TP5-TD Initiation aux probabilités pour le signal

Les probabilités (*a fortiori* la statistique) interviennent de plusieurs manières en analyse (ou synthèse du signal) ; en voici deux.

– Même si un signal analogique enregistré pendant une longue durée temporelle se présente comme un objet déterministe, le choix d'une fenêtre d'échantillonnage (de longueur inférieure à la fréquence d'échantillonnage en Hz) est, lui, un choix entaché d'aléatoire.
– Il arrive aussi bien sûr que le signal lui-même soit un signal aléatoire (par exemple les bruits poissonniens correspondant à des processus d'émission de particules en échographie ou médecine nucléaire).

Pour accompagner (et enrichir) cette feuille de TP 5, on pourra aussi se référer au cours de probabilités pour Math-Info en ligne :

`http://www.math.u-bordeaux1.fr/~yger/Proba6031.pdf`

On pourra aussi consulter (pour ce qui concerne les aspects stochastiques de la Théorie du Signal) le cours rédigé pour le master d'Ingénierie mathématique

`http://www.math.u-bordeaux1.fr/~yger/masterIG.pdf`

ou aussi les chapitres 10, 11, 12 de `Mathématiques Appliquées L3, Pearson,2009`.

Exercice 1 (Quelques lois de probabilité discrètes).

(1) Une variable aléatoire réelle X (prenant ses valeurs dans un sous-ensemble fini de cardinal $N+1$ que l'on peut identifier à $\{0,...,N\}$) suit une *loi binomiale* $\mathscr{B}(N,p)$ ($N \in \mathbb{N}^*$, $p \in [0,1]$) si

$$P(X = k) = \binom{N}{k} p^k (1-p)^{N-k}.$$

Si $N = 1$, la variable X est dite suivre une *loi de Bernoulli* de paramètre p.

– Simulez M réalisations (on prendra ici $M = 10000$) d'une épreuve dont le résultat est matérialisée par une VAR suivant une loi binomiale à l'aide de la routine
```
>> f= binornd(N,p,1,M) ;
>> x=0 : N ;
>> hist(f,x);
```
La dernière commande affiche l'histogramme de ces réalisations.
– L'espérance de X est la moyenne des valeurs prises (ici $0,...,N$) pondérées avec les ≪ poids ≫ avec lesquels ces valeurs sont prises. La variance est l'espérance de la VAR $(X - EX)^2$. Vérifiez que $EX = np$ et que $\operatorname{Var} X = np(1-p)$.

(2) Une variable aléatoire réelle X (prenant ses valeurs dans un ensemble infini dénombrable que l'on peut numéroter comme $\mathbb{N}$) suit une *loi de Poisson de paramètre* $\lambda > 0$ $\mathscr{P}(\lambda)$ si

$$P(X = k) = e^{-\lambda}\,\lambda^k/k!.$$

– Simulez M réalisations (on prendre encore ici $M = 10000$) d'une épreuve dont le résultat est matérialisé par une VAR suivant une loi de Poisson $\mathscr{P}(\lambda)$ à l'aide de la routine

```
>> f= poissrnd (lambda,1,M) ;
>> x= 0 : N ;
>> hist(f,x);
```

– Si $(p_k)_{k\geq 0}$ figure une distribution de probabilité sur $\mathbb{N}$, on appelle fonction génératrice de cette loi de VAR X la fonction

$$G_X : s \in [-1,1] \longmapsto E(s^X) = \sum_{k=0}^{\infty} p_k s^k.$$

Montrer que si X et X^2 sont d'espérance finie, on a $EX = G'_X(1)$ et $\mathrm{Var}(X) = G'_X(1) + G''_X(1) - (G'_X(1))^2$.

– Vérifiez que l'espérance d'une loi de Poisson $\mathscr{P}(\lambda)$ vaut λ ainsi que sa variance.

Exercice 2 (les signaux gaussiens). La loi $\mathscr{N}(\mu,\sigma^2)$ est la loi d'une VAR X cette fois continue et non discrète, ayant pour densité la fonction positive (d'intégrale 1 sur $\mathbb{R}$) :

$$t \in \mathbb{R} \longmapsto f_X(t) := \frac{1}{\sqrt{2\pi}\,\sigma}\exp\Big(-\frac{(t-\mu)^2}{2\sigma^2}\Big),$$

ce qui signifie que, pour tout $T \in \mathbb{R}$, on

$$F_X(T) := P(X \leq T) = \int_{-\infty}^{T} f_X(t)\,dt.$$

Cette fonction croissante F_X (tendant vers 0 si T tend vers $-\infty$ et 1 si T tend vers $+\infty$) est dite *fonction de répartition* de la VAR X. La densité f_x peut, elle, s'interpréter ainsi : si dt figure un accroissement infinitésimal du tempss t, on a

$$P(X \in [t, t+dt]) \simeq f_X(t)\,dt.$$

(1) En admettant que si X est une VAR continue de densité f_X, la moyenne (dite *espérance* $E[\varphi(X)]$) de $\varphi(X)$ (lorsque φ est une fonction de X) est définie par

$$E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(t)\,f_X(t)\,dt$$

(pourvu que l'intégrale impropre $\int_{-\infty}^{\infty} |\varphi(t)|\,f_X(t)\,dt$ converge), calculez espérance et variance d'une VAR X suivant une loi $\mathscr{N}(\mu,\sigma^2)$.

