The Structure of Inverse Problems in Experimental Particle Physics

Sean Gilligan

Abstract

This report provides a survey of some of the common methods used by the high energy physics community to understand and solve ill-posed inverse problems as they pertain to signal distortions that result from imperfect measuring devices and processes. These methods are in general collectively referred to as unfolding. The specifics of data and data collection methods are generalized. Common features are discussed insofar as they contribute to the necessary understanding of the data and implementation of any covered unfolding methods. In order to construct a slightly more wholistic picture some additional topics are briefly touched upon if they relate to other common aspects of data analysis in particle physics, but only during parts of relevant discussions where they would otherwise normally appear.

1 Introduction

A common problem faced in the quantitative sciences and their associated technologies is the introduction of errors during the data collection process. While the possible sources of these errors are as varied as the possible events which the data might describe, significant work has been done to develop methods the can help would-be analysts reconcile them. The requisite understanding of a scenario's underlying systematic and stochastic processes might not allow researchers to truly reverse entropy or make up for the finite resolution of a detector, but it can approximate them with a quantifiable degree of certainty.

The applied mathematics that this involves falls within the general category of **inverse problems**, and there are a variety of labels used to refer to the procedures in its arsenal. There is the colloquially vague **unsmearing**, but there are also names that reference specific applications and methods. For the sake of simplicity, and any necessary physical constraints, the manner of inverse problems addressed here will only have satisfactory solutions that involve linear operations that map from one Hilbert space[1] to another. Symbolically this can be expressed by the equation

$$Az = u,$$

where A is a linear operator acting on an element $z \in Z$, the sought solution, to produce an element $u \in U$, the observed data. Within the context of the methods described herein z and u take the form of continuous or discrete distributions that when integrated or summed over the domain of their arguments result in finite real quantities.

The difficulty of solving for z can be classified into one of two camps. The easiest cases involve conditions that create a **well-posed** problem, which requires that [13]

1. a solution exists $\forall u \in U$,
2. the solution is unique,
3. and if $u_n \longrightarrow u$, $Az_n \longrightarrow u_n$, and $Az \longrightarrow u$, then $z_n \longrightarrow z$.

Conditions 1 and 2 work together to imply that the inverse operator A^{-1} exists, and Condition 3 is often worded to describe the inverse as continuous, which implies that small deviations in u should correspond to similar deviations in z. When one or more of these conditions are not met, the problem is said to be **ill-posed**, and some of the consequences of assuming otherwise should hopefully become clear in the coming pages.

In Section 1.1 the convolution and deconvolution are briefly discussed in the capacity of continuously distributed data. They are then generalized in Section 1.2, and discretization introduced in Section 1.3 establishes a data structure that makes accessible to analysts a host of rigorous statistical and computational methods. This comes in handy as Section 1.4 presents a number of simulated

Entire books have been written on this subject that do not begin to cover the full scope of the methods developed to deal with ill-posed problems. With that in mind the hope for this short paper is for it to serve as an introduction to ill-posed problems while providing some degree of direction for those who would like to know more.

1.1 The Deconvolution

One way to characterize a basic example of a situation suitable for being treated as a convolution would be one that should be very familiar to anyone who has ever taken statistics course. Assume that data collected regarding n statistical events represent the measurement of n independent and identically distributed (i.i.d.) random variables $\boldsymbol{X} = \{X_1, X_2, \ldots, X_n\}$ from a distribution of possible values represented by the probability density function (PDF) $f_X(x)$, such that the probability of a random variable X_i having a value between x_a and x_b is

---

[1] The definition of a Hilbert space is provided in Appendix A.1 for convenience.

