Bruit et traitement du signal

Objectifs

- Explorer les méthodes statistiques pour calculer et évaluer un ensemble de données bruitées;
- Démontrer la relation entre la taille d'un échantillon et la précision de la mesure.

Matériel

Veuillez visionner la vidéo suivante pour une explication du matériel et de son installation: Expérience: Bruit _ HD 1080p *(10 min.)*.

Lors de cette expérience, il est **primordial que vous puissiez importer vos données sur votre ordinateur** afin de les analyser sur Python. Vous devrez donc avoir également:

- Une clé USB (si vous utilisez un oscilloscope pour prendre vos mesures);
- Un ordinateur avec un environnement Python fonctionnel équipé des librairies ***Numpy*** et ***Matplotlib***.

Contexte théorique

Un **signal bruité** est le résultat d'une mesure où la composante significative, appelée signal, est contaminée par des fluctuations aléatoires, ou bruit. Ces fluctuations, inévitables dans toute mesure expérimentale, proviennent de diverses sources : les limites des instruments de mesure (bruit instrumental), les perturbations environnementales (bruit de fond) ou des phénomènes fondamentaux tels que les fluctuations quantiques (bruit de grenaille, ou *shot noise* en anglais).

Un concept clé pour analyser et interpréter un signal bruité est **le rapport signal sur bruit** (Signal-to-Noise Ratio, SNR). Le SNR est une mesure quantitative qui compare la puissance du signal utile à celle du bruit. Plusieurs définitions du SNR existent, mais nous utiliserons la suivante basée sur la **moyenne** μ du signal et son **écart-type** σ**. Pour un signal discret [**$x_1, x_2, x_3, \dots x_N$**]:**

$$\mu = \frac{1}{N} \sum_{i=1}^{N} x_i \qquad \sigma^2 = \frac{1}{N} \sum_{i=1}^{N} (x_i - \mu)^2$$

$$SNR = \mu / \sigma.$$

Une valeur élevée de SNR signifie que le signal domine le bruit et est donc plus facile à détecter et à analyser. En revanche, un faible SNR indique que le bruit est comparable ou supérieur au signal, rendant l'interprétation des données plus difficile (voir fig.1).

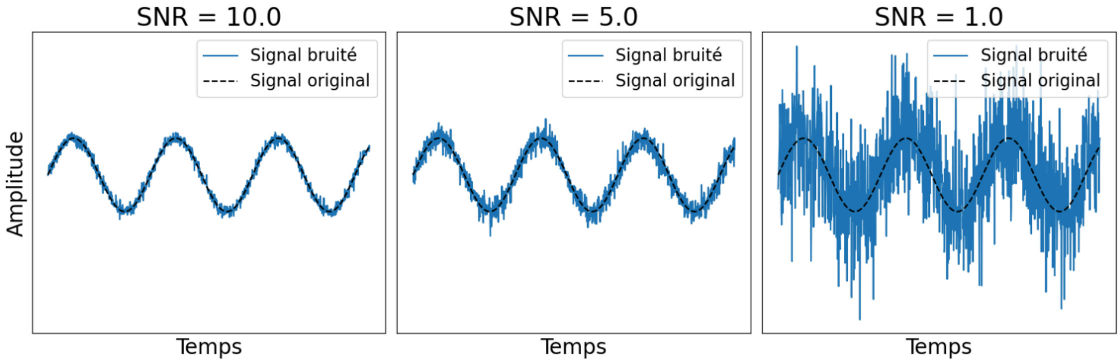

Fig.1 - Exemples de signal sinusoïdal avec différentes valeurs de SNR.

Par la nature aléatoire du bruit, toute mesure expérimentale peut être interprétée comme un échantillonnage d'une distribution aléatoire sous-jacente. La figure 2 illustre le lien qui existe entre un signal bruité et une distribution aléatoire. On observe qu'un signal à un SNR plus faible présente une distribution plus étendue.

