**TRAVAUX PRATIQUES SUR LES LOIS DE PROBABILITES CONTINUES**

**I. FICHIERS DE DONNEES**

**Exemples d'ensembles de données**

Le logiciel Statistics and Machine Learning Toolbox ™ comprend les exemples d'ensembles de données dans le tableau suivant.

Pour charger un ensemble de données dans l'espace de travail MATLAB®, il faut saisir : ‘***’load filename’’***

| **File** | **Description** |
| --- | --- |
| acetylene.mat | Chemical reaction data with correlated predictors |
| arrhythmia.mat | Cardiac arrhythmia data from the UCI machine learning repository |
| carbig.mat | Measurements of cars, 1970–1982 |
| carsmall.mat | Subset of carbig.mat. Measurements of cars, 1970, 1976, 1982 |
| census1994.mat | Adult data from the UCI machine learning repository |
| cereal.mat | Breakfast cereal ingredients |
| cities.mat | Quality of life ratings for U.S. metropolitan areas |
| discrim.mat | A version of cities.mat used for discriminant analysis |
| examgrades.mat | Exam grades on a scale of 0–100 |
| fisheriris.mat | Fisher's 1936 iris data |
| flu.mat | Google Flu Trends estimated ILI (influenza-like illness) percentage for various regions of the US, and CDC weighted ILI percentage based on sentinel provider reports |
| gas.mat | Gasoline prices around the state of Massachusetts in 1993 |
| hald.mat | Heat of cement vs. mix of ingredients |
| hogg.mat | Bacteria counts in different shipments of milk |
| hospital.mat | Simulated hospital data |
| humanactivity.mat | Human activity recognition data of five activities: sitting, standing, walking, running, and dancing |
| imports-85.mat | 1985 Auto Imports Database from the UCI repository |
| ionosphere.mat | Ionosphere dataset from the UCI machine learning repository |
| kmeansdata.mat | Four-dimensional clustered data |
| lawdata.mat | Grade point average and LSAT scores from 15 law schools |
| mileage.mat | Mileage data for three car models from two factories |
| moore.mat | Biochemical oxygen demand on five predictors |
| morse.mat | Recognition of Morse code distinctions by non-coders |
| nlpdata.mat | Natural language processing data extracted from the MathWorks® documentation |
| ovariancancer.mat | Grouped observations on 4000 predictors |
| parts.mat | Dimensional run-out on 36 circular parts |
| polydata.mat | Sample data for polynomial fitting |
| popcorn.mat | Popcorn yield by popper type and brand |
| reaction.mat | Reaction kinetics for Hougen-Watson model |
| spectra.mat | NIR spectra and octane numbers of 60 gasoline samples |
| stockreturns.mat | Simulated stock returns |

Exemple sous Matlab: si on veut charger le fichier des notes d’examens qui comprend cinq ensembles de notes écrites sur cinq colonnes:

**load examgrades**

**x1 = grades(:,1);**

**x2 = grades(:,2);**

**x3 = grades(:,3);**

**x4 = grades(:,4);**

**x5 = grades(:,5);**

**hold on**

**plot(x1, 'b');**

**plot(x2, 'r');**

**plot(x3, 'g');**

**plot(x4, 'y');**

**plot(x5, 'p');**

**title('Notes d’examen');**

**xlabel('Nombre d’étudiants');**

**ylabel('Note sur 100');**

**II. LOI NORMALE OU GAUSSIENNE**

La distribution de probabilité continue la plus largement utilisée en statistique est la distribution de probabilité normale.

Une variable aléatoire est une **variable normale** quand elle dépend d’un grand nombre de causes indépendantes dont aucune n’est prépondérante.

