Information gain in experimental design

Guillermo Torralba Elipe

guillermo.torralba@gssi.it

Gran Sasso Science Institute
&
INFN

7th HERD International Workshop
November, 2018

Outlook

A statistical approach for the comparison of the detector designs

- Theoretical basis
- Some practical examples:
  - Unfolding
  - Composition
- Generalization

- Suppose that we have designed the prototype of a detector to achieve a certain science goal(s).
- Now, we want to change the configuration: how do we know if the changes are good for our analysis?

$\rightarrow$ Put it in another way: we have two possible detector configurations: what is the best configuration for our objectives?

The so-called *relative entropy* or *information gain* allows us to solve this question quantitatively.

Information

The information is a measure of how much one distribution is peaked with respect to the flat distribution: (flat $\equiv$ no information)

$$H = \sum_k p_k \log(p_k) = -S \Rightarrow H = \int p(x) \log p(x) dx$$

The relative information (Kullback-Leibler divergence) is a measure of how much peaked one distribution (p) is respect to other (q):

$$\mathcal{H} = \sum_k p_k \log \left[\frac{p_k}{q_k} \right] = -S \Rightarrow H = \int p(x) \log \left[\frac{p(x)}{q(x)} \right] dx = D(p||q)$$

Prior $\rightarrow$ Posterior: how much information we have gained

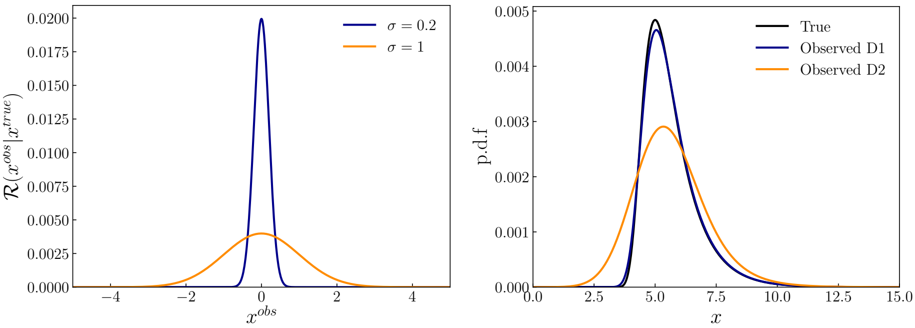
Let θ be the parameter(s) of interest and $\pi(\theta)$ the prior.
The possible values of θ given the measured data are given by $\pi(\theta|D) = \frac{\pi(D|\theta)\pi(\theta)}{\pi(D)}$. The information gained passing from the prior to the posterior is thus

$$\mathcal{H} = \int \pi(\theta|D) \log \left[\frac{\pi(\theta|D)}{\pi(\theta)} \right] d\theta$$

Example 1: Unfolding

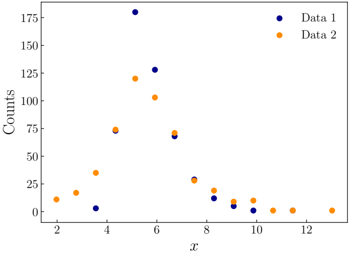
- Two detectors (D1, D2) with different resolutions: $\sigma_1 = 0.2$, $\sigma_2 = 1$.
- $f^{true} = \text{Moyal}(\mu = 5, \lambda = 0.5)$, $f^{obs} = f^{true} \otimes \text{Normal}(0, \sigma)$

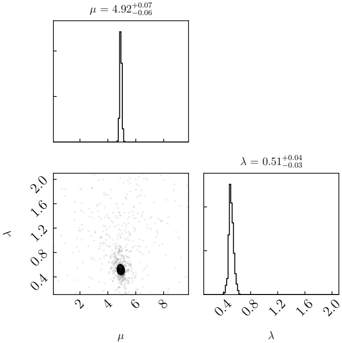
- Goal: to reconstruct the true signal (e.g. E deposited in the detector) $f^{true}(x^{true}|\mu, \lambda)$ from the observed $f^{obs}(x^{obs}|\mu, \lambda, \sigma)$

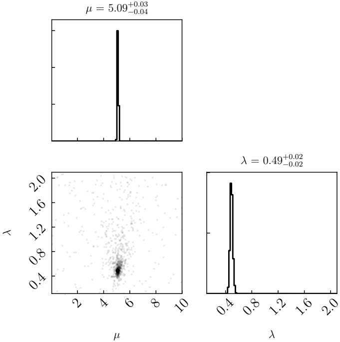


Using $\pi(\mu, \lambda) = \text{Uniform}(0, 10) \times \text{Uniform}(0.1, 2)$

Applying the unfolding to the observed data we obtain the posteriors for μ and λ for each detector.