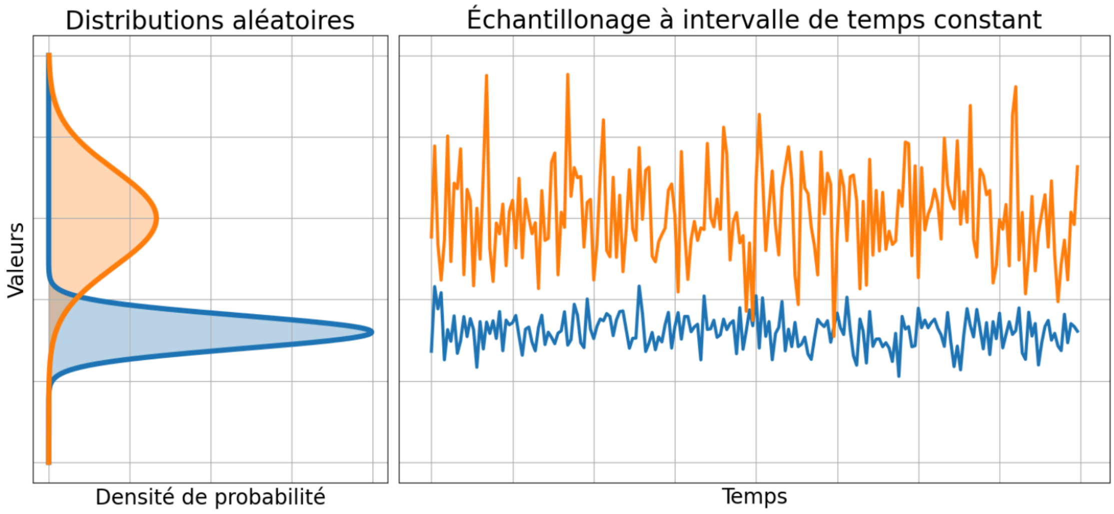

Fig.2 - Deux distributions normales (gauche) sont échantillonnées aléatoirement au fil du temps afin de générer des signaux temporels bruités (droite)

Chaque ensemble de mesures que nous réalisons ne représente qu'une approximation limitée de cette distribution. En effet, pour caractériser parfaitement cette distribution, il serait nécessaire de disposer d'un nombre infini de mesures. Cette limitation introduit une incertitude inévitable, car les paramètres estimés sont eux-mêmes influencés par la taille de l'échantillon (N). **L'erreur standard de la moyenne (SEM)** - qui définit l'erreur sur la mesure de la moyenne - se quantifie à partir de l'écart-type de la distribution aléatoire σ et du nombre de points N acquis:

$$SEM = \sigma / \sqrt{N}$$

Afin de comprendre intuitivement la relation entre l'erreur sur la moyenne et la taille de l'échantillon, la figure ci-dessous compare trois échantillonnages indépendants pour différentes valeurs de *N*. Pour chaque cas, l'histogramme orange représente les données mesurées, tandis que la courbe bleue montre la distribution réelle de la population dont la moyenne μ est indiquée en pointillé. On observe qu'à petite taille d'échantillon (*N*=2), les moyennes estimées varient fortement d'un tirage à l'autre, traduisant une plus grande incertitude statistique. Lorsque *N* augmente (*N=10* puis *N=50*), les histogrammes deviennent plus centrés autour de μ et la dispersion des moyennes estimées diminue.