Densité de probabilité d’une loi normale est donnée par:



Elle est notée **N(µ,σ) où µ est la moyenne statistique et σ l’écart type, c’est-à-dire :**

**E(X) = µ V(X) = σ2**

Sous Matlab, nous avons plusieurs fonctions et instructions qu’on peut utiliser dans le cas d’une loi normale. Par exemple en utilisant le fichier **examgrades**. A partir de ses données, soit x1, x2, x3, x4 ou x5, nous allons créer une distribution normale en essayant de les ajuster. Pour cela nous devons commencer la fonction Matlab suivante :

***pN = fitdist(x,'Normal')***

Cette fonction nous permet alors d’obtenir la valeur moyenne mu et l’écart type sigma qui nous donnerons la loi normale. Chacun de ces deux paramètres sera calculé avec un intervalle de confiance à 95%.

Ensuite, il suffit d’utiliser la fonction ***y = normpdf(x,mu,sigma)*** pour obtenir la loi normale souhaitée

**1er exemple**

Comme nous avons, dans l’exemple précédent, cinq groupes de notes nous pouvons créer alors cinq densités de probabilités normales ou gaussiennes

***load examgrades***

***x1 = grades(:,1);***

***x2 = grades(:,2);***

***x3 = grades(:,3);***

***x4 = grades(:,4);***

***x5 = grades(:,5);***

***figure;***

***hold on***

***plot(x1, 'b');***

***plot(x2, 'r');***

***plot(x3, 'g');***

***plot(x4, 'y');***

***plot(x5, 'p');***

***title('Notes d’examen');***

***xlabel('Nombre d’étudiants');***

***ylabel('Note sur 100');***

***hold off***

***pN1 = fitdist(x1,'Normal')***

***mu1= pN1.mu;***

***sigma1= pN1.sigma;***

***pN2 = fitdist(x2,'Normal')***

***mu2= pN2.mu;***

***sigma2= pN2.sigma;***

***pN3 = fitdist(x3,'Normal')***

***mu3= pN3.mu;***

***sigma3= pN3.sigma;***

***pN4 = fitdist(x4,'Normal')***

***mu4= pN4.mu;***

***sigma4= pN4.sigma;***

***pN5 = fitdist(x5,'Normal')***

***mu5= pN5.mu;***

***sigma5= pN5.sigma;***

***x\_values\_1 = round(mu1-2\*sigma1):1: round(mu1+2\*sigma1);***

***x\_values\_2 = round(mu2-2\*sigma2):1: round(mu2+2\*sigma2);***

***x\_values\_3 = round(mu3-2\*sigma3):1: round(mu3+2\*sigma3);***

***x\_values\_4 = round(mu4-2\*sigma4):1: round(mu4+2\*sigma4);***

***x\_values\_5 = round(mu5-2\*sigma5):1: round(mu5+2\*sigma5);***

***y1 = normpdf(x\_values\_1,mu1,sigma1);***

***y2 = normpdf(x\_values\_2,mu2,sigma2);***

***y3 = normpdf(x\_values\_3,mu3,sigma3);***

***y4 = normpdf(x\_values\_4,mu4,sigma4);***

***y5 = normpdf(x\_values\_5,mu5,sigma5);***

***figure;***

***hold on***

***plot(x\_values\_1, y1, 'b');***

***plot(x\_values\_2, y2, 'r');***

***plot(x\_values\_3, y3, 'g');***

***plot(x\_values\_4, y4, 'y');***

***plot(x\_values\_5, y5, 'o');***

***title('Densités de probabilités normales');***

***xlabel('Note sur 100');***

***ylabel('Probabilités');***

***hold off***

***% Pour calculer la fonction de repartition on peut utiliser la commande cdf***

***Y1=cdf('Normal', x\_values\_1,mu1,sigma1);***

***Y2=cdf('Normal', x\_values\_2,mu1,sigma2);***

***Y3=cdf('Normal', x\_values\_3,mu1,sigma3);***

***Y4=cdf('Normal', x\_values\_4,mu1,sigma4);***

***Y5=cdf('Normal', x\_values\_5,mu1,sigma5);***

***figure;***

***hold on***

***plot(x\_values\_1, Y1, 'b');***

***plot(x\_values\_2, Y2, 'r');***

***plot(x\_values\_3, Y3, 'g');***

***plot(x\_values\_4, Y4, 'y');***

***plot(x\_values\_5, Y5, 'o');***

***title('Fonctions de répartitions');***

***xlabel('Note sur 100');***

***ylabel('Probabilités cumulées');***

***hold off***



 

**2ème exemple**

Nous allons maintenant charger un autre exemple de données, correspondant aux poids des patients.