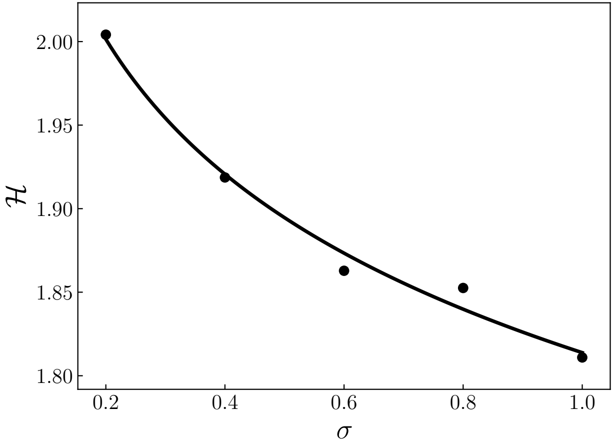




We compute the value of information:

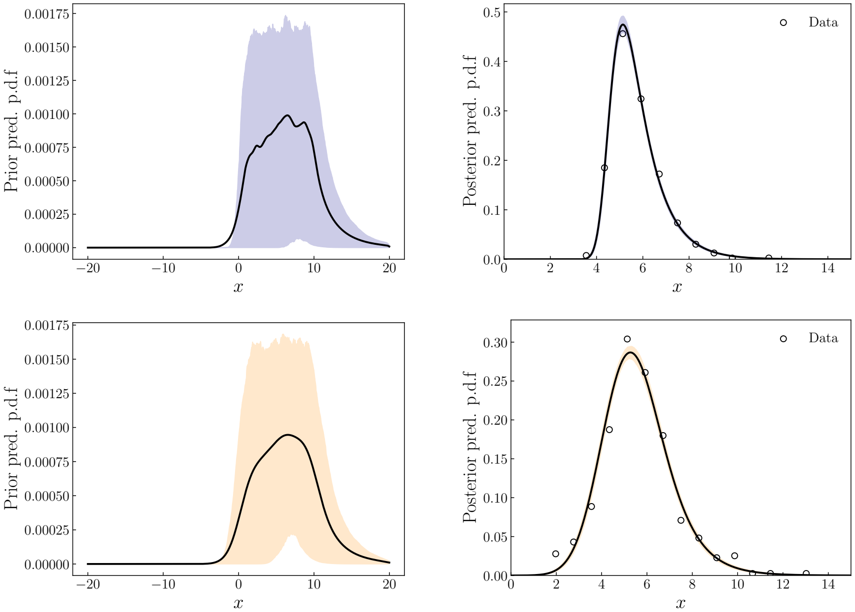
$$\mathcal{H}_1 = 2$$
$$\mathcal{H}_2 = 1.8$$

We extract more information from the best detector

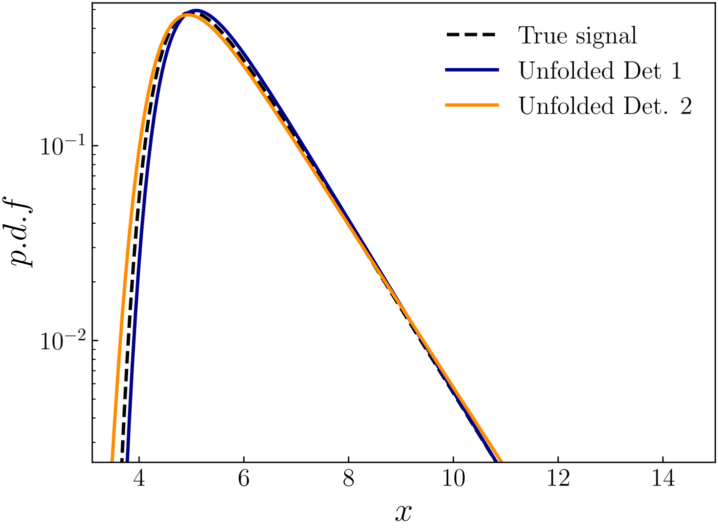


... as expected from intuition

To extrapolate the information to measured signal we can use the predictive distributions (which take into account all the possible values of μ and λ):



Result:

