Analyse des données en sciences de la Terre
Partiel 3 : Statistique

9/12/2010

Exercice 1 : distribution laplacienne (7 points)

La densité de probabilité normalisée d'une distribution laplacienne de médiane m et de paramètre d'échelle b s'exprime de la manière suivante :

$$f(x) = \frac{1}{2b} \exp\left[-\frac{|x-m|}{b}\right].$$

1. En ajoutant une légende aux axes, représenter la densité de probabilité $g(x)$ d'une distribution laplacienne réduite ($m = 0$, $b = 1$) pour x entre -10 et 10 par pas de 0,01.
2. Calculer sa fonction de répartition $h(x) = \int_{-10}^{10} g(x)dx$ et la représenter dans une sous-figure juste en dessous de la densité de probabilité. Mettre une légende aux axes.
3. Par une méthode de votre choix, calculer la probabilité d'être inférieur ou égal à $-b$.
4. Par une méthode de votre choix, calculer la probabilité d'être supérieur à $+b$.
5. Calculer ou déduire des questions précédentes la probabilité d'être dans l'intervalle $]-b\ ;\ +b]$.
6. BONUS : Quelle valeur attendriez vous pour l'intégrale $\int_{-\infty}^{+\infty} f(x)dx$?

Exercice 2 : comparaison de deux moyennes (6 points)

1. Comment définit-on une fonction sous matlab ? Comment s'y prend-on ensuite pour l'utiliser ?
2. Ecrire une fonction nommée "comparaison" qui compare deux moyennes expérimentales à l'aide d'un test de Student. Cette fonction devra prendre en entrée deux arguments : la série de valeurs x et la série de valeurs y. Elle devra fournir en sortie trois arguments : le nombre d'éléments de x, le nombre d'éléments de y et la valeur de la statistique.
3. Utiliser cette fonction avec les vecteurs A et B puis calculer la valeur critique de la statistique pour $\alpha = 0.05$.
4. BONUS : Vérifier à l'intérieur de votre fonction que les arguments d'entrée x et y sont bien des vecteurs colonnes.

Exercice 3 : ANOVA et moindres carrés (7 points)

L'ANOVA peut également être utilisée pour dire si un modèle de moindres carrés est pertinent pour expliquer les données. Au lieu de comparer la variance inter-groupe et la variance intra-groupe, on compare la variance du modèle (MSM) et la variance des résidus (MSD).

1. Soit le fichier "donnees.txt" qui contient en première colonne une série d'abscisses x et en deuxième colonne une série d'ordonnées y. Lire les données et calculer un modèle $\widetilde{y}$ consistant en une tendance quadratique.
2. Calculer les deux sommes suivantes :

$$\begin{aligned} SSM &= \sum_i^N (\widetilde{y}_i - \overline{y})^2 \\ SSD &= \sum_i^N (y_i - \widetilde{y}_i)^2 \end{aligned}$$

3. Calculer les degrés de liberté suivants : $DFM = I - 1$ et $DFD = N - I$, où I correspond au nombres de paramètres du modèle et N correspond au nombres de données. I et N doivent être déterminés à partir des dimensions de la matrice normale des moindres carrés.
4. Calculer $MSM = SSM/DFM$, $MSD = SSD/DFD$, puis la statistique de Fisher $F = MSM/MSD$.
5. Vous obtenez une statistique $F = 1.18$ et une valeur critique $F_c = 3.03$ (pour un niveau de confiance $\alpha = 0.05$). La commande La commande `fp(F,DFM,DFD)` donne par ailleurs 0.69. Que pouvez-vous en conclure ?
6. BONUS : Représenter sur le même graphique le modèle en bleu et les données en rouge.

Rappels sur les moindres carrés

$$\left\{ \begin{array}{lcl} y_1 &=& a + bx_1 \\ y_2 &=& a + bx_2 \\ & ... & \\ y_n &=& a + bx_n \end{array} \right. \quad \Leftrightarrow \quad \begin{pmatrix} y_1 \\ y_2 \\ ... \\ y_n \end{pmatrix} = \begin{pmatrix} 1 & x_1 \\ 1 & x_2 \\ ... & ... \\ 1 & x_n \end{pmatrix} \begin{pmatrix} a \\ b \end{pmatrix} \quad \Leftrightarrow \quad y = Ax$$

solution des moindres carrés sous Matlab : `x=A\y`