1 Introduction

During this session, we put into action the concepts presented during the course with experiments simulated with `Python` and synthetic data. Don't hesitate to ask any question. To learn more about this subject, you may refer to the following articles [1, 2, 3] published in 2024.

2 Entropy estimation

Let's consider a sample of observations $x_1, x_2, ..., x_N$, assumed i.i.d.

1/ First run the program `main_entropy` to recover the examples from the course presentation for "normal", "mixture" and "VonMises" distributions. For each chosen density, a figure pops-up and a pdf file is generated in the current directory.

2/ For the Gaussian mixture, when the difference between both Gaussian means increases, what happens to the standard deviation? What happens to the entropy? Interpret this result.

3/ Check the option so that the histogram of the generated sample is plotted. Find the sample length (N) and the number of bins (K) that lead to "nice" approximations of the probability density functions.

4/ Check the option so that the program performs a Monte Carlo simulation that analyze the accuracy of the following two techniques that estimate the differential entropy:

$$\widehat{h}_1(X) = -\frac{1}{N}\sum_n \log p_X(x_n) \tag{1}$$

where p_X is the probability density function (pdf) of X. The second estimator is

$$\widehat{h}_2(X) = -\sum_k P_k \log_2 P_k + \log_2 \Delta \tag{2}$$

where P_k is the histogram height at bin k and Δ is the bin size. [1]

5/ Check the influence of the sample size (N), of the number of bins (K) on the entropy estimation. Which estimator do you prefer? Justify your answer.

3 Conditional entropy and mutual information

3.1 Measurement model

Let's consider a physical parameter (e.g. $\log N(\mathrm{H}_2)$) assumed to be distributed along a uniform distribution $X \sim \mathcal{U}_{[20,\,24]}$. Assume one observes an integrated intensity Y related to X by

$$Y = m(X) + N \tag{3}$$

where $N \sim \mathcal{N}(0, \sigma^2)$ is an additive white gaussian noise (AWGN) and

$$m(X) = \begin{cases} A.\mathrm{asinh}\left(\frac{X-C}{B}\right) & \text{if } X > C \\ 0 & \text{Otherwise} \end{cases} \tag{4}$$

where A, B and C are assumed to be known. In particular, we assume in the following that C depends on the considered line.

[1] Remember that $P_k \simeq \Delta\, p_X(a_k)$ where a_k is the bin location and $\sum_k P_k = 1$.

3.2 Line selection

1/ By running the program `main_condional_entropy_asinh`, you should recover the plots from the course: the 2D histogram of the pair X, Y, but also some density of $X|Y = y$ for some particular values of y.

2/ Note that to save the memory space, the 2D histogram is computed within a loop, which means that several 2D histograms are actually generated and summed. Check the influence of the number of histograms on the estimated distributions by changing `nb_histo`. Identify the tradeoff between accuracy and the cpu time of your computer.

3/ Note that in this program, the entropy is estimated based on Eq. (2). Why don't we use the estimator from Eq. (1)?

More precisely, for each bin k (of X) and l (of Y), let's note $\widehat{P}_{kl}$ the heigh of the observed 2D-histogram. The program computes the following estimations

$\widehat{P}_k = \sum_l \widehat{P}_{kl}$	the distribution of X
$\widehat{P}_l = \sum_k \widehat{P}_{kl}$	the distribution of Y
$\widehat{Q}_{k\|l} = \widehat{P}_{kl}/\widehat{P}_l$	the distribution of X conditioned by the knowledge of Y
$\widehat{h}_2(X\|Y = b_l) = -\sum_k \widehat{Q}_{k\|l} \log_2 \widehat{Q}_{k\|l} + \log_2 \Delta$	the entropy of $X\|Y = b_l$

where $\widehat{.}$ emphasize that these are all estimations. Furthermore, $h(X|Y)$ is estimated based on

$$\widehat{h}_2(X|Y) = \sum_l \widehat{h}_2(X|Y = b_l)\, \widehat{P}_l \tag{5}$$

whereas the mutual information is estimated based on

$$\widehat{I}(X;Y) = \sum_{k,l} \widehat{P}_{kl} \log_2 \frac{\widehat{P}_{kl}}{\widehat{P}_k \widehat{P}_l}$$

4/ Equation 5 is a complex way estimate the conditional entropy. Can you propose another solution to estimate the conditional entropy based on the estimation of the mutual information? What is the physical interpretation of this relation?

