Séance n°1 : estimation de la densité de probabilité

Module de Génie Informatique

`Laboratoire LTSI - UMR INSERM 642 - Université de Rennes1`

Ce TP aura pour objet de familiariser les étudiants avec quelques notions de bases de probabilités et d'initier à la programmation sous Matlab. Matlab est un logiciel de programmation adapté au calcul matriciel et disposant de nombreuses routines mathématiques notamment statistiques. Ce sont ces dernières que nous exploiterons lors de ce TP afin de permettre l'estimation de la densité de probabilité d'une variable aléatoire à partir de N réalisations de cette dernière.

1) Tout d'abord, lançons l'application "Matlab.exe", et ouvrons un nouveau fichier à l'aide de l'éditeur de Matlab. Ce dernier prendra l'extension ".m" lors de l'archivage. Tout script possédant l'extension ".m" peut être exécuter sous Matlab à partir de la fenêtre de dialogue. Ainsi on tapera "tp1" suivi du "retour chariot" afin d'exécuter le programme sauvegardé sous le nom "tp1.m".

2) Puis générerons $N = 100$ réalisations d'une variable aléatoire gausienne X d'espérance nulle et de variance unité. Ceci se fait à l'aide de la commande :

$$x = \text{randn}(1, N);$$

Quelle est la nature de la quantité x ? Que représente la quantité $x(n)$? De plus, quel est l'effet du point virgule placé derrière la commande "randn" ?

3) A présent nous allons représenter graphiquement les N réalisations obtenues à l'aide du script suivant :

figure(1);
plot(x);

Quel est le rôle de chacune des commandes précédentes ? On pourra s'aider de la fonction "help" de Matlab afin d'obtenir le descriptif détaillé des commandes en question, en tapant par exemple dans la fenêtre de dialogue :

help figure

De plus, nous pourrons habiller le graphe obtenu à l'aide des commandes suivantes :

title('Représentation graphique d un processus');
xlabel('Nombre d''échantillons');
ylabel('Valeurs du processus');

4) A présent, représentons l'histogramme[1] du vecteur x à l'aide de la commande "hist". Pour cela, exécutons les commandes suivantes :

$M = 10$;
figure(2);
hist(x, M);

[1] L'histogramme d'un tableau de données donne l'occurrence de chaque valeur du tableau.

Puis :

$pas = 1\mathrm{e}{-1};$
$dx = -2.9 : pas : 2.9;$
$\mathrm{hist}(x, dx);$
$\mathrm{figure}(3);$

Quelle différences observe-t-on entre les figures 2 et 3 ? Que se passe-t-il lorsque l'on augmente le nombre M ou bien lorsque l'on diminue le nombre *pas* ?

5) Dès lors, nous allons tâcher de comprendre ce que permet d'estimer le vecteur de données défini par "hist(x, M)" ou bien par "hist(x, dx)". Pour cela, représentons graphiquement la courbe asociée à la densité de probabilité[2] de la variable X. On rappelle que, X étant une variable aléatoire gaussienne, sa densité de probabilité est définie par :

$$\forall\ t \in \mathbb{R}, \qquad f(t) = \frac{1}{\sigma\sqrt{2\pi}}\ \mathrm{e}^{-\frac{1}{2}\left(\frac{t-m}{\sigma}\right)^2}$$

X étant supposée centrée réduite, que valent en fait les quantités m et σ ? Construisons alors, de deux manières distinctes, le tableau *tabf* contenant les valeurs prises par la fonction f en différents points.
– utilisons tout d'abord une boucle "for (...) end" ;
– puis, construisons le tableau sans utiliser de boucle, uniquement par produit et addition matriciels.
En effet, alors que la première manière de programmer est plus intuitive, la seconde est rendue possible car Matlab possède des opérateurs matriciels non définis en Mathématiques tels que ".*" et "./" évitant entre autres l'emploi du boucles "for (...) end" et réduisant ainsi le coût de calcul. Ces opérateurs permettent d'effectuer une opération terme à terme entre deux matrices. On rencontre également l'opérateur ".^". Quel est en particulier l'effet de la commande "randn(N, N)^3" ? Enfin, que permet d'estimer le vecteur de données défini par "hist(x, M)" ou bien par "hist(x, dx)" ? que remarque-t-on lorsque le nombre N augmente ?

[2] On appelle densité de probabilité d'une variable aléatoire réelle continue une fonction f i) positive ou nulle sur $\mathbb{R}$, ii) intégrable sur $\mathbb{R}$ et iii) vérifiant $\int_{\mathbb{R}} f(t)dt = 1$.