$$P(\, x_a < X_i < x_b \,) = \int_{x_a}^{x_b} f_X(x)\, dx$$

and

$$\int_{\mathcal{X}} f_X(x)\, dx = 1,$$

where $\mathcal{X}$ represents the domain of x. The error introduced during the measurement process is similarly represented by a set of i.i.d. random variables $\boldsymbol{\varepsilon} = \{\varepsilon_1, \varepsilon_2, \ldots, \varepsilon_n\}$ with a PDF $f_\varepsilon(\varepsilon)$, where the sets $\boldsymbol{\varepsilon}$ and $\boldsymbol{X}$ are typically assumed to be independent of each other. The set of measured/reconstructed values $\boldsymbol{Y} = \{Y_1, Y_2, \ldots, Y_n\}$ then are also i.i.d. and can be defined in terms of the preceding sets of variables such that for event $i \in \{1, \ldots, n\}$,

$$\begin{aligned} Y_i &= g(X_i, \varepsilon_i) \\ &= X_i + \varepsilon_i. \end{aligned} \tag{1}$$

In light of this relationship, the corresponding PDF $f_Y(y)$ can be found explicitly through an operation on $f_X(x)$ and $f_\varepsilon(\varepsilon)$ using the mathematics of functional analysis. Stated in more general terms, the empirical density function f_Y is formed from the **convolution** of the true density function f_X and the error density function f_ε, and is defined by [10]

$$f_Y \equiv f_X * f_\varepsilon \tag{2}$$

$$\begin{aligned} f_Y(y) &\equiv \int_{\mathcal{X}} f_X(x) f_\varepsilon(\varepsilon)\, dx \\ &= \int_{\mathcal{X}} f_X(x) f_\varepsilon\Big(g_x^{-1}(y)\Big) \Big| J_{g_x^{-1}}(y) \Big|\, dx \\ &= \int_{\mathcal{X}} f_X(x) f_\varepsilon(y - x)\, dx, \end{aligned} \tag{3}$$

where J represents the Jacobian of the transformation involved in performing the change of basis on f_ε from ε to x, which is necessary for the evaluation of the integral for a given y. The magnitude of the Jacobian for transformation of ε to $y - x$ through the manipulation of Equation (1) happens to be 1.

As the collection of measured values $\boldsymbol{Y}$ accumulates an estimate of empirical density $\hat{f}_Y$ can readily be formed. However, a major goal in an analysis of data like this is typically to develop an accurate estimate of the true density $\hat{f}_X$. Using the information contained in

$\hat{f}_Y$ to accomplish this necessarily requires some attempt at finding an inverse process to the convolution, i.e. the **deconvolution**.

For cases in the form of this particular example there are a variety approaches, but they commonly involve the Fourier transform of the density functions $\{f_X, f_\varepsilon, f_Y\}$ into their corresponding characteristic functions $\{\phi_X, \phi_\varepsilon, \phi_Y\}$ [9][10]. Minor aspects of the definition for the Fourier transform can vary slightly between applications, resulting primarily from the use of different scale factors and sign conventions. Here it will be defined for some random variable $U \in \mathbb{R}$ with density function $f_U(u)$ and random variable $T \in \mathbb{R}$ as

$$\phi_T(t) = \int_{-\infty}^{\infty} f_U(u)\, e^{itu}\, du. \tag{4}$$

When conditions permit the inverse Fourier transform can be found via

$$f_U(u) = \int_{-\infty}^{\infty} \phi_T(t)\, e^{-itu}\, dt. \tag{5}$$

The Fourier transform is important in deconvolution methods because when you apply it to the convolution of two density functions the link between their respective characteristic functions becomes purely multiplicative, i.e.

$$f_Y = f_X * f_\varepsilon \implies \phi_Y = \phi_X \phi_\varepsilon.$$

An instructional proof of this result is provided on page 447 of [4]. The steps so far characterize a typical deconvolution scheme, with later steps consisting of various ways to perform density estimation and addressing issues similar to those that will be seen ahead [9].

