

Utilisation du théorème de la limite centrale sur un signal expérimental

V. CROQUETTE

Janvier 2009

Résumé

Nous considérerons l'application du théorème de la limite centrale (TLC) à un signal expérimental. Nous rappellerons d'abord quelques notions sur les lois de probabilité et de leurs propriétés caractéristiques. En utilisant un signal expérimental nous montrerons que les conditions de validité du TLC ne sont pas toujours réunies. Nous discuterons du traitement que l'on peut appliquer dans ce cas pour retrouver les conditions de validité.

1 Distribution de probabilité

Une variable aléatoire $x_i(t)$ est caractérisée par sa distribution de probabilité, c'est-à-dire la probabilité $P(x)$ qu'elle prenne la valeur x . Si nous avons un signal aléatoire inconnu, on peut se faire une idée de sa distribution de probabilité est en faisant l'histogramme des valeurs. Cette méthode est évidemment la plus convaincante, cependant elle demande de disposer dans grand nombre d'échantillons indépendants. Souvent les points de mesures sont longs à obtenir et on n'en dispose pas d'un nombre suffisant pour établir un histogramme, on préfère alors calculer les moments de la variable aléatoire.

Le moment d'ordre n est défini par $\mu_n = \int_{-\infty}^{\infty} t^n x_i(t) dt$. Pour une série discrète l'intégrale devient une somme. Le moment d'ordre 1 est la valeur moyenne souvent noté μ , le moment d'ordre 2 est le carré de l'écart type σ^2 . Lorsque la symétrie $x, -x$ existe les moments impairs sont nuls (sauf le moment d'ordre 1).

Les signaux expérimentaux physiques présentent des distributions de probabilité particulières. La plus classique est évidemment la gaussienne, son origine est une conséquence du théorème de la limite centrale que nous discutons ci-dessous. Cependant un certain nombre de phénomènes physiques sont caractérisés par des distributions de type loi de Poisson.

1.1 La distribution gaussienne

Elle est caractérisée par $P(x) \frac{1}{\sigma\sqrt{2\pi}} = \exp - \left(\frac{(x-\mu)^2}{2\sigma^2} \right)$, elle est également appelée distribution normale : $\mathcal{N}(\mu, \sigma)$. Si nous excluons le moments d'ordre 1 (distribution centrée), nous observons que tous les moments impairs sont nuls et que tous les moments pairs peuvent s'exprimer en fonction de σ^2 . $\mu_{2k} = ((2k)!/2^k!) \sigma^{2k}$

Par exemple, le moment d'ordre 4 $\mu_4 = 3\mu_2 \cdot \mu_2 = 3 \cdot \sigma^4$. En vérifiant cette égalité on peut vérifier la nature gaussienne d'une distribution.

De façon générale on note $\gamma_1 = \mu_3/\sigma^3$ l'asymétrie d'une distribution (skewness) qui est nulle pour une gaussienne. De façon analogue on définit également l'aplatissement (kurtosis) $\beta_2 = \mu_4/\sigma^4$ qui vaut 3 pour une gaussienne.

1.2 La distribution de Poisson

Elle prend la forme générale $P_k(X) = \frac{(X)^k}{k!} e^{-X}$, le plus souvent en physique on considère le cas $k = 0$ et la variable X s'écrit plutôt $X = x/a$. La probabilité de distribution a alors une forme exponentielle $P(x) = a^{-1} e^{-x/a}$.

La propriété remarquable d'une distribution de Poisson est que la valeur moyenne est égale à l'écart type.

2 Théorème de la limite centrale

Il est très utile de simuler des signaux expérimentaux avec un ordinateur, les algorithmes utilisant des fonctions non-linéaires permettent de réaliser des fonction pseudo-aléatoires rapides et de bonne qualité. Dans leur version la plus simple

c'est fonctions génèrent un nombre entier aléatoire appartenant à un intervalle $\in [N_0, N_1]$. La distribution de probabilité de ces générateurs est ainsi une fonction rectange. Des algorithmes plus sophistiqués permettent de fabriquer n'importe quelle distribution de probabilité.

