TP n°1 : Entropies discrète et continue, information mutuelle

Aarab Wassim - Dlimi Mohammed - Ettaki Mohammed Amine

Décembre 2023

Table des matières

1 Lien entre entropie discrète et continue

X est une variable aléatoire de densité f_X continue.
$\forall \Delta > 0$, $X_\Delta = \sum_{i \in \mathbb{Z}} x_i \mathbb{1}_{[i\Delta,(i+1)\Delta]}(X)$ où $x_i \in [i\Delta, (i+1)\Delta]$.

$$1 - \quad \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f_X(x)\,dx = \frac{1}{\Delta}(F_X((i+1)\Delta) - F_X(i\Delta)) = \frac{F_X((i+1)\Delta) - F_X(i\Delta)}{(i+1)\Delta - i\Delta}$$

D'après le théorème des accroissements finis appliqués sur F_X sur $[i\Delta, (i+1)\Delta]$:

$$\forall i \in \mathbb{Z}, \exists x_i \in]i\Delta, (i+1)\Delta[, \quad f_X(x_i) = \frac{1}{\Delta} \int_{i\Delta}^{(i+1)\Delta} f_X(x)\,dx$$

$$2 - \quad P(X_\Delta = x_i) = P(i\Delta \leq X \leq (i+1)\Delta) = \int_{i\Delta}^{(i+1)\Delta} f_X(x)\,dx = \Delta f_X(x_i)$$

$$\begin{aligned} 3 - \quad H(X_\Delta) &= E(-\log(P_X \Delta)) \\ &= \sum_{i \in \mathbb{Z}} -P(X_\Delta = x_i) \log(P(X_\Delta = x_i)) \\ &= -\sum_{i \in \mathbb{Z}} \Delta f_X(x_i) \log(\Delta f_X(x_i)) \\ &= -\Delta \left(\sum_{i \in \mathbb{Z}} f_X(x_i) \log(f_X(x_i)) + \log(\Delta) \sum_{i \in \mathbb{Z}} f_X(x_i) \right) \\ &= -\Delta \sum_{i \in \mathbb{Z}} f_X(x_i) \log(f_X(x_i)) - \log(\Delta) \end{aligned}$$

2 Loi gaussienne

2.1 Loi gaussienne univariée.

on considère maintenant une variable aléatoire réelle X qui suit une loi gaussienne de moyenne μ et de variance σ^2 (ie. $X \sim \mathcal{N}(\mu, \sigma^2)$), on veut maintenant vérifier numériquement le résultat théorique obtenu précédemment pour une variable qui suit une loi gaussienne. Il suffit donc de considérer un nombre très grand de réalisations différentes de X notées x_i pour qu'on puisse définir X_Δ avec un Δ très petit, car plus le nombre des x_i qu'on choisit est grand plus que la distance entre ses éléments qui est égale à Δ est petite. On choisit par exemple $n = 10000$ réalisations. Pour vérifier maintenant numériquement que la densité de cette loi est bien $f_X = \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)$, on trace l'histogramme de ces n réalisations x_i d'une aire normalisée à 1 et la densité f_X dans une même figure. On trouve la figure ci-dessous.