1.2 Generalizing

The remainder of this paper is dedicated to a more generalized study of these type of problems. With the understanding that even experts can be fairly loose and inconsistent with their vocabulary, this paper will do its best to provide clear definitions. To begin, while most literature on deconvolution methods do use the word "convolution", this operation is also referred to by the German word *faltung* [12]. The latter's English translation, **folding**, is featured prominently in the particle physics community, but refers to a more generalized process than what is described by Equation (3) [7][1][2]. In general, folding and **unfolding** refer to two sets of processes within which the sets of convolution and deconvolution processes form proper respective subsets.

One way to arrive at the desired generalization is with the help of conditional probability.

Thinking of $\{X, Y\}$ as a continuous bivariate random vector with joint PDF $f(x, y)$ and marginal PDFs $f_X(x)$ and $f_Y(y)$, we can define the conditional PDF of Y given that $X = x$ as function of y, $f(y\,|x)$ [5]. The relationship between these PDFs is sufficient to define any one of them in terms of operations involving one or more of the others. As such, for $f_Y(y)$ it can be shown

$$\begin{aligned} f_Y(y) &= \int_{\mathcal{X}} f(x, y)\, dx \\ &= \int_{\mathcal{X}} f(y\,|x) f_X(x)\, dx \\ &= \int_{\mathcal{X}} K(x, y) f_X(x)\, dx. \end{aligned} \tag{6}$$

While integrating over x, $f(y\,|x)$ is implicitly treated as a function of both x and y. Acknowledging this allows for understanding Equation (6) as a Fredholm integral of the first kind with a Kernel function $K(x, y)$ that reflects the physical measurement process [3]. The relationship between x and y in $K(x, y)$ is not defined, but when the kernel is a function of the difference of its arguments, such that $K(x, y) = K(y - x)$, Equation (6) becomes the convolution described in Equation (3).

In particle physics experiments, analysts make use of Monte-Carlo (MC) simulations to estimate detector response to randoms samples from some true distribution $f_X(x)^{\mathrm{MC}}$, which is itself estimated by way of MC simulations using models that typically contain theory being tested by the experiment in question. The resulting measured distribution $f_Y(y)^{\mathrm{MC}}$ grants implicit knowledge of $K(x, y)$ by way of Equation (6) [2]. Finding the inverse of this Kernel is then the goal, as it should in theory allow for the mapping of experimental observations $\boldsymbol{Y}$, as randomly sampled from $f_Y(y)$, back to their true values $\boldsymbol{X}$.

1.3 Discretization

In practice researchers are only ever dealing with estimates $\hat{f}_X$, $\hat{f}_Y$, $\hat{f}_X^{\mathrm{MC}}$, and $\hat{f}_Y^{\mathrm{MC}}$, and the sets of data that contribute to these estimates are organized by bin into histograms that form unnormalized granular approximations of their true distributions. Thinking in terms of these histograms allows for the reformulation of Equation (6) into the linear matrix equation:

$$\boldsymbol{\nu} = \boldsymbol{R\mu}. \tag{7}$$

The vectors $\boldsymbol{\nu}$, $\boldsymbol{\mu}$ and matrix $\boldsymbol{R}$ relate to their continuous counterparts by [2]:

$$\begin{aligned} \text{true distribution } f_X(x) &\longrightarrow \boldsymbol{\mu} \in \{\mathcal{U} \equiv \mathbb{R}_+^M \cup \mathbf{0}\} \text{ the unknown true bin counts,} \\ \text{measured distribution } f_Y(y) &\longrightarrow \boldsymbol{\nu} \in \mathcal{V} \equiv \{\mathbb{R}_+^N \cup \mathbf{0}\} \text{ the measured bin counts,} \\ \text{Kernel } K(x,y) &\longrightarrow \boldsymbol{R} \text{ the rectangular } N\text{-by-}M \textbf{ response matrix}. \end{aligned}$$

