Travaux pratique n°1
Aléatoire et information
Cours TS115

1. Introduction

Ce premier TP a pour objectif de faire la synthèse de plusieurs notions introduites en cours de probabilités, processus aléatoires, théorie de l'information et de mathématiques au semestre 1 de la formation en filière Télécommunications.

En particulier, il s'agit d'étudier différentes distributions de variables aléatoires (VA) à l'aide de matlab et d'analyser leur pertinence au travers de différentes applications. Ce TP permet ainsi de se familiariser avec les bases du traitement du signal, qui sont utilisées au cours de ce semestre. En effet, le traitement du signal permet de manipuler du son, de la parole et de l'audio, mais aussi d'analyser des signaux biomédicaux, d'aborder des problématiques liées aux signaux sonar et radar, à la localisation et à la navigation GPS, au tracking de missiles, d'étudier des systèmes de communications mobiles, etc. Dans ces différentes applications, des problèmes d'estimations sont à résoudre et reposent souvent sur l'emploi des statistiques.

Le travail peut être mené en binôme.

Evaluation

L'évaluation du travail repose sur un rapport et une note de travail continu. Le rapport de six de pages maximum doit être dactylographié (sous *Word* ou *latex*). Sous Word, les équations doivent être générées avec l'éditeur d'équations et numérotées. Les commentaires doivent être pertinents et tout résultat justifié. Les programmes matlab peuvent être mis en annexe. Des références bibliographiques peuvent être introduites et seront regroupées dans une section en fin de rapport.

<u>Rappel :</u> lorsque l'on utilise les fonctions sous matlab, la commande *help* dans la fenêtre de commande permet d'obtenir des informations sur les fonctions et la manière de les utiliser.

2. Variable aléatoire gaussienne univariée

Préambule : pourquoi utiliser une hypothèse gaussienne ?

La loi gaussienne est une hypothèse couramment utilisée dans le traitement du signal et des images pour décrire la distribution d'une grandeur à estimer. En effet, de nombreux phénomènes naturels réels peuvent être décrits à partir d'une loi gaussienne de moyenne μ et de variance σ^2 définie comme suit :

$$f_X(x) = \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2}\left(\frac{x-\mu}{\sigma}\right)^2\right). \tag{1}$$

Par exemple, en télécommunications, le bruit lié à la transmission d'un message est modélisé comme un processus gaussien de moyenne nulle. Pour valider les performances d'un décodeur (TNT par exemple), en vue de sa commercialisation, il faut analyser sa robustesse pour des variances de bruit élevées par rapport à celle du signal informatif. L'hypothèse gaussienne est également présente dans la navigation GPS, pour

le traitement des signaux émis par les satellites qui sont perturbés par un bruit blanc gaussien dû principalement à la traversée de l'atmosphère.
Cette notion est donc très importante et vous la retrouverez présente tout au long de votre cursus à l'ENSEIRB.

Manipulation n°1 : Générer *N = 10000* réalisations d'une VA suivant une loi gaussienne $N(\mu,\sigma^2)$ avec ($\mu=2$) et $\sigma^2=9$ en utilisant la fonction *randn* de Matlab.

Tracer l'histogramme des valeurs prises par cette VA à l'aide de la fonction *hist*.

Sur une même figure, représenter la fonction densité $f_X(x)$ calculée à l'aide de (1) et l'histogramme estimé. Montrer ainsi visuellement que ces quantités sont équivalentes à une constante de normalisation près (voir l'annexe).

Manipulation n°2 : Retrouver numériquement l'expression de l'entropie théorique de la variable gaussienne à partir des valeurs de l'histogramme (écrire une fonction qui calcule l'entropie discrète à partir de l'histogramme). Vous prendrez la loi gaussienne de moyenne 2 et de variance 9 et 10000 échantillons.

3. Variable multidimensionelle aléatoire gaussienne

Génération d'une variable bivariée gaussienne

Question n°1 : Retrouver l'expression de la loi gaussienne à P composantes indépendantes dont les variances sont identiques.

Question n°2 : *Démontrer que si on prend la transformation* $\boldsymbol{y} = \Sigma_y^{1/2}\boldsymbol{x} + \boldsymbol{\mu}$ avec $\boldsymbol{x} \sim N(\mathbf{0}, \mathrm{I})$ alors nous générons des échantillons du vecteur $\boldsymbol{y}$ selon la loi $N(\boldsymbol{\mu}, \Sigma_y)$.

