CHAPTER 1

Unfolding Methods

1 What is unfolding (maybe in Theory-chapter?)

- The inverse problem $f = \boldsymbol{R}u$
- The physics behind?
- The detector/experimental setup
- SiRi, OSCAr
- The raw data matrix, how it is produced?

1.1 The Response Matrix

The response matrix contains what we know about the circumstances of the experiment, a large part of which are the characteristics of the detector. The response matrix has elements given by:

$$R_{ij} = P(\text{measured in bin i} \mid \text{true in bin j}) \tag{1.1}$$

[1]. This can be read as the probability of observing an event in energy bin i, given the true event happening in bin j. In a nutshell, the response matrix describes how a signal is smeared over the other bins in the spectrum.

- Multiplying with $\boldsymbol{R}^{-1}$ leads to fluctuations because we cannot assume the observed data equals the *expectation values* for the data. Statistical fluctuations in the data is assumed to come from a real structure in the true spectrum and will be magnified. http://www-library.desy.de/preparch/desy/proc/proc14-02/P52.pdf

2 The Folding Iteration Method

The following section describes the methods developed by Guttormsen et al. [2].

The folded spectrum f is on the form

$$f = \boldsymbol{R}u, \tag{1.2}$$

where $\boldsymbol{R}$ is the response matrix and u is the expectation values for the true spectrum. The iterative method can then be described in 4 parts.

- First we use the measured spectrum r as an initial guess,
  $u_0 = r$
- We then fold this with the response matrix,
  $f_0 = \boldsymbol{R}u_0$
- The difference between the folded and the raw spectrum is calculated and added to the initial guess, and we end up with the next trial spectrum,
  $u_1 = u_0 + (r - f_0)$
- This is then repeated according to the following iteration scheme,
  $u_{i+1} = u_i + (r - f_i)$

This method is performed until $f_i \approx r$ within the experimental uncertainties [2]. It is important to note that for each new iteration, the oscillations between channels increase, as the solution approaches the inverted matrix solution $u = \boldsymbol{R}^{-1}r$, which exhibits large oscillations. [1][2]

2.1 The Compton Subtraction Method

As the resulting spectrum from the folding iteration method often contains some degree of fluctuations, the Compton subtraction method is performed to obtain a significantly more stable spectrum.

The first step is to define a new spectrum $v(i)$ as the observed data excluding the Compton contribution:

$$v(i) = p_f(i)u(i) + w(i)\,, \tag{1.3}$$

where $u(i)$ is the spectrum obtained from the folding iteration method, which multiplied with p_f gives the full energy contribution. The remaining contributions are contained in $w(i) = u_s + u_d + u_a = p_s(i)u(i) + p_d(i)u(i) + \sum p_{511}(i)u(i)$, representing single escape, double escape and annihilation (note the missing Compton contribution "u_c"). To match the observed energy resolution, each contribution is then smoothed with a Gaussian function. Next, we subtract this from the raw spectrum to obtain the Compton background spectrum:

$$c(i) = r(i) - v(i)\,. \tag{1.4}$$

This spectrum may exhibit significant oscillations, and is thus further smoothed. This smoothing carries a low risk of loss of important information due to the nature of the

spectrum not containing any sharp, narrow peaks. After this smoothing procedure on the individual contributions, we now "return" to the unfolded spectrum like so:

$$u(i) = \frac{r(i) - c(i) - w(i)}{p_f}. \tag{1.5}$$

Finally, to get closer to the true number of events, we correct for the total detector efficiency:

$$U(i) = \frac{u(i)}{\epsilon_{tot}(i)}. \tag{1.6}$$

This final spectrum shows higher stability compared to the result of the iteration method, while keeping similar statistical fluctuations to the raw spectrum. [2]

3 Fully Bayesian Unfolding

Bayes' theorem succinctly describes what we are asking for in the problem of unfolding, showing the relation between the truth spectrum $\boldsymbol{T}$, and the data we have obtained $\boldsymbol{D}$.

