

Statistique

Ce chapitre a pour but d'introduire quelques notions de statistiques. L'objectif de la statistique est de déterminer les lois que suivent une variable aléatoire à partir d'un échantillon de mesure. Nous aborderons tout d'abord les principales lois de probabilité que l'on peut rencontrer en physique. Dans une deuxième partie, nous parlerons des incertitudes en physique et en particulier comment calculer ces incertitudes. Nous aborderons un point essentiel qui est celui de l'estimation des paramètres d'une loi à partir d'un échantillon. Nous appliquerons ces notions à l'ajustement des données (fit).

1 Variables aléatoires

Une variable aléatoire et une grandeur physique dont la valeur dépend d'un certain nombre d'événements. L'origine de ce hasard peut être fondamental (par exemple, le nombre de désintégration en une seconde d'un échantillon radioactif) ou lié à des paramètres que l'on ne connaît pas (quelle sera la température à Tokyo le 15 juillet 2045?).

Dans la suite de ce cours, on notera X une variable aléatoire. Nous ne considérerons dans un premier temps que des variables aléatoires à valeur réelle. Elles seront soit discrètes entières soit continues. Nous noterons par k la valeur d'une v.a. discrète et par x la valeur d'une v.a. continue. Nous verrons ensuite le cas de vecteurs aléatoires.

1.1 Variables aléatoires discrètes

Une variable aléatoire discrète est caractérisée par sa fonction de masse qui est la probabilité que la v.a. vaut k :

$$\mathcal{P}(X = k) \tag{1}$$

Plusieurs lois de probabilités sont rencontrées en physique

1.1.1 Loi de Bernoulli

La loi de Bernoulli est la loi qui dépend du paramètre $p \in [0,1]$ pour laquelle on a $\mathcal{P}(X = 1) = p$, $\mathcal{P}(X = 0) = 1 - p$ et $\mathcal{P}(X = k) = 0$ sinon.

C'est par exemple la loi qui apparaît en mécanique quantique. Si un atome est dans une superposition de deux états $|\psi\rangle = \alpha|0\rangle + \beta|1\rangle$ (avec $|\alpha|^2 + |\beta|^2 = 1$), alors la probabilité de mesurer l'atome dans l'état $|1\rangle$ suit une loi de Bernoulli de paramètre $p = |\beta|^2$.

1.1.2 Loi binomiale

C'est la loi obtenue en additionnant plusieurs v.a. qui suivent une loi de Bernoulli. C'est à dire que si X_1, \dots, X_n sont des v.a. qui suivent la loi de Bernoulli de paramètres p , alors $X = X_1 + \dots + X_n$ suit la loi binomiale de paramètre p et n .

On peut montrer facilement que :

$$\mathcal{P}(X = k) = C_n^k p^k (1 - p)^{n-k} \quad (2)$$

En reprenant l'exemple d'un atome dans une superposition de deux états, la loi binomiale apparaît lorsque l'on fait la mesure sur un nombre n connu d'atomes.

1.1.3 Loi de Poisson

Cette loi dépend d'un paramètre λ . Elle a pour fonction de masse :

$$\mathcal{P}(X = k) = \frac{\lambda^k}{k!} e^{-\lambda} \quad (3)$$

Elle est très présente en physique. Lorsque l'on a une source contenant de façon aléatoire un grand nombre d'émetteur de particule (source radioactive, source de lumière), la statistique d'émission de ces particules suit en général un processus Poissonien. Lorsque l'on compte le nombre de particules sur un temps donné, alors la variable aléatoire suit une loi Poissonienne.

Une raison de la présence de la loi Poissonienne est sa stabilité : si X_1 et X_2 sont deux v.a. indépendantes qui suivent une Poissonienne de paramètre λ_1 et λ_2 alors la v.a. $X = X_1 + X_2$ suit une distribution Poissonienne de paramètre $\lambda_1 + \lambda_2$. Réciproquement, si on atténue une v.a. de loi de Poisson, alors le nouveau processus sera lui aussi Poissonien.

1.2 Variable aléatoire continue

Une variable aléatoire continue est caractérisée par sa densité de probabilité noté $f(x)$ définie de telle sorte que :

$$\forall \alpha, \beta \in \mathbb{R}, \mathcal{P}(x \in [\alpha, \beta]) = \int_{\alpha}^{\beta} f(x) dx \quad (4)$$

Il existe un grand nombre de distributions continues. Nous allons ici en présenter deux : la loi uniforme sur un intervalle et la distribution normale (gaussienne).