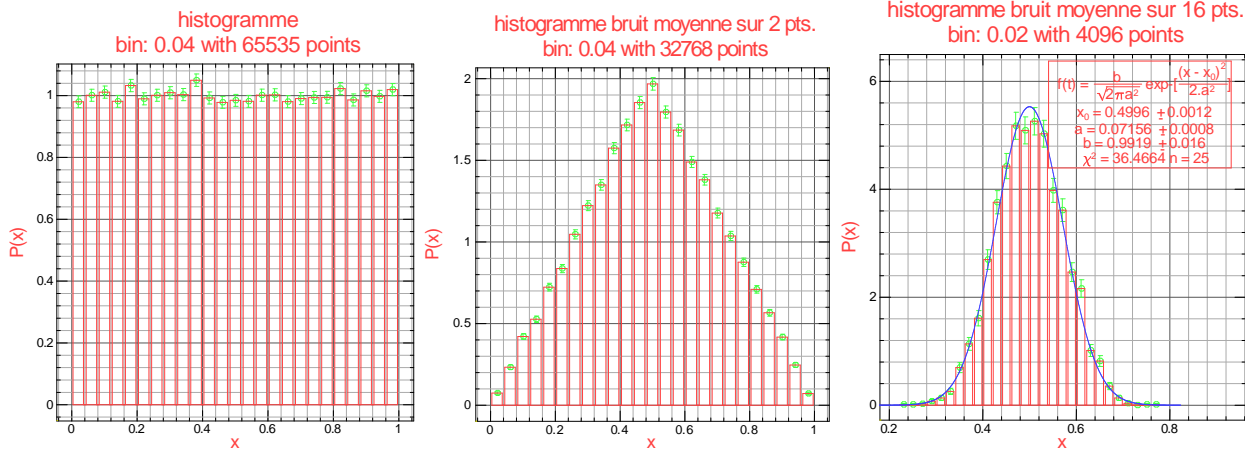


FIGURE 1 – Illustration du théorème centrale limite sur un signal aléatoire de distribution plate entre $[0, 1]$: A gauche, densité de probabilité du signal. Au centre densité de probabilité du même signal avec chaque point correspondant à la moyenne de deux points du signal original. A droite même chose avec une moyenne sur 16 points. L’histogramme de ce signal indique que la distribution des positions est presque gaussienne.

Le théorème nous dit que si X_1, X_2, \dots est une suite de variables aléatoire de même distribution de probabilité (μ, σ) , dans la mesure où ses variables sont indépendantes et où σ est défini et non nul ; alors $S_n = X_1 + X_2 + \dots + X_n$ est une variable aléatoire dont la distribution tend vers une gaussienne de valeur moyenne $n\mu$ et d’écart type $\sqrt{n}\sigma$.

La convergence de la densité de probabilité vers une gaussienne est une propriété remarquable puisqu’elle est vrai quelque soit la densité de probabilité de X_i comme l’illustre la figure 1.

3 Application à un signal expérimental

Le signal expérimental que nous utilisons est caractérisé par un bruit important. Il correspond ici à la mesure de la position $x(t)$ d’une bille magnétique de $1\mu m$ de diamètre évoluant dans l’eau et attachée à une seule molécule d’ADN de $16\mu m$ de longueur. Elle est tirée par une force magnétique de quelques pN suivant la direction z . Ce système se comporte comme un pendule amorti, le mouvement Brownien fait bouger la bille de façon aléatoire tandis que la molécule d’ADN ramène la bille vers sa position d’équilibre. Cependant, ce que nous allons discuter sur ce signal sera valide pour la plupart des signaux expérimentaux. Notre signal expérimental a une valeur moyenne de $21.53\mu m$ et un écart type de $0.34\mu m$.

Ce signal présente une distribution gaussienne car le mouvement brownien est le résultat du très grand nombre de chocs des molécules d’eau sur la bille, la distribution de probabilité de chacun de ces chocs n’est pas gaussienne, mais leur moyenne est piloté par le TLC et devient gaussien.

Nous nous intéresserons également à un signal simulé (Fig. 2) par un ordinateur présentant des caractéristiques analogues à notre signal expérimental généré en utilisant les fonctions aléatoires de distribution gaussienne de *numerical recipe*.

4 Moyenner pour diminuer le bruit

Pour réduire le bruit d’une expérience on moyenne le signal. Nous nous proposons de moyenner les points de nos signaux par paquets et de prédire l’écart type de ce ces séries.

Si $x_0(i)$ est le signal original avec $i \in [0, N]$ et $\sigma_0 = 0.34\mu m$, notre série moyennée sur k points s’écrit : $x_k(i) = \sum_{j=ki}^{k(i+1)} x_0(j)/k$ avec $i \in [0, N/k]$, son écart type est noté σ_k .