$$P(\boldsymbol{T}|\boldsymbol{D}) \propto L(\boldsymbol{D}|\boldsymbol{T}) \cdot P(\boldsymbol{T}) \tag{1.7}$$

The expected truth spectrum T and the raw spectrum D are binned with N_t and N_r bins, respectively. Each bin in T is assigned a prior probability distribution, describing our belief of the number of events expected to be present. We assume the data follows a Poisson distribution, meaning

$$L(\boldsymbol{D}|\boldsymbol{T}) = \prod_{r=1}^{N_r} \frac{f_r^{D_r}}{D_r!} e^{-f_r} \tag{1.8}$$

where

$$f_r = \sum_{t=1}^{N_t} T_t \cdot R_{rt}. \tag{1.9}$$

Here, R_{tr} is the element of the response matrix $R^{N_r \times N_t}$, corresponding to the probability that an event produced in the truth bin t is observed in the response bin r: $P(r|t)$. If we wish to include the background, all we have to do is add it to the sum:

$$f_r = \boldsymbol{B}_r + \sum_{t=1}^{N_t} T_t \cdot R_{rt}. \tag{1.10}$$

The next step is to employ a sampling scheme of the parameter space, usually a MCMC algorithm, to calculate $L(\boldsymbol{D}|\boldsymbol{T}) \cdot P(\boldsymbol{T})$ and arrive at a posterior distribution per bin in the expected truth spectrum.

3.1 Priors

- Uniform

There is practically an infinite amount of choices one can make for assigning a prior, depending on what knowledge one has beforehand. If one wishes to make the least amount of assumptions about the truth, a *uniform* prior is suitable. This flat distribution assigns equal probability to every outcome in the space of possibilities. The only assumption to be made here is determining boundaries on this space. Complete ignorance would strictly be represented with a uniform prior without any boundaries. This would mean we believe all numbers on the interval $[-\infty, \infty]$ to be equally likely in a one-dimensional space. Such a space is of course not possible to explore completely, and otherwise extremely large limits will be computationally unfeasible. This is especially true considering the fact that many problems are complex and demand multidimensional parameter spaces. In addition to this, unfolding in physics is often related to physical experiments pertaining to the counting of a number of events measured by a detector. In these cases, the existence of negative counts is unphysical, meaning a lower prior limit can safely be set to 0 (Discuss possibility for negative counts?).

Choosing the upper limit is not as straightforward. The ideal choice would be the largest possible limit that still allows for reasonable computational performance. Of course, if we have some knowledge about the domain of the possible truth-values, there is no need to pick a limit located significantly beyond this domain. Make sure the folded of prior contains the raw spectrum? In this thesis, an upper prior limit of 10 times the raw data will be used for the uniform prior. In other words, we say that the true value must be contained within an area relatively close to the observed value.

- Image of uniform distribution

- Logarithmic (using 'interpolate' in pymc3)

Another prior distribution we will use is the Logarithmic distribution. This will be implemented using the 'interpolated' class in PyMC3.

- Image of logarithmic distribution

- Explain prior dist for each bin

Every bin in the data spectrum is assigned a prior distribution as well as a likelihood for which we apply Bayes' theorem.

3.2 Multiplying response with (total) efficiencies

3.3 Likelihood

- Poisson (already mentioned above)

- Modified Poisson to take into account the total amount of counts

3.4 Sampling

- NUTS

There are several sampling methods possible for the problem of unfolding, a common example being Markov Chain Monte Carlo (MCMC) algorithms such as the Metropolis-Hastings algorithm. In the PyFBU-package, a variant of a Hamiltonian Monte Carlo (HMC) Markov Chain Monte Carlo algorithm is the default sampler. HMC aims to be much more efficient than regular MCMC algorithms by avoiding both sensitivity to correlated parameters and random walk tendencies [3]. A drawback to this is a significant sensitivity to step size as well as the number of steps, requiring manual tuning of these parameters. To circumvent this, Hoffman and Gelman created the No U-turn Sampler (NUTS), a variant of HMC which removes having to specify the number of steps. They also implemented an adaptive step size, meaning no manual tuning is necessary for running NUTS. Furthermore, they observed similar to better performance than other fine-tuned HMC algorithms [3]. The NUTS algorithm is implemented in the PyMC3 package [4] and is the default sampling algorithm in PyFBU.