1.2.1 Distribution uniforme

Elle dépend de deux paramètres, une valeur centrale μ et une largeur L . Elle est déterminée par la densité $f(x) = 1/L$ si $x \in [\mu - L/2, \mu + L/2]$ et 0 sinon.

1.2.2 Distribution gaussienne

Elle dépend d'une valeur centrale μ et d'une dispersion σ . On a :

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad (5)$$

1.3 Caractéristique d'une v.a.

Nous considérerons ici deux grandeurs pour caractériser une variable aléatoire, son espérance et son écart type.

1.3.1 Espérance

Elle peut se définir comme la moyenne d'un nombre infini de réalisation d'une variable aléatoire. Dans le cas d'une distribution discrète elle vaut :

$$E(X) = \sum_k kP(X = k) \quad (6)$$

et pour une variable continue :

$$E(X) = \int_{-\infty}^{\infty} xf(x)dx \quad (7)$$

Il est possible que l'espérance d'une v.a. ne soit pas définie lorsque la somme ou l'intégrale ne converge pas.

1.3.2 Variance et écart type

La variance est définie comme l'espérance de la variable aléatoire $(X - E(X))^2$:

$$\text{Var}(X) = E((X - E(X))^2) \quad (8)$$

L'écart type se définit comme la racine de la variance.

Notons que l'on distingue sémantiquement la notion d'espérance d'une variable aléatoire de la moyenne d'un échantillon de cette variable (lequel tend vers l'espérance pour un grand nombre d'échantillon). On utilise le même mot pour désigner la variance d'une v.a. et la variance d'un échantillon statistique ($\langle (x - \bar{x})^2 \rangle$).

1.3.3 Théorème central limite

On considère N variables aléatoires X_i qui suivent une même loi et on note μ l'espérance de cette loi et σ^2 sa variance. On note Y la variable aléatoire $Y = \sum_i X_i$. Le théorème central limite nous dit que Y tend vers une distribution normale de paramètre $N\mu$ et de variance $N\sigma^2$.

Par exemple, une loi de poisson de paramètre $N\lambda$ va tendre vers une loi normale d'écart type $\mu = N\lambda$ et d'écart type $\sqrt{N\lambda}$ lorsque $N \rightarrow \infty$. De même la loi binomiale s'approche d'une gaussienne lorsque $n \rightarrow \infty$.

En physique, le bruit est souvent la somme d'une multitude de bruit que l'on ne connaît pas. C'est pour cela qu'il est souvent décrit par une loi normale.

1.4 Composition des incertitudes

L'incertitude d'une mesure quantifie la dispersion ou l'écart-type que l'on a estimé sur la mesure qui a été faite. Souvent, une grandeur est calculée à partir de plusieurs grandeurs elles même entachées d'une incertitude. Il est alors possible de calculer l'incertitude de la grandeur calculée.

On définit la covariance de deux v.a. X et Y de la façon suivante :

$$\text{Covar}(X, Y) = E((X - E(X))(Y - E(Y))) \quad (9)$$

On a donc la relation $\text{Var}(X) = \text{Covar}(X, X)$. La covariance est une fonction bilinéaire. On en déduit donc que

$$\text{Var}(\lambda_1 X_1 + \dots + \lambda_N X_N) = \sum_{i,j} \lambda_i \lambda_j \text{Covar}(X_i, X_j) \quad (10)$$

Le plus souvent, les variables sont indépendantes, c'est à dire que $\text{Covar}(X_i, X_j) = 0$ si $i \neq j$, on a donc la relation :

$$\text{Var}(\lambda_1 X_1 + \dots + \lambda_N X_N) = \sum_i \lambda_i^2 \text{Var}(X_i) \quad (11)$$

Dans le cas général où l'on a $X = f(X_1, \dots, X_N)$, on linéarise f autour de l'espérance des X_i , ce qui revient à remplacer les λ_i par les dérivées partielles de f :

$$\text{Var}(f(X_1, \dots, X_N)) = \sum_{i,j} \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} \text{Covar}(X_i, X_j) \quad (12)$$