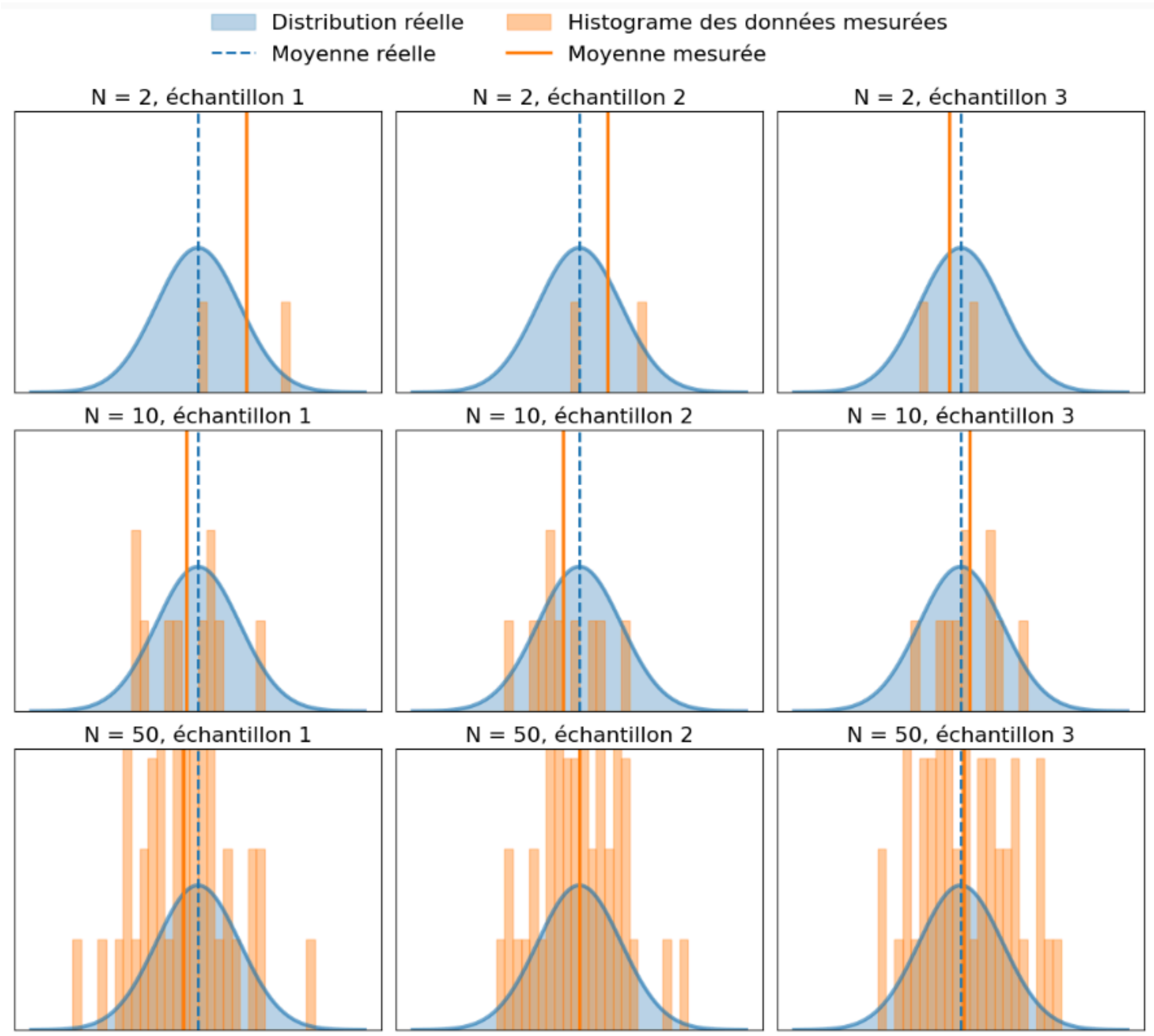

Fig.3 - Impact de la taille d'échantillon sur la précision de la moyenne. Trois tirages indépendants sont illustrés pour *N=2, 10, 50*. Pour les petits échantillons, les moyennes estimées fluctuent largement autour de la moyenne réelle, tandis qu'avec des échantillons plus grands, elles s'en rapprochent.

Le **théorème central limite (TCL)** est un résultat fondamental en probabilité et en statistique. Il stipule que, sous certaines conditions, la somme (ou la moyenne) d'un grand nombre de variables aléatoires indépendantes et identiquement distribuées tend vers une distribution normale, *quelle que soit la distribution initiale des variables*. La figure 4 illustre les densités de probabilité associées à un lancer d'un ou plusieurs dés et en calculant la somme des valeurs obtenues. La densité de probabilité d'un seul dé est uniforme: chaque résultat est également probable. Par contre, lorsque l'on lance plusieurs dés à la fois et que l'on calcule la somme, on observe que la densité de probabilité converge vers une distribution normale.