***load hospital***

***x = hospital.Weight;***

***figure;***

***hold on***

***plot(x, 'b');***

***title('Poids des patients d’un hôpital');***

***xlabel('Patient');***

***ylabel('Poids en livres');***

***hold off***

***pN = fitdist(x,'Normal')***

***mu= pN.mu;***

***sigma= pN.sigma;***

***x\_values = round(mu-4\*sigma1):1: round(mu+4\*sigma1);***

***y = normpdf(x\_values,mu,sigma);***

***figure;***

***plot(x\_values, y, 'b');***

***title('Densités de probabilités normales');***

***xlabel('Poids en livres');***

***ylabel('Probabilités');***

***hold off***

***% Pour calculer la function de repartition on peut utilizer la commande cdf***

***Y=cdf('Normal', x\_values,mu,sigma)***

***% Y est la function de repartition ou bien encore la densité de probabilité cumulée***

***figure;***

***plot(x\_values, Y, 'b');***

***title('Fonctions de répartitions');***

***xlabel('Poids en livres');***

***ylabel('Probabilités cumulées');***

  

**LOI DE STUDENT vs LOI NORMALE CENTREE REDUITE**

La distribution de Student est une famille de courbes dépendant d’un seul paramètre ν (les degrés de liberté). Lorsque les degrés de liberté ν se rapprochent de l'infini, la distribution t se rapproche de la distribution normale standard.

La loi normale standard est obtenue pour un mu=0 et sigma=1 à savoir N(0,1) ou loi normale centrée réduite. Sous Matlab on utilise la fonction z = normpdf(x,0,1)pour générer la densité de probabilité de N(0,1);

Sous Matlab la densité de probabilité de Student peut être calculée à l’aide de la fonction **y = tpdf(x,nu);** ou nu est le degré de liberté, par exempe :

1. Loi normale centrée réduite vs Loi de Student avec nu=1

***x = -50:1:50;***

***y = tpdf(x,1);***

***z = normpdf(x,0,1);***

plot(x,y,'-.',x,z,'--',x,y-z,'-')

legend('Loi de Student avec \nu=1', ...

'Loi Normale Standard', ...

'Différence entre Loi normale standard et Student','Location','best')

xlabel('Observation')

ylabel('Densité de Probabilité')

title('Comparaison entre loi normale standard et Loi de Student')

******

1. Loi normale centrée réduite vs Loi de Student avec nu=5

***x = -50:1:50;***

***y = tpdf(x,5);***

***z = normpdf(x,0,1);***

plot(x,y,'-.',x,z,'--',x,y-z,'-')

legend('Loi de Student avec \nu=5', ...

'Loi Normale Standard', ...

'Différence entre Loi normale standard et Student','Location','best')

xlabel('Observation')

ylabel('Densité de Probabilité')

title('Comparaison entre loi normale standard et Loi de Student')

******

1. Loi normale centrée réduite vs Loi de Student avec nu=10

***x = -50:1:50;***

***y = tpdf(x,10);***

***z = normpdf(x,0,1);***

plot(x,y,'-.',x,z,'--',x,y-z,'-')

legend('Loi de Student avec \nu=10', ...

'Loi Normale Standard', ...

'Différence entre Loi normale standard et Student','Location','best')

xlabel('Observation')

ylabel('Densité de Probabilité')

title('Comparaison entre loi normale standard et Loi de Student')

******