Remarque : le package `uncertainties` de Python permet de gérer automatiquement les incertitudes. Ce package peut calculer automatiquement les incertitudes en tenant compte des corrélations entre variables.

```
1 import uncertainties
2 from uncertainties.umath import sin, cos
3
4 x = uncertainties.ufloat(2.3, 0.1)
5 y = uncertainties.ufloat(1.3, 0.2)
6
7 z = x+y
8 print(z) # 3.60+/-0.22
9 z1 = sin(x)
10 z2 = cos(x)
11 print(z1, z2) # 0.75+/-0.07 -0.67+/-0.07
12 print(z1**2 + z2**2) # 1.0+/-0
```

2 Statistique

2.1 Estimateur

On considère une variable aléatoire X dont le type de la loi est connue mais dont on ne connaît pas les paramètres. On note $\theta = \theta_1, \dots, \theta_p$ les p paramètres de la loi.

On réalise N mesures indépendantes, x_1, \dots, x_N : c'est l'échantillon statistique. Notre objectif est d'estimer la valeur de θ en fonction de ces N mesures : c'est ce que l'on nomme l'estimateur.

Un estimateur est une fonction de \mathbb{R}^N dans \mathbb{R}^p . Cette fonction pourra avoir différentes formes : une fonction analytique ou un algorithme ! On la notera $\hat{\theta}(x_1, \dots, x_N)$.

L'estimateur $\hat{\theta}(X_1, \dots, X_N)$ est une variable aléatoire. On pourra donc calculer son espérance ou son écart type.

2.2 Exemple

On prend une loi uniforme déterminée par la densité $f(x) = 1/L$ si $x \in [\mu - L/2, \mu + L/2]$ et 0 sinon. On souhaite estimer μ à partir d'un échantillon x_1, \dots, x_N . Voici trois estimateurs possibles :

- $\hat{\mu}_1(x_1, \dots, x_N) = x_1$
- $\hat{\mu}_2(x_1, \dots, x_N) = \frac{\sum x_i}{N}$
- $\hat{\mu}_3(x_1, \dots, x_N) = \frac{\max(x_i) + \min(x_i)}{2}$

2.3 Caractéristique d'un estimateur

— Un estimateur est dit non biaisé si

$$E(\hat{\theta}(X_1, \dots, X_N)) = \theta \quad (13)$$

Les trois estimateurs ci-dessus sont non biaisés.

L'incertitude statistique d'un estimateur non biaisé est donné par son erreur quadratique :

$$\sigma_{\hat{\theta}} = \sqrt{E((\hat{\theta} - \theta)^2)} \quad (14)$$

— Un estimateur est convergent si

$$\hat{\theta}(X_1, \dots, X_N) \rightarrow_{N \rightarrow \infty} \theta \quad (15)$$

Il s'agit d'une convergence de loi de probabilité – notion qui n'est pas universelle. Par exemple on peut prendre l'erreur quadratique : un estimateur est convergent si $E((\hat{\theta}_N - \theta)^2) \rightarrow_{N \rightarrow \infty} 0$.

Seul les deux derniers estimateurs sont convergents.

La vitesse de convergence peut se calculer à partir de l'incertitude statistique de l'estimateur. Le dernier estimateur (μ_3) converge plus rapidement.

2.4 Comment trouver un estimateur

Un estimateur couramment utilisé est le maximum de vraisemblance. A partir d'un échantillon, on peut appliquer la loi de Bayes et calculer la probabilité d'avoir un paramètre θ :

$$\mathcal{P}(\theta | x_1, \dots, x_N) = \frac{\mathcal{P}(x_1, \dots, x_N | \theta) \mathcal{P}(\theta)}{\mathcal{P}(x_1, \dots, x_N)} \quad (16)$$

Dans cette expression, le seul terme que l'on connaît est $\mathcal{P}(x_1, \dots, x_N | \theta)$ (c'est la loi de probabilité). En général on ne connaît pas $P(\theta)$, on suppose que c'est uniforme et indépendant de θ . Quand au terme $\mathcal{P}(x_1, \dots, x_N)$, il est indépendant de θ (coef de normalisation).

Sous couvert de l'hypothèse que $P(\theta)$ est indépendant de θ , on obtient donc la loi de distribution de θ . Un estimateur peut être obtenu à partir du maximum de cette probabilité. C'est le maximum de vraisemblance.