The components of vectors $\boldsymbol{\nu}$ and $\boldsymbol{\mu}$ represent the number of events that have occurred within the regions of x and y that define the components' corresponding bins. For $i = 1, \ldots, N$ and $j = 1, \ldots, M$ the components of matrix $\boldsymbol{R}$ are defined by the conditional probability [6]

$$\begin{aligned} R_{ij} &= P(\text{measured value in bin } i | \text{true value in bin } j) \\ &= \frac{P(\text{measured value in bin } i \text{ and true value in bin } j)}{P(\text{true value in bin } j)} \\ &= \frac{\int_{\text{bin } i} \int_{\text{bin } j} K(x,y) f_X(x) dx\, dy}{\int_{\text{bin } j} dx\, f_X(x)} \\ &\equiv P(\nu_i | \mu_j). \end{aligned} \tag{8}$$

In terms of $P(\nu_i|\mu_j)$ the full response matrix then has the form

$$\boldsymbol{R} = \begin{pmatrix} P(\nu_1|\mu_1) & P(\nu_1|\mu_2) & \ldots & P(\nu_1|\mu_N) \\ P(\nu_2|\mu_1) & P(\nu_2|\mu_2) & \cdots & P(\nu_2|\mu_N) \\ \vdots & \vdots & \ddots & \vdots \\ P(\nu_M|\mu_1) & P(\nu_M|\mu_2) & \ldots & P(\nu_M|\mu_N) \end{pmatrix}. \tag{9}$$

With these definitions Equation (7) tells us that an event produced in bin μ_j has some probability ≥ 0 of being measured in each of the N bins of $\boldsymbol{\nu}$, and that each bin count ν_i receives potential contributions from each of the M bins in $\boldsymbol{\mu}$, i.e.

$$\nu_i = \sum_{j=1}^{M} R_{ij} \mu_j. \tag{10}$$

The number of bins are typically set such that $M \leq N$, with the convention $N = M + 1$ being common. A higher number of bins in the measured distribution reflects that the measuring process is expected to map some events in $\boldsymbol{X}$ to values of $\boldsymbol{Y}$ that are outside the region of values that define the initial M bins. These one or more extra bins are intended to account for all the possible values that a particular event could be mapped to, such that for a given event starting in bin j one might expect the probabilities of it being measured in each of the N final bins to sum to 1.

However, in practice there are a variety of constraints on events that can either result in them not being included for analysis or even prevent them from being detected at all. For example, an analyst might cut events observed in regions of a detector that result in insufficient data collection, or maybe some event information carriers miss the detector entirely, resulting in such events going unseen. In either case the effect of these missing events is described using the detector **efficiency**, and represented mathematically by the N-vector $\boldsymbol{\epsilon}$, where component ϵ_j is the efficiency of the jth true bin defined[2] by [6]:

$$\sum_{i=1}^{N} P(\nu_i|\mu_j) = \sum_{i=1}^{N} R_{ij} = \epsilon_j \leq 1. \tag{11}$$

In contrast to this are contributions to measured counts from **background** processes. Just as events produced in a region of interest can be smeared out of it, events produced out of it can be smeared into it. The crossed barrier could correspond to the variable of interest, but it can also include events excluded from analysis due to assigned constraints on other variables that describe the event. Background processes are often studied and dealt with prior to the unfolding procedures described in the paper. It is briefly mentioned here to provide a slightly more holistic picture of particle physics analysese. Mathematically, background would be included by modifying Equation (10) to read

$$\nu_i = \sum_{j=1}^{M} R_{ij}\mu_j + \beta_i, \tag{12}$$

where β_i is the ith component of the N-vector $\boldsymbol{\beta}$, which represents the binned background counts. This leads to equations like $\nu_i^{\text{sig}} = \nu_i - \beta_i$ in order to specify the expected number of measured counts that are from the signal of interest. Going forward background will be assumed to already have been accounted for, and ν_i will refer to the expected signal counts of bin i.