5/ Assume that the value of C (see Eq. (4)) depends on the line observed (e.g. ^{12}CO, ^{13}CO, C^{18}O, ...). Which line (i.e. which value of C) provides the most of information about X? You may simply modify line 16 of the provided program to run several values of C.

6/ Does this "optimal" value of C depends on other parameters? For example, what happens when A does from 3 to 1? Interpret this result.

7/ How would you proceed to adapt this methodology to another physical model (e.g. Radex)? What challenges will occur during the implementation of this technique?

8/ Assume one observed a map of lines (e.g. in Orion B) and that in the same area, one also observed with a spatial telescope (e.g. Herschel) a map of $N(\mathrm{H}_2)$. What can be done with these data?

4 Estimation performance analysis with the Cramér-Rao lower bound

Let's consider the same model than before, but we now use "estimation theory" notations:

$$X = m(\theta) + N \tag{6}$$

where X is the measurement, $N \sim \mathcal{N}(0, \sigma^2)$ is a AWGN and

$$m(\theta) = \begin{cases} A.\text{asinh}\left(\frac{\theta - C}{B}\right) & \text{if } \theta > C \\ 0 & \text{otherwise} \end{cases} \tag{7}$$

where A, B and C are assumed to be known and the parameter to estimate is θ.

For an additive gaussian noise as in Eq. (6), the probability density function of X is

$$\mathbb{P}_X(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp - \frac{[x - m(\theta)]^2}{2\sigma^2}$$

and it can be shown that the calculus of the Fisher information leads to

$$F(\theta) = \left(\frac{m'(\theta)}{\sigma}\right)^2 \quad \text{where } m' \text{ is the derivative of } m$$

Thus, the Cramér-Rao lower bound (CRB) is

$$\mathcal{B}(\theta) = \left(\frac{\sigma}{m'(\theta)}\right)^2$$

Remember that, for any unbiased estimator $\widehat{\theta}$,

$$\mathbb{V}\text{ar}(\widehat{\theta}) \geq \mathcal{B}(\theta)$$

which means that the CRB provides a precision of reference, independant of the choice of the estimator. With Eq. (7), it is straightforward to show that

$$m'(\theta) = \frac{A}{B\sqrt{1 + \left(\frac{\theta - C}{B}\right)^2}}$$

Thus, the precision of reference is

$$\boxed{\mathcal{B}(\theta) = \sigma^2 \frac{B^2 + (\theta - C)^2}{A^2}} \tag{8}$$

1/ The program `main_CRB_asinh` simulates P realisations $(\widehat{\theta}^{(p)})_{p=1,\dots,P}$ of an estimator of θ defined by $\widehat{\theta}(x) = m + B\sinh(x/A)$. Based on the shown histogram, comment on the performance of this estimator.

2/ By changing line 23 of the program, one can check the influence of θ on the S/N ($\frac{m(\theta)}{\sigma}$), on the CRB ($\mathcal{B}(\theta)$), but also on the bias and mean square error defined by

$$\widehat{\text{bias}}(\theta, \widehat{\theta}) = \frac{1}{P}\sum_{p=1}^{P} \widehat{\theta}^{(p)} - \theta \quad \widehat{\text{MSE}}(\theta, \widehat{\theta}) = \frac{1}{P}\sum_{p=1}^{P} (\widehat{\theta}^{(p)} - \theta)^2$$

How do you explain that the S/N and the MSE both increase with θ? What happens when you increase the value of P from 10^2 to 10^4? Explain this phenomenon.

3/ Can the CRB help to select the "best" line (i.e. the "best" value of C)? If so, how would you proceed to estimate it? Will it necessarily provide the same result than the mutual information?

4/ How would you proceed to adapt this methodology to another physical model (e.g. Radex)?

5 Next steps

Assume now that in Eq. (7), A is unknown, but we have some *a priori* on A characterized by a prior distribution $A \sim \mathcal{N}(\mu_A, \sigma_A^2)$.

1/ Can we still use the mutual information technique to select the best line? If yes, how would you proceed?

2/ Can we still use the CRB to select the best line? If yes, how would you proceed?

3/ What if B is unknown, but we have some *a priori* on B characterized by a prior distribution $B \sim \mathcal{N}(\mu_B, \sigma_B^2)$.

4/ Conclude on this hands on.

References

[1] Einig, L., Palud, P., Roueff, A., et al. 2024, A&A, 691

[2] Roueff, A., Pety, J., Gerin, M., et al. 2024, A&A, 686

[3] Segal, L., Roueff, A., Pety, J., et al. 2024, A&A, 692