(2) Sous `MATLAB`, la routine

```
>> f= normrnd (mu, sigma, 1, M);
```

Générez 10^7 réalisations d'une loi gaussienne $\mathscr{N}(100p, 100p(1-p))$ pour diverses valeurs de p entre 0 et 1.

(3) Lorsque $N > 30$ et $\min(Np, N(1-p)) > 5$ (ou $N > 30$ et $Np(1-p) > 9$), la loi binomiale $\mathscr{B}(N,p)$ est approchée par la loi $\mathscr{N}(Np, Np(1-p))$. Validez ce résultat empiriquement en dressant des histogrammes comparés (sur 10^7 valeurs) de ces deux lois pour des valeurs de p entre 0 et 1 restant dans les normes où l'approximation est validée. Retrouvez la densité de la loi $\mathscr{N}(\mu, \sigma^2)$ en observant que si $dt = .1$ est le pas de l'histogramme utilisé pour la visualisation des résultats de la simulation de la loi gaussienne $\mathscr{N}(100p, 100p(1-p))$, on a

$$f_X(t) \simeq \frac{N_h(t)}{10^7} \times \frac{1}{.1} = \frac{N_h(t)}{10^6},$$

où $N_h(t)$ désigne le nombre de simulations rangées dans la colonne au dessus de base $[t, t+.1]$ dans cet histogramme. Faites de même avec l'histogramme des réalisations de la loi binomoiale $\mathscr{B}(N,p)$, en prenant cette fois un pas égal à 1.

(4) Générez sous `MATLAB` 10000 réalisations d'une variable aléatoire réelle s'écrivant

$$X = \frac{1}{100}(X_1 + \cdots + X_{100})$$

où $X_1, ..., X_{100}$ sont des VAR mutuellement indépendantes suivant toutes une même loi de Poisson [1] de même paramètre $\lambda = 30$, puis affichez l'histogramme de ces réalisations. Cette VAR X est un VAR que l'on appelle *moyenne empirique* (à l'ordre ici 100) de la VAR X_1 ; ceci est raisonnable car on vérifie immédiatement que $E[X] = E[X_1]$ (*cf.* plus loin) ; la moyenne empirique à l'ordre N deviendra, lorsque N augmente, un *estimateur* de la moyenne de X_1 (ceci résulte tu Théorème Limite Central, *cf.* plus loin). Que vaut l'espérance de la VAR X ? Que vaut sa variance ? (on rappelle que la variance d'une somme de VAR deux-à-deux indépendantes est égale à la somme des variances de ces VAR). Générez d'autre part 10000 réalisations d'une VAR suivant une loi $\mathscr{N}(30, .3)$ et affichez l'histogramme de ces réalisations. Comparez le avec l'histogramme des réalisations de X. Pourquoi ce résultat est-il conforme au *théorème limite central* qui assure que si $(X_k)_{k\in\mathbb{N}}$ est une suite de VAR mutuellement indépendantes et de même loi (d'espérance μ et de variance σ^2), la suite de VAR

$$\frac{\sqrt{N}}{\sigma}\Big(\frac{X_1 + \cdots + X_N}{N} - \mu\Big)$$

converge vers une VAR suivant une loi $\mathscr{N}(0,1)$.

(5) On reprend $X_1, ..., X_{100}$ comme à la question précédente (VAR mutuellement indépendantes et suivant toutes une loi de Poisson de paramètre 30) et

$$X = \frac{1}{100}(X_1 + \cdots + X_{100}).$$

Les VAR $X_1, ..., X_{100}$ et X sont-elles mutuellement indépendantes ? De combien de degrés de liberté dépendent les VAR que sont $(X_1 - X)^2$, ...,

1. Observez pour cela, en utilisant par exemple les fonctions génératrices (*cf.* la question 2) que la somme de 100 VAR indépendantes et suivant des lois de Poisson de paramètres respectifs $\lambda_1, ..., \lambda_{100}$ est une VAR suivant une loi de Poisson de paramètre $\sum_1^{100} \lambda_j$.