As all these variables so far have been derived from the exact continuous distributions $f_X(x)$ and $f_Y(y)$, they correspond to the expectation values that researchers are estimating during data collection and analysis. As this is a counting process the components of the observed number of signal events $\boldsymbol{n}$, an N-vector, are often related to the components of the expected number of observed counts $\boldsymbol{\nu}$ as a collection of N separate and independent Poisson processes. That is to say the observed counts n_i in bin i are treated as i.i.d. random variables with the probability mass function

---

[2] In the continuous case it is typically written as $\epsilon(x)$, and understood to be the conditional probability of an event producing any measured value given it has a true value of x. It is typically absorbed into $K(x, y)$ where it goes on to manifest within $\boldsymbol{R}$ in the manner shown in Equation (11) [2].

$$P(n_i|\nu_i) = \frac{\nu_i^{n_i} e^{-\nu_i}}{n_i!}. \tag{13}$$

The counts n_i would in theory then form the estimate $\hat{\nu}_i$ of the expected counts ν_i by

$$\begin{aligned} \nu_i &= \mathrm{E}[\hat{\nu}_i] = \mathrm{E}[n_i] \\ &= \mathrm{Var}[\hat{\nu}_i] = \mathrm{Var}[n_i]. \end{aligned}$$

Understanding the probability distribution of $\boldsymbol{n}$ allows for unfolding methods that involve the use of maximum likelihood estimation. For methods based on least squares it becomes necessary to find the covariance matrix $\boldsymbol{\Sigma}^{\nu}$ of the observations, which for independent Poisson processes has components of the form

$$\begin{aligned} \Sigma_{ij}^{\nu} &= \mathrm{Cov}[n_i, n_j] \\ &= \delta_{ij} \nu_i, \end{aligned} \tag{14}$$

where δ_{ij} is the Kronecker delta[3]. The path to the covariance matrix of the estimated true distribution $\hat{\boldsymbol{\mu}}$ can be considered briefly by considering the maximum log-likelihood, where it can be shown

$$\begin{aligned} \log L(\boldsymbol{\mu}) &= \sum_{i=1}^{N} \log \left(\frac{\nu_i^{n_i} e^{-\nu_i}}{n_i!} \right) \\ &= \sum_{i=1}^{N} \left(n_i \log \nu_i - \nu_i - \log n_i! \right) \\ \frac{\partial \log L}{\partial \mu_k} &= \sum_{i=1}^{N} \frac{\partial \log L}{\partial \nu_i} \frac{\partial \nu_i}{\partial \mu_k} \\ &= \sum_{i=1}^{N} \left(\frac{n_i}{\nu_i} - 1 \right) R_{ik} = 0. \end{aligned}$$

Some minor algebra here reproduces the estimate $\hat{\boldsymbol{\nu}} = \boldsymbol{n}$, which has been assumed up until now. Continuing with an additional derivative shows

---

[3] The Kronecker delta δ_{ij} is a piecewise function of variables i and j defined by $\delta_{ij} = \begin{cases} 0 & \text{if } i \neq j \\ 1 & \text{if } i = j \end{cases}$.

$$\begin{aligned}\frac{\partial^2 \log L}{\partial \mu_k \partial \mu_l} &= -\sum_{i=1}^{N}\left(\frac{n_i}{\nu_i^2}\frac{\partial \nu_i}{\partial \mu_l}\right) R_{ik} \\ &= -\sum_{i=1}^{N}\frac{n_i R_{il} R_{ik}}{\nu_i^2},\end{aligned}$$

the negative of the expectation value of which is the Fisher information. The Fisher information's relationship with the Cramér-Rao lower bound can, as its name implies, create a lower bound on the covariance matrix of an unbiased estimator of $\boldsymbol{\mu}$. For such an unbiased estimator $\hat{\boldsymbol{T}}_{\mu\text{nbiased}}$ we have