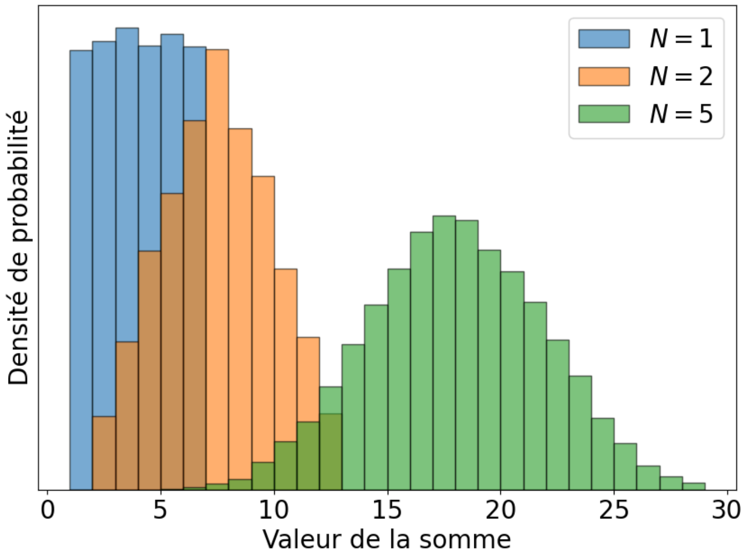

Fig.4 - Illustration de l'application du théorème central limite. Les distributions aléatoires illustrées ci-haut ont été obtenues en lançant N dés et en calculant la somme des valeurs obtenues.

Supposons une variable aléatoire X possédant une moyenne μ et un écart-type σ. La distribution aléatoire obtenue par la *somme* des $X_1, X_2, \ldots, X_N$ - que l'on note $\overline{X}$ - possède une moyenne $\overline{\mu}$ et un écart-type $\overline{\sigma}$ définit comme

$$\overline{\mu} = N\mu \qquad\qquad \overline{\sigma} = \sqrt{N}\sigma$$

Le TCL est à la base de nombreuses méthodes statistiques. Il garantit que les moyennes d'échantillons, même issues de distributions non normales, deviennent approximativement normales si la taille de l'échantillon est suffisamment grande. Cela justifie l'utilisation fréquente des modèles gaussiens en analyse de données expérimentales. En physique, cela explique également pourquoi de nombreux phénomènes mesurés suivent une distribution normale, même si les processus sous-jacents ne le sont pas nécessairement.

Manipulations

Dans ce laboratoire, vous travaillerez avec un montage qui simule une mesure bruitée. Dans cette simulation, on considère que vous avez un appareil qui capte une augmentation de signal par rapport au bruit de fond. Votre but est de quantifier cette augmentation de la manière la plus juste et la plus précise possible.

1. Installer le montage tel que présenté dans la vidéo suivante: Expérience: Bruit_HD 1080p.

2. Mesurer le signal émis et importer les données sur votre ordinateur.
    a. Quelle est la valeur du SNR du signal que vous observez?
    Ici, il est important que vous considériez le signal mesuré par rapport au bruit de fond.
    b. Caractériser la distribution des données obtenues.
    Astuce: Utiliser le signal de synchronisation afin de différencier les parties "signal haut" et "bruit de fond" de vos données. La fonction numpy.where() pourrait vous être utile.

3. Évaluer l'intensité de l'augmentation du signal par rapport au bruit de fond. Considérez adéquatement vos incertitudes sur la mesure. Tentez d'obtenir un résultat le plus précis possible.

4. Démontrer expérimentalement la formule de l'erreur standard de la moyenne (SEM).
    À ce stade-ci, vous devriez avoir sous la main un très grand nombre de données. Si ce n'est pas déjà le cas, reprenez un jeu de données de la plus grande taille possible. ***Utilisez uniquement les données lorsque le signal est en mode "haut".***
    a. Choisir au hasard deux valeurs de votre jeu de données et calculer la moyenne de ces points. Répéter un grand nombre de fois ce processus.
    Astuce: La fonction numpy.random.choice() pourrait vous être utile et vous permettrait de réaliser ce processus un grand nombre de fois.
    b. Caractériser la distribution des moyennes obtenues de cette manière. Évaluer la variation de cette distribution.
    c. Augmenter graduellement le nombre de points utilisé pour calculer la moyenne et tracer la variation en fonction du nombre de points N. *Est-ce que la relation erreur vs. N correspond à ce à quoi vous vous attendiez? Démontrez-le quantitativement*

5. À partir de la technique de votre choix, trouver une manière d'améliorer le plus possible le SNR du signal mesuré.
    a. Évaluer de manière quantitative l'impact de votre approche sur la qualité du signal.
    b. Évaluer la différence d'intensité du signal en considérant les incertitudes sur vos mesures. Est-ce que ce résultat diffère de celui que vous avez obtenu en 2c?