$(X_{100} - X)^2$? Générez 10000 réalisations de la VAR

$$\widetilde{X} := \frac{1}{99}\sum_{j=1}^{100}(X_j - X)^2.$$

Montrez que l'espérance de $\widetilde{X}$ est égale à la variance des X_j, soit ici 30. Générez 10000 réalisations de la VAR

$$\frac{10}{\sqrt{\widetilde{X}}}\,(X - 30).$$

La routine

```
>> f= trnd (99, 1, M) ;
```

génère M réalisations d'une VAR suivant une loi de Student à 99 degrés de liberté, c'est-à-dire s'écrivant commee le quotient d'une VAR suivant une loi $\mathscr{N}(0,1)$ par la racine carrée d'une somme

$$\frac{1}{99}\sum_{j=1}^{99} Y_j^2$$

où $Y_1, ..., Y_{99}$ sont des VAR mutuellement indépendantes suivant toutes aussi une loi $\mathscr{N}(0,1)$. Comparez l'histogramme d'une VAR suivant une loi de Student à 99 degrés de liberté avec celui de la VAR

$$\frac{10}{\sqrt{\widetilde{X}}}\,(X - 30)$$

obtenu précédemment.

Exercice 3 (Autorégressivité d'un signal, génération de signaux à partir de bruits blancs et de filtres `AR`).

(1) Soit $(s_k)_{k\in\mathbb{Z}}$ un signal digital (tous les $s(k)$ étant nuls sauf un nombre fini) que l'on suppose régi de manière approchée par une équation aux différences

$$s(k) \simeq \sum_{m=1}^{M-1} a_m\, s(k-m),$$

où $M \geq 2$ est supposé connu, mais les paramètres $a_1, ..., a_M$ inconus. Comment convient-t'il de choisir les $M-1$ paramètres $a_1, ..., a_{M-1}$ de manière à ce que la quantité

$$\sum_{k\in\mathbb{Z}} \Big| s(k) - \sum_{m=1}^{M-1} a_m\, s(k-m)\Big|^2$$

soit minimale ?

(2) On suppose maintenant que le signal digital $(s(k))_{k\in\mathbb{Z}}$ est régi de manière approchée par une équation aux différences :

$$s(k) - a_0 \simeq \sum_{m=1}^{M-1} a_m(s(k-m) - \alpha_0),$$

où $M \geq 2$ et $a_0, ..., a_{M-1}$ désignent M paramètres inconnus. Comment convient-il de choisir les M paramètres $a_0, ..., a_{M-1}$ de manière à ce que la quantité :

$$\sum_{k \in \mathbb{Z}} \Big| s(k) - a_0 - \sum_{m=1}^{M-1} a_m \left(s(k-m) - a_0\right)\Big|^2$$

soit minimale ?

(3) Étant donné un signal digital `s = [s(1),...,s(N)]`, écrivez une routine `MATLAB` `AutoReg1.m` qui, étant donné un entier `M` entre 2 et `N`, calcule les `M` paramètres $a_0, ..., a_{M-1}$ de manière à minimiser la quantité

$$\sum_{k=1}^{\mathtt{N-M}} \Big| \mathtt{s}(k+\mathtt{M}) - a_0 - \sum_{m=1}^{\mathtt{M}-1} a_m \, s(k+\mathtt{M}-m) - a_0)\Big|^2.$$

(4) En utilisant la routine `AutoReg1.m` précédente, écrivez une routine `AutoReg2.m` :

```
>> f= AutoReg2(M, pas, s, sigma) ;
```

qui calcule `M` paramètres d'autorégressivité $a_0, ..., a_{M-1}$ du signal digital (de longueur `N`) `s` et affiche le graphe de la fonction

$$\omega \in [0, 1/2] \longmapsto \frac{1}{\Big|1 - \sum\limits_{m=1}^{M-1} a_m \, e^{-2i\pi m\omega}\Big|^2 + \sigma^2}.$$

Si vous avez des difficultés, ces deux routines `AutoReg1` et `AutoReg2` sont dans le répertoire `Routines-feuilleTP5` dans `MATLABSignal`. Étudiez en la syntaxe et exploitez les. Testez le résultat pour des signaux générés avec la routine `testsignal`, que vous aurez préalablement téléchargé depuis le répertoire `test10.11` de

`http://www.math.u-bordeaux1.fr/~yger/MATLABSignal`

(5) Étudiez la syntaxe de la routine `welch.m` que vous aurez préalablement téléchargé sur le même site. Appliquez cette routine aux signaux `testA` et `testB` téléchargés (toujours depuis le répertoire `test10.11`), puis sur des signaux que vous aurez similé avec `signaltest`. Comparez avec les résultats obtenus à la question précédente.

(6) Simulez un bruit blanc avec la routine `rand` ou `randn`. Filtrez ce bruit blanc avec le filtre `AR` régi par la relation aux différences :

$$s(k) - \sum_{m=1}^{M-1} a_m \, s(k-m) = e(k) \quad \forall \, k \in \mathbb{Z}$$

en utilisant la commande `filter` (les coefficients $a_1, .., a_{M-1}$ étant ceux calculés à partir des signaux `testA`, `testB`, ou plus généralement de signaux digitaux générés par la routine `signaltest`. Qu'observez vous ?