Treatment of systematic uncertainties with Bayesian networks

Dark Matter direct detection as a case study

Giovanni Grilli di Cortona, Andrea Messina, Stefano Piacentini

Introduction

- The propagation of the **systematic uncertainties** on the final results is often a challenging task in frontier physics
- Considering **Dark Matter** (DM) **direct detection** experiments, we developed a **general method** to fulfill this task using Bayesian networks and linear algebra
- The **final spectrum** is expressed as an **analytical function** of the **calibration parameters**, allowing to simultaneously fit both the parameters of interest and the calibration parameters
- We implemented the statistical aspects using **BAT** [1], and linear algebra on **GPU** with **CUDA** [2]

[1] F. Beaujean et al. "Bayesian analysis toolkit: BAT." https://github.com/bat/bat
[2] NVIDIA, CUDA, https://developer.nvidia.com/cuda-toolkit

Bayesian Networks

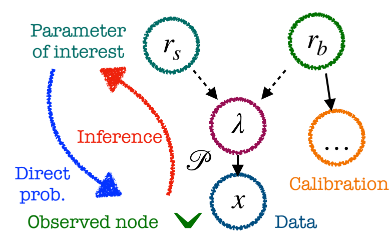
- Bayesian networks are a **graphical way** to represent the **probability**
- The probabilistic (solid arrows) or deterministic (dashed arrows) **connections** between observables are made **explicit** in the description of the likelihood (direct probability: cause → effect)
- The **Bayesian inference** is an **information flow** from the observed nodes to the parameter of interest

Probability decomposition

$$p(x, r_s) = p(x\,|\,r_s)p(r_s)$$

Inference after observation

$$p(r_s\,|\,x) = \frac{p(x\,|\,r_s)p(r_s)}{p(x)}$$



Systematic uncertainties in DM direct detection

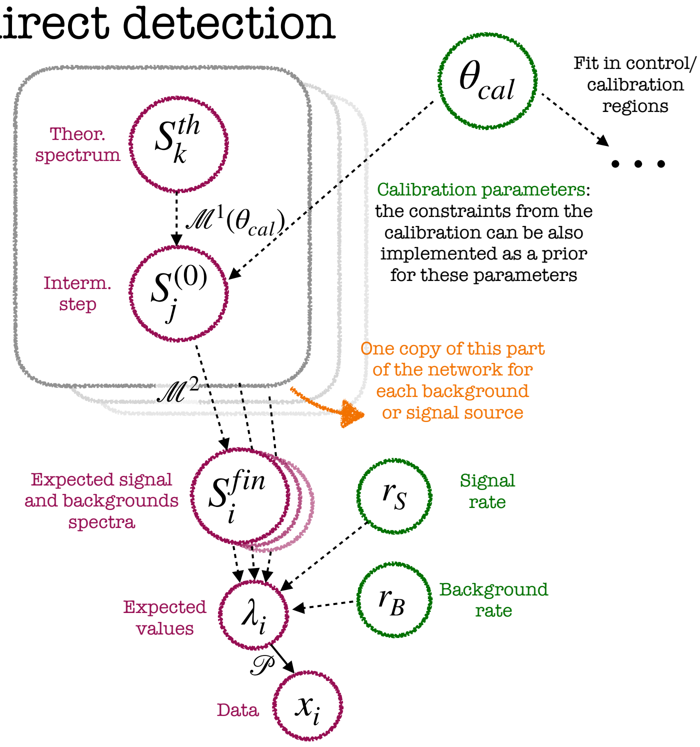
Typical **DM direct detection** experiment:

- **Working principle**: counting the number of possible DM interactions in the active volume that produce a certain amount of detectable quanta depending on the recoil energy
- **Data**: a spectrum of detectable quanta N_q, measure of the energy of the interaction
- **Response model**: relates the event original observable (e.g. the energy release) to the number of the detected quanta by means of a set of **calibration** parameters.

We developed a **method** to compute the **expected spectrum as an analytical function** of the **theoretical** spectrum and the **calibration parameters** implemented with CUDA on **GPUs**

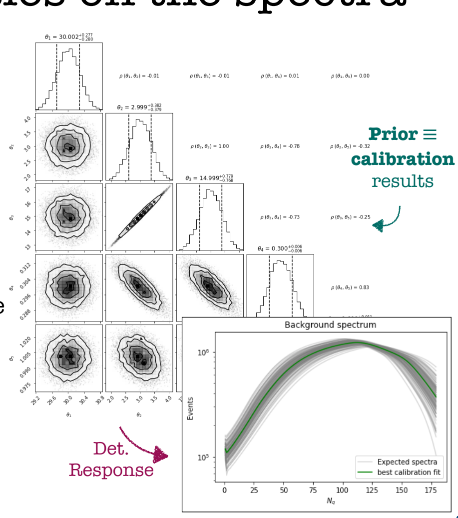
$$\underbrace{S_i^{fin}(N_q = i, \theta_{cal})}_{\text{Final expected spectrum}} = \sum_j \sum_k \underbrace{\mathscr{M}^2_{ij}\mathscr{M}^1_{jk}(\theta_{cal})}_{\text{Detector response}} \; \underbrace{S_k^{th}(E = E_k)}_{\text{Theor. spectrum}}$$

$$\begin{cases} \mathscr{M}^1_{jk}(\theta_{cal}) = p(N_q^{(0)} = j\,|\,E_k, \theta_{cal}) & \leftrightarrow \text{ quenching factor, ionization, recombination, Fano factor, ...} \\ \mathscr{M}^2_{ij} = p(N_q = i\,|\,N_q^{(0)} = j) & \leftrightarrow \text{ Efficiency, resolution, ...} \end{cases}$$



Impact of the systematic uncertainties on the spectra

- Our method **allows computing** the expected spectra **for any detector response** configuration sampling the prior $p(\theta_{cal})$
- It is possible to take into account the **bin-by-bin correlations**, and propagate the uncertainties and correlations of θ_{cal}



Resulting sensitivity

$$\underbrace{p(r_S, \boldsymbol{\theta}\,|\,\{x_i\})}_{\text{Posterior}} = N \; \overbrace{p(\{x_i\}\,|\,r_S, \boldsymbol{\theta})}^{\text{Likelihood}} \; p(r_S)\, p(\boldsymbol{\theta})$$

Prior ≡ calibration results

- The procedure tries to **fit** the data **weighting** the results with the p.d.f. of the **specific response model** returning the posterior p.d.f. (implemented in **BAT**)
- The propagation of the uncertainties on the final result is **automatic** by marginalizing the posterior
- The fit **updates the** knowledge on the **calibration** parameters.
- The typical impact of the uncertainties on the **sensitivity** is as large as a factor of 3