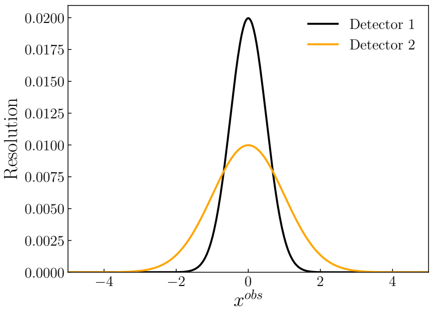


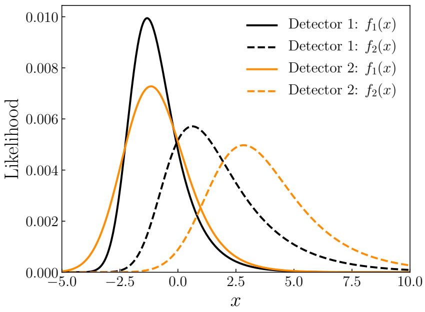
True signal
Unfolded Det 1
Unfolded Det. 2
10^{-1}
10^{-2}
$p.d.f$
4
6
8
10
12
14
x

Example 2: Composition

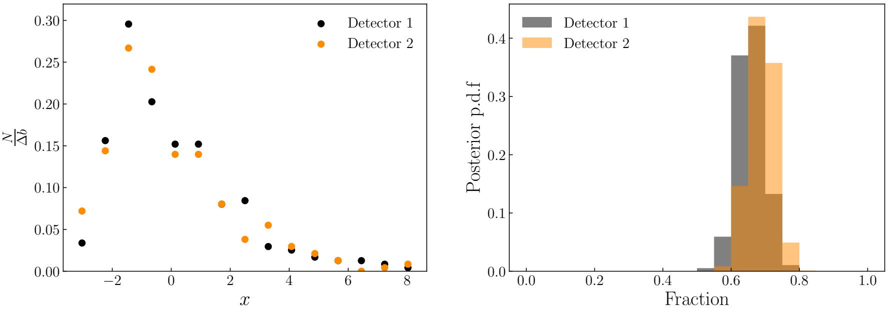
- Two detector configurations
- Detector 2 increases the distances between maximums

- Goal: infer the composition (e.g: p, He; with true proton fraction 0.7)





Second detector has worst resolution but increases de distance between maximums.



	$< \alpha >$	$\sigma[\alpha]$	$\mathcal{H}$
Detector 1	0.66	$1.7 \cdot 10^{-3}$	1.7
Detector 2	0.69	$1.5 \cdot 10^{-3}$	1.85

TABLE: Inferences both for detectors. The true fraction is 0.7

If we want to reconstruct the true signal, Det. 1 is the best option (Example 1) but for the composition inference the best is Det. 2

Generalization and comments

Good resolution, accurate arrival direction, precise composition inference, etc.

Let $\{A_i\}$ the set of analyses that we are thinking to carry out with the detector. Our best choice for the detector configuration is the one that maximizes $\mathcal{H}(A_1, A_2, ...)$ - If they are not independent -

Useful priors are those based from previous experiments (spectrum, composition,...)

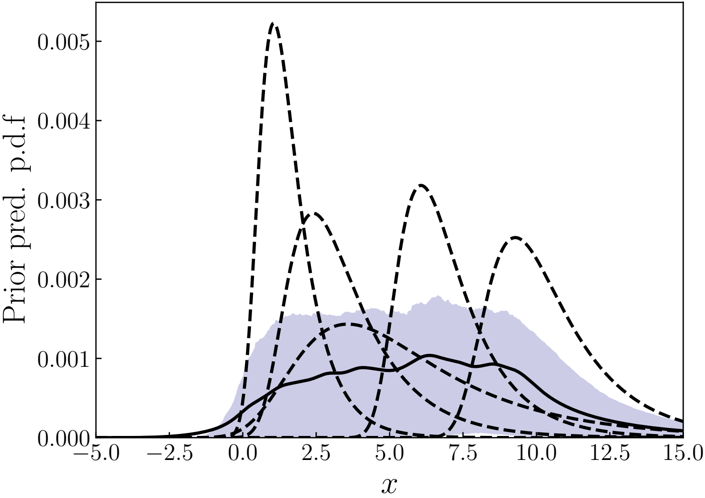
For all the analyses, we can compute the *expected information gain* repeating the analyses sampling from the prior and integrating over data:

$$\langle \mathcal{H} \rangle = \int \mathcal{H} dD$$

The method allows to obtain the best configuration for all the hadronic models used in simulations.

In HERD, the method can be implemented at the end of the simulation chain to evaluate the goodness of the detector for the different analyses that will be carried on.

Thanks


Prior pred. p.d.f
0.000
0.001
0.002
0.003
0.004
0.005
−5.0
−2.5
0.0
2.5
5.0
7.5
10.0
12.5
15.0
x