On introduit la vraisemblance :

$$\mathcal{L}(\theta) = \mathcal{P}(x_1, \dots, x_N | \theta) \quad (17)$$

et le maximum de vraisemblance

$$\hat{\theta} = \arg \max \mathcal{L}(\theta). \quad (18)$$

Dans le cas où les mesures sont indépendantes, on a : $P(x_1, \dots, x_N | \theta) = \prod_i \mathcal{P}(x_i | \theta)$. En général, on maximise alors $\sum_i \ln(\mathcal{P}(x_i | \theta))$.

Exemple de la loi normale : les X_i suivent une loi normale de moyenne μ et d'écart type σ . Calculez analytiquement le maximum de vraisemblance pour μ et σ . L'estimateur est-il biaisé ? Est-il convergent ?

2.5 Méthode Monte-Carlo

L'outil numérique peut être indispensable pour d'une part implémenter un estimateur et aussi le caractériser.

Pour implémenter le maximum de vraisemblance, on pourra utiliser les fonctions d'optimisation du module `scipy.optimize`.

Pour caractériser un estimateur, on utilise en générale la méthode Monte-Carlo : avec des paramètres connus, on va simuler un jeu N de donnée sur lequel on va ensuite évaluer l'estimateur. Cette simulation sera répétée un grand nombre M de fois. On pourra alors estimer l'espérance et l'incertitude de notre estimateur à partir de la moyenne et l'écart type du résultat des simulations.

3 Ajustement

3.1 Variables aléatoires multidimensionnelles

Vecteur aléatoire

On note :

$$\vec{X} = \begin{pmatrix} X_1 \\ X_2 \\ \vdots \\ X_n \end{pmatrix} \quad (19)$$

La distribution de probabilité est une mesure sur \mathbb{R}^n :

$$\mathcal{P}(\vec{X} \in D) = \int_D f(\vec{x}) d\vec{x} \quad (20)$$

On peut définir l'espérance :

$$E(\vec{X}) = \int_{\mathbb{R}^n} \vec{x} f(\vec{x}) d\vec{x} \quad (21)$$

On a bien évidemment : $E(\vec{X})_i = E(X_i)$.

On définit aussi la matrice de variance :

$$\text{Var}(\vec{X}) = E\left((\vec{X} - E(\vec{X})) \cdot^t (\vec{X} - E(\vec{X}))\right) \quad (22)$$

Changement de variable

On considère la variable scalaire $Y = aX + b$. Que vaut f_Y en fonction de f_X . Il faut écrire $f_X(x)dx = f_Y(y)dy$ avec $dy = a dx$ on en déduit que :

$$f_Y(y) = \frac{1}{a} f_X\left(\frac{y-b}{a}\right) \quad (23)$$

Dans le cas d'un vecteur aléatoire, $\vec{Y} = A\vec{X} + \vec{B}$, on obtient une formule similaire :

$$f_Y(\vec{y}) = \frac{f_X(A^{-1}(\vec{y} - \vec{B}))}{|\det A|} \quad (24)$$

Loi normale multidimensionnelle

Elle est définie par :

$$f(\vec{x}) \propto e^{-\vec{x} M \vec{x}} \quad (25)$$

où M est une matrice symétrique définie positive.

On peut alors montrer que :

$$\text{Var}(\vec{X}) = \text{Covar}(X_i, X_j) = \frac{1}{2M} \quad (26)$$

Cette propriété est triviale si M est diagonale (cas de variables indépendante). On se ramène à ce cas en diagonalisant M , c'est à dire en créant une combinaison linéaire de variables initiales qui soit indépendantes : $Y = PX$ et tPMP diagonale.

3.2 Ajustement

Qu'est ce qu'un ajustement (fit)

On mesure un signal y en fonction d'un paramètre. En général ce paramètre est une variable réelle x , mais rien n'empêche d'avoir un paramètre pris dans une ensemble quelconque (par exemple pixel d'un image). Le signal que l'on mesure dépend à la fois de ce paramètre x (connu), mais aussi de paramètres \vec{p} que l'on ne connaît pas. Dans le cas le plus courant, chaque mesure y_i est donnée par une variable aléatoire Y_i qui suit une loi normale :

$$y_i = f(x_i, \vec{p}) + \epsilon_i \quad (27)$$

où ϵ_i suit une loi gaussienne de variance σ (indépendant de i) et centrée en 0. La généralisation aux cas où les σ dépendent de i sera immédiate.