$$\begin{aligned}\operatorname{Cov}\left(\hat{\boldsymbol{T}}_{\mu\text{nbiased}}\right) \geq \boldsymbol{I}(\boldsymbol{\mu})^{-1} &= \left(-E\left[\frac{\partial^2 \log L}{\partial \mu_k \partial \mu_l}\right]\right)^{-1} \\ &= \left(\sum_{i=1}^{N}\frac{E[n_i] R_{il} R_{ik}}{\nu_i^2}\right)^{-1} \\ &= \left(\sum_{i=1}^{N}\frac{R_{il} R_{ik}}{\nu_i}\right)^{-1} \\ &= \left(\boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\right)^{-1} \\ &= \boldsymbol{R}^{-1} \boldsymbol{\Sigma}_\nu \boldsymbol{R}^{-1^T}. \end{aligned} \tag{15}$$

Indeed, this matrix must then be lower bound for the covariance matrix of any unbiased estimator of $\boldsymbol{\mu}$. This is some good insight to have before getting into the weeds of working on actual data.

1.4 A Simulated Example

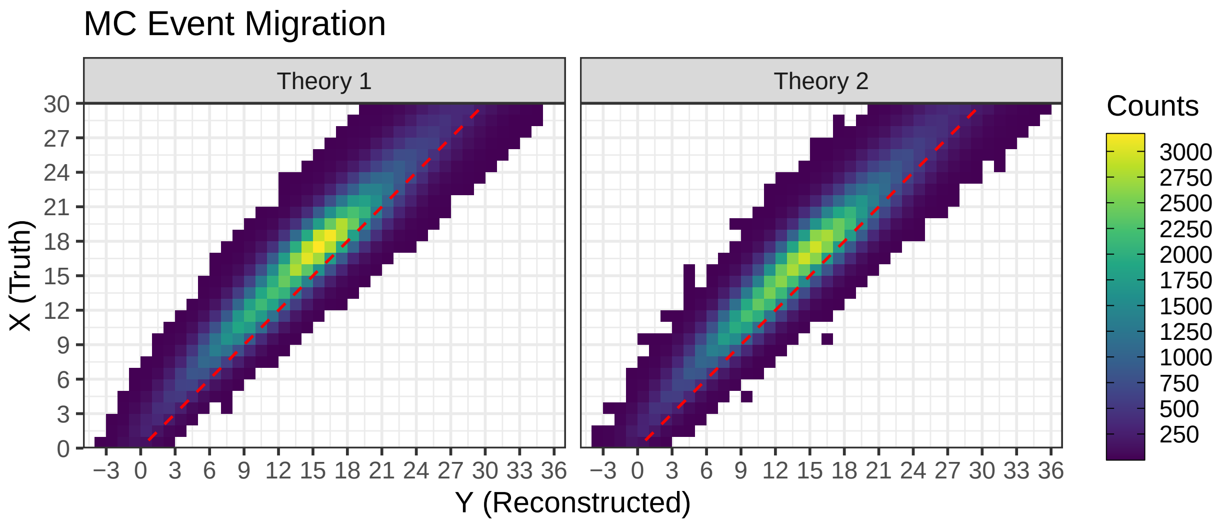
Consider the following three sets of i.i.d. random variables from separate Cauchy distributions.

$$\begin{aligned}X_{1,i} &\sim \text{Cauchy}(x_0^{(1)}, \gamma_1) \\ X_{2,i} &\sim \text{Cauchy}(x_0^{(2)}, \gamma_2) \\ X_{3,i} &\sim \text{Cauchy}(x_0^{(3)}, \gamma_3)\end{aligned}$$

Current models predict that some class of physics events observed in past detectors are solely coming from the first two processes, such that for n events the number coming from the first process is an i.i.d. random variable from the binomial distribution $B(n, p)$. Meanwhile, a new model has been developed that suggests that the third process has been occurring this whole time but has been incorrectly categorized as one or the other of the first two. Napkin math has estimated a contribution rate that is reflective of some probability p_3, such that the binomial distribution is actually a multinomial distribution with probabilities $\boldsymbol{p} = \{p_1, p_2, p_3\}$.

