *histogram*. Le représenter sur le même graphique que la vraie densité de X et que la densité estimée $N(\bar{x}, \bar{\sigma})$.
   P1:=histogram([Ech]);
   P2:=plot(statevalf[pdf,normald[175,10]],140..200);
   P3:=plot(statevalf[pdf,normald[moy,eca]],140..200);
   plots[display](P1,P2,P3);

6. Construire un tableau qui contienne les quantiles de 0 à 1 par pas de 0.05. Tracer une fonction de répartition empirique à l'aide de ce tableau.
   qu:=array[1..20];
   for i from 1 to 20 do qu[i]:=describe[quantile[i/20]]([Ech]); end do:
   plot([seq([qu[i],i/20],i=1..20)]);

3 Illustration de la loi des grands nombres

La loi des grands nombres est le théorème qui sera bientôt vu en Probabilités, qui dit que si $(X_n)_{n \in N}$ est une suite de variables aléatoires indépendantes de même loi, d'espérance finie m, alors pour presque tout ω, $\frac{\sum_{i=1}^{n} X_i(\omega)}{n}$ converge vers m : la moyenne arithmétique converge vers l'espérance. Cela entraine que si de plus $E(X_i^2)$ est finie, l'écart-type empirique converge presque sûrement vers le vrai écart-type. D'autre part, ici, on a également convergence des fonctions de répartition empiriques vers la fonction de répartition théorique.

1. Générer un échantillon de X de taille 5000 et obtenir deux tableaux dont le ième terme soit respectivement la moyenne empirique, l'écart-type empirique des i premiers termes.
   On pourra par exemple vérifier et utiliser $\sum_{i=1}^{n}(X_i - (\frac{1}{n}\sum_{j=1}^{n} X_j))^2 = \sum_{i=1}^{n} X_i^2 - \frac{1}{n}(\sum_{i=1}^{n} X_i)^2$
   On peut aussi construire des sous-échantillons : par exemple
   [seq(Ech[i],i=1..10)] construit un sous échantillon formé des 10 premiers termes de Ech.

2. Représenter pour des tailles n allant de 30 à 5000 par pas de 10 la moyenne empirique $\bar{x}$ en fonction de n ainsi qu'une droite horizontale correspondant à l'espérance; faire la même chose pour l'écart-type.

3. Représenter les fonctions de répartition empiriques pour les tailles 100, 500,1000 et 5000 ainsi que la fonction de répartition théorique.

4. On considère maintenant une variable aléatoire suivant une loi de Cauchy de paramètres $(0,1)$ de densité $\frac{1}{\pi(1+x^2)}$. Est-elle d'espérance finie ? Effectuer les mêmes opérations que sur X.

4 Théorème central limite

1. On l'a vu, la moyenne empirique converge vers l'espérance. Représenter pour des tailles n allant de 30 à 5000 par pas de 10 $\frac{\sqrt{n}}{\sigma}(\bar{x}-\mu)$.

   Cette quantité ne semble pas converger, mais elle reste d'ordre 1. Néanmoins, le théorème central limite affirme que si (X_n) est une suite de variables aléatoires indépendantes de même loi et d'espérance μ et de variance σ^2 finies, $\frac{\sqrt{n}}{\sigma}(\bar{x}-\mu)$, où $\bar{x}$ est la moyenne arithmétique des n premiers X_i, converge en loi vers une loi normale centrée réduite. Cette quantité reste donc aléatoire, mais on sait comment elle est distribuée asymptotiquement.

2. Ici les X_i suivent une loi gaussienne, donc on connaît exactement la loi de $\frac{\sqrt{n}}{\sigma}(\bar{x}-\mu)$. Quelle est-elle ?

3. Construire une procédure qui prend en entrée deux entiers n et N_{obs}, génère N_{obs} n-échantillons de X, calcule pour chacun de ces échantillons $\frac{\sqrt{n}}{\sigma}(\bar{x}-\mu)$ et trace sur un même graphique un histogramme de ces grandeurs ainsi que la densité de la loi normale centrée réduite.

   Rappel : une procédure se présente sous la forme :
   $Bidule := proc(parametres)$
   $local\, i :: integer$: $(ligne\, pas\, indispensable)$
   $instructions$;
   $end\, proc$;

4. Tester la procédure avec $Nobs = 1000$ et $n = 30, 50, 100, 300, 1000$.

5. L'intérêt de ce théorème est qu'il marche avec toute loi de variance finie. Soit Z une variable aléatoire qui suit une loi binomiale $B(n,p)$.

   (a) Rappeler l'espérance de Z et sa variance. Pour $n = 100$ et $p = 0.3$, obtenir le graphe des $(k, P(Z = k))$ à l'aide de pf. Comparer avec la densité d'une variable aléatoire normale de même espérance et variance.

   (b) Z suit la même loi qu'une somme de n variables de Bernoulli indépendantes de paramètre p. On peut donc appliquer le théorème central limite. Donner une procédure qui prend n, p et N_{obs} en entrée, génère un N_{obs} échantillon $(Z_1, \dots Z_{Nobs})$ de loi binomiale $B(n,p)$ et trace un histogramme des $\frac{Z_i - np}{\sqrt{np(1-p)}}$ ainsi que la densité de la loi normale centrée réduite. Lancer la procédure avec $N_obs = 1000$, $n = 100$ et $p = 0.1$ puis 0.3, 0.5, 0.99. Recommencer avec $n = 300$.

   (c) On peut voir aussi la convergence en loi sur le graphe de la loi de probabilité. Tracer pour différentes valeurs de n et p le graphe formé des $(\frac{k-np}{\sqrt{np(1-p)}}, \sqrt{np(1-p)}P(Z = k))$, ainsi que la densité de la loi normale centrée réduite.

5 Challenge

J'ai mis sur ma page web (http://www.math.u-bordeaux1.fr/ chabanol), dans le répertoire Enseignement puis Statistiques 6 échantillons obtenus en simulant des lois classiques. Récupérez-les et faites-les lire par maple à l'aide de la commande *importdata('toto')*. Serez-vous capable de deviner quelle loi a généré chaque échantillon ?