Manipulation n°3 : A partir de cette transformation, générer 1000 échantillons $\boldsymbol{y}$ de covariance $\Sigma_y = \begin{pmatrix} 2 & 1 \\ 1 & 2 \end{pmatrix}$ *et* $\boldsymbol{\mu} = (0.2,\ 0.2)^T$. (Pour obtenir $\Sigma_y^{1/2}$ il faut utiliser [U,D]=eig(A); B=U*D.^(0.5)*U';)

Manipulation n°4 : Visualiser l'histogramme 2D (vous pouvez utiliser les fonction matlab suivantes : histogram2, contour, contour3, surf). Comparer à l'histogramme de $\boldsymbol{x} \sim N(\mathbf{0}, \mathrm{I})$ où I est la matrice identité de taille (2x2)?

Manipulation n°5 : Estimer la valeur du coefficient de corrélation à l'aide de la fonction corrcoef.

Mesures quantitatives

Les données traitées dans ce paragraphe représentent les enregistrements des hauteurs des précipitations relevées dans trois villes différentes Bordeaux, Nantes et Santiago (Chili). Il s'agit des hauteurs pluviométriques relevées toutes les heures pendant 50 jours. L'objectif de cette analyse est de comparer les régimes pluviométriques en évaluant les corrélations et les similarités existantes entre les trois villes considérées.

<u>Manipulation n°6 :</u> Charger le vecteur X_pluv.mat en utilisant la commande matlab load. Chaque colonne de ce vecteur présente les données d'une ville spécifique : (X1: ville de Bordeaux, X2 : ville de Nantes, X3 : ville de Santiago). Pour récupérer les colonnes de X_pluv sous matlab, utiliser pour X1 : X1= X_pluv(:,1).

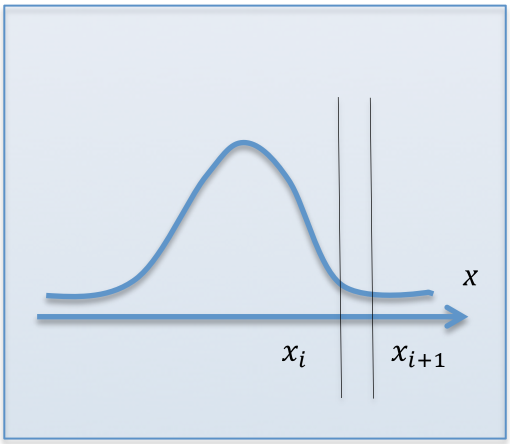
<u>Manipulation n°7 :</u> Calculer les matrices de la covariance pour les variables suivantes **Y1**=[X1, X2], **Y2**=[X1, X3], **Y3**=[X2, X3]. Utiliser la fonction matlab cov. Que peut-on déduire sur les corrélations existantes entre les précipitations des différentes villes ?

<u>Manipulation n°8 :</u> Calculer sous Matlab, l'information mutuelle pour les couples **Y1** et **Y2** puis pour **Y1** et **Y3** à l'aide de leurs statistiques d'ordre 2. Ces résultats confirment-ils la théorie ?

Annexe 1 : Approximation discrète d'une densité (loi d'une variable continue)

L'approximation discrète d'une densité notée $f(x)$ est appelée histogramme ou loi empirique car elle est construite à partir de l'analyse des données. Un histogramme est créé selon deux étapes.

Etape n°1 : définition des classes (ou bin), elles sont définies d'une part en fixant le nombre de classes, Nc et en considérant, une partitionnement uniforme de l'amplitude de la variable aléatoire. La longueur de chaque classe, $\Delta x = x_{i+1} - x_i$ peut être définie comme suit : $C_i \stackrel{\text{def}}{=} \left\{x \in \left[\left((i-1).\frac{(max_x - min_x)}{Nc}\right) \quad \left((i).\frac{(max_x - min_x)}{Nc}\right)\right]\right\}$ $\forall i \epsilon [1, Nc]$, ou selon la règle suivante : $C_i \stackrel{\text{def}}{=} \{x \in (|x - \bar{x}_i| < 3\sigma \ )\}$ $\forall i \epsilon [1, Nc]$ avec $\bar{x}_i$ le centre de la classe.



Etape n°2 : Pour chaque classe C_i, compter combien de fois la variable x tombe dans le segment $C_i = x_{i+1} - x_i$.

$$P_i = \frac{1}{N} \sum_{k=1}^{N-1} \delta \left[x(k) \in C_i\right]$$

Du fait de l'approximation, nous avons la relation suivante :

$$P_i = \int_{x_i}^{x_{i+1}} f(x)dx \approx f(x_i)\Delta x$$