A new experiment is being designed and funded to test this new theory, on top of many others, and simulations are being performed to give analyzers plenty of opportunities to develop their collaboration's analysis framework, perform calibrations, and make ready a myriad of studies that hope to shed light on many an unanswered question. The corresponding MC simulations performed during this time include such simulations for the physics events of interest, but also for the detector. The purposes of this paper a relatively simple set of simulations have been performed, leading to 400,000 simulated events per theory that are meant to represent the MC simulations, as well as a set of 20,000 simulated events for each theory the are meant to represent hypothetical data as they might be produced.

Please see Appendix B for information regarding the simulations performed. Please see Figure



2 Unfolding in particle physics

Intro to section 2

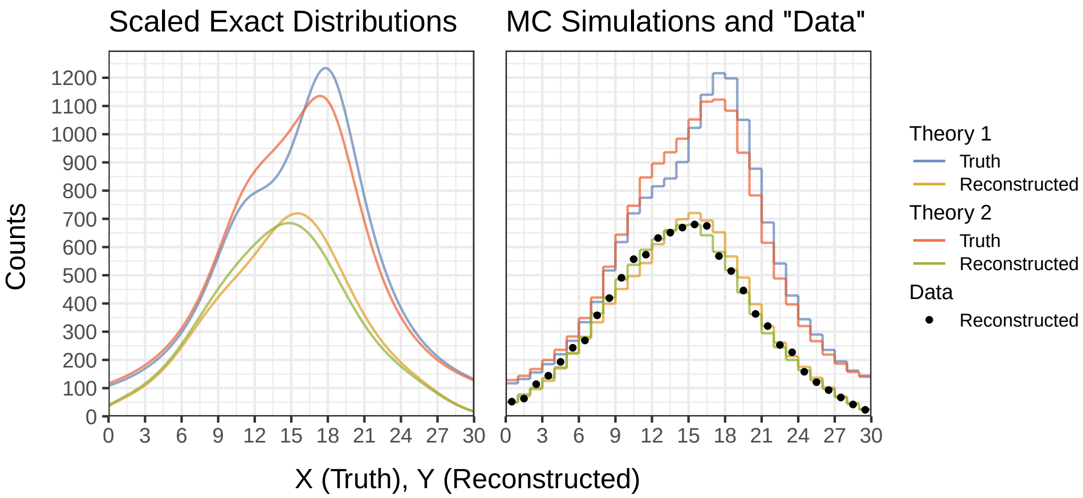


Figure 1: *The above plots feature two bimodal gamma distributions depicting events before and after after detector effects. (Top Left) The two continuous distributions correspond to the theoretical PDFs of the two distributions rescaled to correspond with the counts from 20,000 events. The histograms are calculated from the PDFs and correspond to the expected event counts from 20,000 simulated events. (Bottom Left) These four histograms consist of the same expected event counts as well as one instance of actual counts resulting from 20,000 simulated events. (Right) A visual study of simulated detector efficiency is provided by a side-by-side comparison of two heat maps that demonstrate the skewness and dispersion added by a simulated measurement process for detected and undetected events. This study is not intended to meaningfully represent a hypothetical distribution of undetected events in any real detector. Actual detector efficiencies are almost certainly governed by much more complicated collections of parameters.*