L'objectif est d'obtenir une estimation des paramètres \vec{p} .

Fit des moindres carrés

On peut appliquer le principe du maximum de vraisemblance pour estimer le paramètre \vec{p} .

La vraisemblance est donnée par :

$$\mathcal{L}(\vec{p}) = e^{-\frac{1}{2\sigma^2} \sum_i (y_i - f(x_i, \vec{p}))^2} \quad (28)$$

La maximum de vraisemblance est donc obtenu en minimisant $\sum_i (y_i - f(x_i, \vec{p}))^2$. C'est le principe des moindres carrés.

On peut aussi obtenir la covariance des paramètres : soit \vec{p}_{opt} les paramètres optimaux. On peut développer la vraisemblance autour de \vec{p}_{opt} et obtenir

$$\mathcal{L}(\vec{p}_{\text{opt}} + \vec{dp}) = \mathcal{L}(\vec{p}_{\text{opt}}) e^{-\frac{1}{2\sigma^2} \sum_i (\vec{\nabla} f(x_i) \cdot \vec{dp})^2} \quad (29)$$

On peut estimer σ par le résiduel du fit. En introduisant la matrice M :

$$M_{k,l} = \sum_i \left. \frac{\partial f}{\partial p_k} \right|_{x_i} \left. \frac{\partial f}{\partial p_l} \right|_{x_i} \quad (30)$$

on obtient que

$$\text{Covar}(p_i, p_j) = \frac{\sigma^2}{2M} \Big|_{i,j} \quad (31)$$

La valeur de σ sera estimée à partir du résiduel du fit. C'est ce que fait la fonction `curve_fit` de python.

Estimation de l'erreur d'un fit

Dans le cas où les paramètres sont indépendants, M est diagonale. L'incertitude sur le paramètre i est donc donnée par $\sigma_i^2 = \frac{\sigma^2}{2M_{i,i}}$.

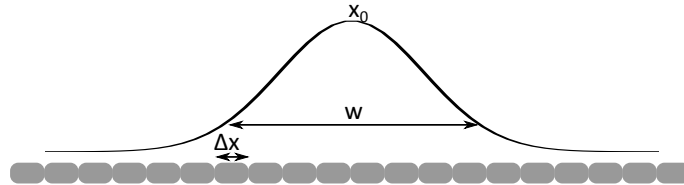
Si on a effectué un nombre N de mesures avec des paramètres x_i repartis uniformément entre a et b , alors on peut approximer $M_{i,i}$ par

$$M_{i,i} \simeq N \frac{\int_a^b \left(\frac{\partial f}{\partial p_i} \right)^2 dx}{b - a} \quad (32)$$

Exemple : une camera enregistre l'intensité d'un faisceau Gaussien de "waist" w et centré en x_0 . Cette intensité est donnée par :

$$y(x) = e^{-\frac{2(x-x_0)^2}{w^2}} \quad (33)$$

On note Δx la taille de chaque pixel. Le rapport entre le bruit de chaque pixel et le maximum du signal est noté σ . L'objectif est de déterminer la position x_0 . Calculer l'ordre de grandeur de l'incertitude sur la mesure de x_0 .



Fit en Python

On peut utiliser la fonction `curve_fit` du package `scipy.optimize`. Cette fonction permet aussi d'obtenir l'incertitude des paramètres sous forme d'une matrice de corrélation. Cette fonction s'utilise de la façon suivante :

```
p_opt, cor_mat = curve_fit(fonction_de_fit, data_x, data_y, p_ini)
```

où

- `fonction_de_fit(x, p1, p2, ..., pn)` est la fonction de fit. Les variables `p1`, ..., `pn` sont les paramètres de la fonction de fit.
- `data_x` et `data_y` sont les points de mesure.
- `p_ini` sont les paramètres initiaux (sous forme d'une liste/tuple/array).
- `p_opt` seront les paramètres optimaux et `cor_mat` est la matrice de corrélation entre les paramètres.

Il est aussi possible d'utiliser le package `uncertainties` pour créer des variables corrélés : `p_opt_uncert = uncertainties.correlated_values(p_opt, cor_mat)`