Le TLC nous dit que l’écart quadratique moyen décroît come \sqrt{k} , si $k = 64$ l’écart type doit être réduit d’un facteur 8.

La mesure de σ_{64} faite sur la série des 256 points de $x_{64}(i)$ donne les résultats suivants : Pour le signal expérimental $\sigma_{E64} = 0.249\mu m$, pour le signal simulé $\sigma_{S64} = 0.0434\mu m$.

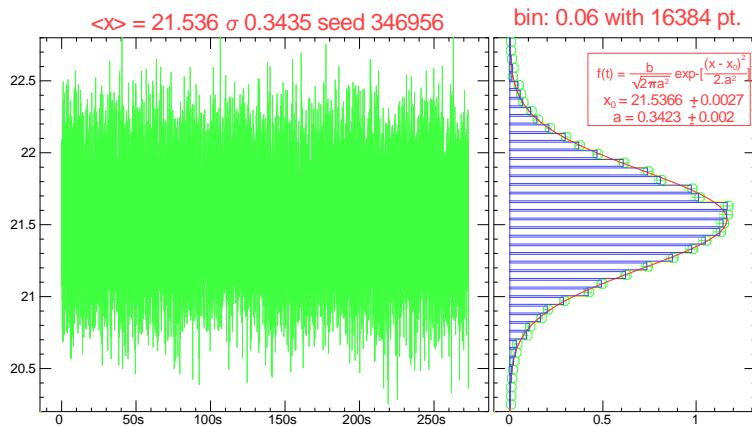


FIGURE 2 – A gauche, signal simulé ayant la même valeur moyenne et le même écart type que le signal expérimental. A droite, histogramme du signal de gauche indiquant que la distribution des positions est gaussienne.

Le signal simulé se comporte comme nous l’attendons ce qui n’est pas le cas pour notre signal expérimental. La raison c’est qu’une des hypothèse du TLC n’est pas valide dans le cas du signal expérimental.

5 Comportement de σ_k en fonction de k

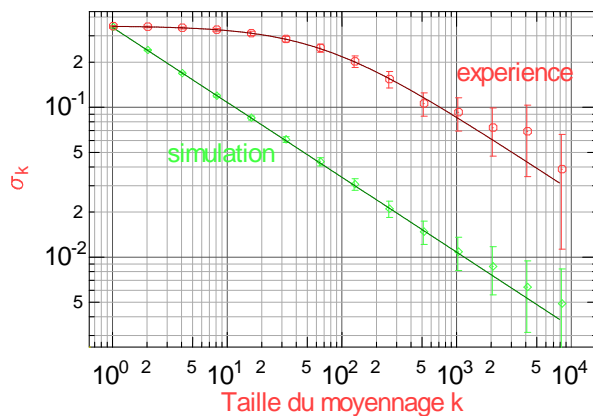


FIGURE 3 – Evolution de l’écart type en fonction du nombre de points sur lequel est fait le moyennage. Les points correspondent au calcul de l’écart type pour les signaux expérimentaux et simulés. Le trait plein ajusté aux points du signal simulé correspond à $\sigma_k = \sigma_0/\sqrt{k}$.

La figure 3 présente l’évolution de $\sigma_k(k)$ pour les deux types de signaux avec $k = 2^j$. Pour le signal simulé issu du générateur de nombres aléatoires, nous retrouvons le comportement prédit par le théorème de la limite centrale.

Au fur et à mesure que nous augmentons k , le nombre d’échantillons diminue, il en va de même de la précision de la mesure de l’écart type.

Le signal expérimental fait apparaître deux comportements distincts, pour $k < 64$ l’écart type ne varie que très peu avec k , par contre lorsque $k > 64$ l’écart type décroît comme le signal simulé c’est-à-dire avec \sqrt{k} .

Ce qui ne va pas dans notre signal expérimental c’est que les points de mesure ne sont pas indépendants ! c’est le cas presque dans toutes les expériences. Les points sont corrélés pendant un certain temps, ici $n_c = 64$ correspond à environ une seconde. Quand nous moyennons pendant un temps court, l’écart type ne change pas car le signal change peu durant le moyennage. Si nous moyennons plus d’une seconde, les échantillons ne sont plus corrélés, ils deviennent indépendants et le TLC s’applique.

la forme de la fonction qui ajuste les donnée est $\sigma_k = \frac{1}{\sqrt{1+k/n_c}}$.

L’origine physique du temps de corrélation est ici le temps caractéristique de retour à l’équilibre de la bille qui subit un frottement visqueux important.