- ADVI?

Another sampling method which will be used in this thesis is called Automatic Differentiation Variational Inference (ADVI) [5].

It is included in PyMC3 as a possible choice for the **initialization** of NUTS. In some cases, the use of this initialization will help when FBU would otherwise crash. Maybe due to the 'jitter' part of 'auto'?

3.5 Posterior inference

Now that the unfolding has been performed, how do we interpret the resulting posterior distribution? While other methods may only return a point value, not necessarily accompanied by the uncertainties, FBU allows us to directly look at the final distribution per bin, and thus observe the result and its corresponding degree of belief. Of course, we are able to quantify these concepts in multiple ways. Here, we take a look at some of the methods of posterior inference.

3.5.1 Point estimates

We will consider three different point estimates, the posterior mean, median and mode:

- Posterior mean: The mean of the posterior distribution which minimizes the mean squared error (MSE) [6].
- Posterior median: The median of the posterior distribution which minimizes the expected absolute error [6].
- Posterior mode: The mode of the posterior distribution, also called the Maximum A Posteriori (MAP), which represents the most likely value for the parameter in question. This does not take into account any skewness of the posterior nor the existence of multiple modes of similar magnitudes.

3.5.2 Credible intervals

The credible interval is the Bayesian version of the frequentist confidence interval. It depends on the posterior and is defined as any interval that encompasses a certain percent of the posterior density. The difference between confidence and credible intervals is subtle, but not negligible. In the case of frequentist inference, the parameter in question, lets say θ, is treated as an unknown, but fixed value. The limits of the confidence interval are treated as random variables. Therefore, a confidence level of 95% means that for 100 repeated experiments, 95 of the confidence intervals will contain θ. Note that this does not mean there is a 0.95 probability of finding θ in every confidence interval. [7][8]

For bayesian inference however, the random-trait is switched, with credible interval limits being fixed, and θ treated as the random variable. The credible interval takes our prior belief into account, while the confidence interval relies only upon the data. A 95% credible interval covers 95% of the posterior and can then be said to contain θ with a probability of 0.95. [7][8]

There are many types of credible intervals, the only requirement being that it covers a certain amount of area of the posterior. Some examples of ways of constructing credible intervals are:

- Using the posterior mean as the interval center.
- Making sure the probability of being outside the interval is equal on all sides (equal tailed).
- Making the interval as narrow as possible, the Highest Posterior Density interval (HPD). This will include the most likely values, as well as the mode of the posterior if it is unimodal.

We will be using the HPD interval which, together with the point estimates mentioned above, will give a solid basis of comparison against the true energies of the γ-ray spectrum.

4 The FBU-package?

5 PyMC3?

- Built-in distributions possible to be used for both likelihoods and priors, such as Uniform, Normal, Poisson etc, as well as Truncated versions of some.
- Custom distribution! 'Interpolated' using x-values and pdf-values.

5.1 Theano?

6 Error statistics? MAE (MSE), R2-score, others?

- Compare unfolded with truth (if known)

- Compare unfolded*response with raw
- Residual plots for both

7 1-dimensional test spectrum

8 2-dimensional test spectrum

- Compare posterior with likelihood*prior

9 28 Si spectrum

- Compare result with Valas, using logscale prior, modified likelihood?
- Background?

Valsdóttir found that, when the background was known, including it in the unfolding significantly improved the results.

10 146 Nd spectrum

- 250 channels
- Background, nanoseconds
- 500 channels
- Background? nanoseconds?
- 453 keV

  Discrepancy at lower energies, unavoidable with this response, we have information to correct this. Maybe on the other states too? Results should match OMPy and thats what we want?
- 1-1.4 MeV
- 6-6.2 MeV