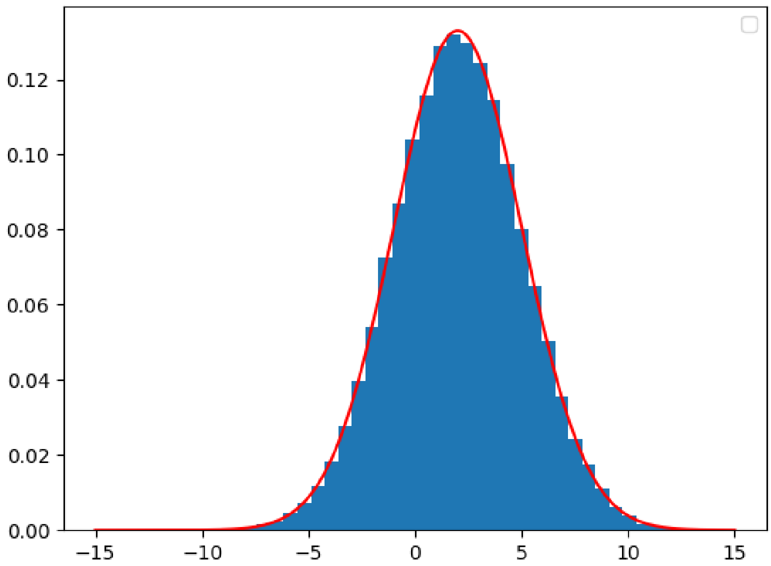

FIGURE 1 – histogramme-densité

ceci prouve que la densité de X est bien f_X puisque les deux graphes sont compatibles,il nous reste maintenant à vérifier le résultat$\mathbb{H}(X_\Delta) + \log(\Delta) \underset{\Delta \to 0}{\to} \mathbb{H}(X)$,pour se faire,on calcule tout d'abord le membre gauche de la relation à l'aide d'un code python avec $\Delta = 1/n$ qui tend vers 0 fourni en annexe,et en testant on trouve la valeur 0.21314555932018384,ensuite en calcule théoriquement $\mathbb{H}(X)$ comme ci dessous :

$$\begin{aligned}
\mathbb{H}(X) &= -\int_{-\infty}^{\infty} f_X \log(f_X(x))\, dx \\
&= -\int_{-\infty}^{\infty} \frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \log\left(\frac{1}{\sigma\sqrt{2\pi}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right) dx \\
&= \int_{-\infty}^{\infty} \left(\frac{\log(3\sqrt{2\pi})}{3\sqrt{2\pi}} \exp\left(-\frac{(x-2)^2}{18}\right) + \frac{1}{3\sqrt{2\pi}}\left(\frac{(x-2)^2}{18}\right) \exp\left(-\frac{(x-2)^2}{18}\right)\right) dx
\end{aligned}$$

et on a d'après les intégrales de Gauss : $\int_{-\infty}^{\infty} \exp -x^2 = \sqrt{\pi}$ et $\int_{\infty}^{\infty} x^2 \exp(-x^2) = \frac{\sqrt{\pi}}{2}$ et donc par conséquent :

$$\mathbb{H}(X) = \frac{1}{2} + \log(3\sqrt{2\pi})$$
$$\mathbb{H}(X) = 2.51755$$

cette valeur est très loin de la valeur obtenue numériquement,la relation n'est pas donc satisfaite pour une loi gaussièene univariée

2.2 Loi gaussienne multivariée.

Soit $\mu \in \mathbb{R}^p$ et $X \in \mathcal{M}_p(\mathbb{R})$, pour générer des réalisations de X, le processus se déroule en deux étapes :

1. Dans la première étape, la fonction `np.random.randn` est utilisée pour créer une matrice Z où chaque ligne représente un échantillon indépendant d'une loi normale univariée $N(0,1)$.

$$Z = \begin{bmatrix} Z_{1,1} & Z_{1,2} & \dots & Z_{1,p} \\ Z_{2,1} & Z_{2,2} & \dots & Z_{2,p} \\ \vdots & \vdots & \ddots & \vdots \\ Z_{n,1} & Z_{n,2} & \dots & Z_{n,p} \end{bmatrix}$$

2. Dans la deuxième étape, une transformation linéaire est appliquée à ces échantillons en utilisant la décomposition de Cholesky de la matrice de covariance R. Cela donne des échantillons de la loi normale multivariée recherchée $X = \mu + Z \cdot L^T$, où Z est la matrice des échantillons de la loi normale univariée et L est la matrice triangulaire inférieure de la décomposition de Cholesky de R.

3 Analyse de données

A Annexe A

```python
import numpy as np
from numpy import random as rnd
from matplotlib import pyplot as plt
import math

n = 10000
mu = 2
sigma = np.sqrt(9)

def realisation(x):
    return 1/(sigma*np.sqrt(2*np.pi))*np.exp(-(x-mu)**2/(2*sigma**2))

X = sigma*rnd.randn(n)+mu
nbins = 40
hist, bin_edges = np.histogram(X, bins=nbins, density=True)
hist = hist/(np.sum(hist)*(bin_edges[1]-bin_edges[0]))
plt.figure()
plt.bar(bin_edges[:-1], hist, align='edge', width=bin_edges[1] - bin_edges[0])

t = np.linspace(-15,15,1000)
y = [realisation(x) for x in t]
plt.plot(t,y,color='red')
plt.show()
```

FIGURE 2 – python histogramme-densité

B Annexe B

```python
def entropie_numérique_plus_log(X):
    s=0
    for i in range(n):
        s-=realisation(X[i])*np.log(realisation(X[i]))
    s*=1/n
    return s
s=entropie_numérique_plus_log(X)
print(s)
```

FIGURE 3 – entropie