2.1 Inverting the Response Matrix

In the event of Equation (7) being well-posed the obvious approach would be to construct the unique inverse of the response matrix $\boldsymbol{R}^{-1}$ and map the reconstructed counts back to an estimate of the true counts via

$$\hat{\boldsymbol{\mu}} = \boldsymbol{R}^{-1}\boldsymbol{n}. \tag{16}$$

A statistical justification for this comes from performing generalized least squares [8] fit to estimate $\boldsymbol{\mu}$, which relies on approximating bin count n_i as normally distributed with mean ν_i

and variance $1/\nu_i$. Minimizing the sums of squares yields

$$
\begin{align}
\min_{\boldsymbol{\mu}} \nabla_{\boldsymbol{\mu}} \boldsymbol{\chi}^2(\boldsymbol{\mu}) &= \nabla_{\boldsymbol{\mu}} (\boldsymbol{R}\boldsymbol{\mu} - \boldsymbol{n})^T \boldsymbol{\Sigma}_\nu^{-1} (\boldsymbol{R}\boldsymbol{\mu} - \boldsymbol{n}) \notag \\
&= \nabla_{\boldsymbol{\mu}} (\boldsymbol{\mu}^T \boldsymbol{R}^T - \boldsymbol{n}^T) \boldsymbol{\Sigma}_\nu^{-1} (\boldsymbol{R}\boldsymbol{\mu} - \boldsymbol{n}) \notag \\
&= \nabla_{\boldsymbol{\mu}} (\boldsymbol{\mu}^T \boldsymbol{R}^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{R}\boldsymbol{\mu} - \boldsymbol{\mu}^T \boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{n} - \boldsymbol{n}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\boldsymbol{\mu} + \boldsymbol{n}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{n}) \notag \\
&= \nabla_{\boldsymbol{\mu}} (\boldsymbol{\mu}^T \boldsymbol{R}^T \boldsymbol{\Sigma}_n^{-1} \boldsymbol{R}\boldsymbol{\mu} - 2\boldsymbol{\mu}^T \boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{n} + \boldsymbol{n}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{n}) \notag \\
&= 2\boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\boldsymbol{\mu} - 2\boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{n} \notag \\
&= 0 \notag \\
\implies \boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\boldsymbol{\mu} &= \boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{n} \notag \\
\implies \hat{\boldsymbol{\mu}} &= (\boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\boldsymbol{\mu})^{-1} \boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{n} = \boldsymbol{R}^+ \boldsymbol{n} \tag{17} \\
\implies \boldsymbol{R}^+ &= (\boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\boldsymbol{\mu})^{-1} \boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1}, \tag{18}
\end{align}
$$

where $\boldsymbol{R}^+$ is the Moore-Penrose generalized inverse (or pseudo-inverse) [2], and it is assumed that $\boldsymbol{R}^T \boldsymbol{\Sigma}_\nu^{-1} \boldsymbol{R}\boldsymbol{\mu}$ is not singular. The corresponding covariance matrix is more quickly calculated by way of

$$
\boldsymbol{\Sigma}_\mu = \mathrm{Cov}[\hat{\boldsymbol{\mu}}, \hat{\boldsymbol{\mu}}] = \mathrm{Cov}[\boldsymbol{R}^+\boldsymbol{n}, \boldsymbol{R}^+\boldsymbol{n}] = \boldsymbol{R}^+ \mathrm{Cov}[\boldsymbol{n}, \boldsymbol{n}] \boldsymbol{R}^{+^T} = \boldsymbol{R}^+ \boldsymbol{\Sigma}_\nu \boldsymbol{R}^{+^T}.
$$

One such inverse matrix was calculated from each of the MC Theory simulations and both were applied to every set of the Reconstructed counts. The residuals of these unfolding calculations are paired with a 1σ uncertainty from the estimated Truth counts. They are then scaled by the expected Truth for the assumed theory and plotted on a log10 scale in Figure 2.

This degree of failure is disastrous, but one thing to notice here is that this problem does not satisfy Condition 3 of a well-posed problem as described in Section 1. That the inverse calculated here maps $\boldsymbol{n}$ to negative counts indicates that the estimate $\hat{\boldsymbol{\mu}} \notin \mathcal{U}$, the space of allowable solutions defined at the beginning of Section 1.3.

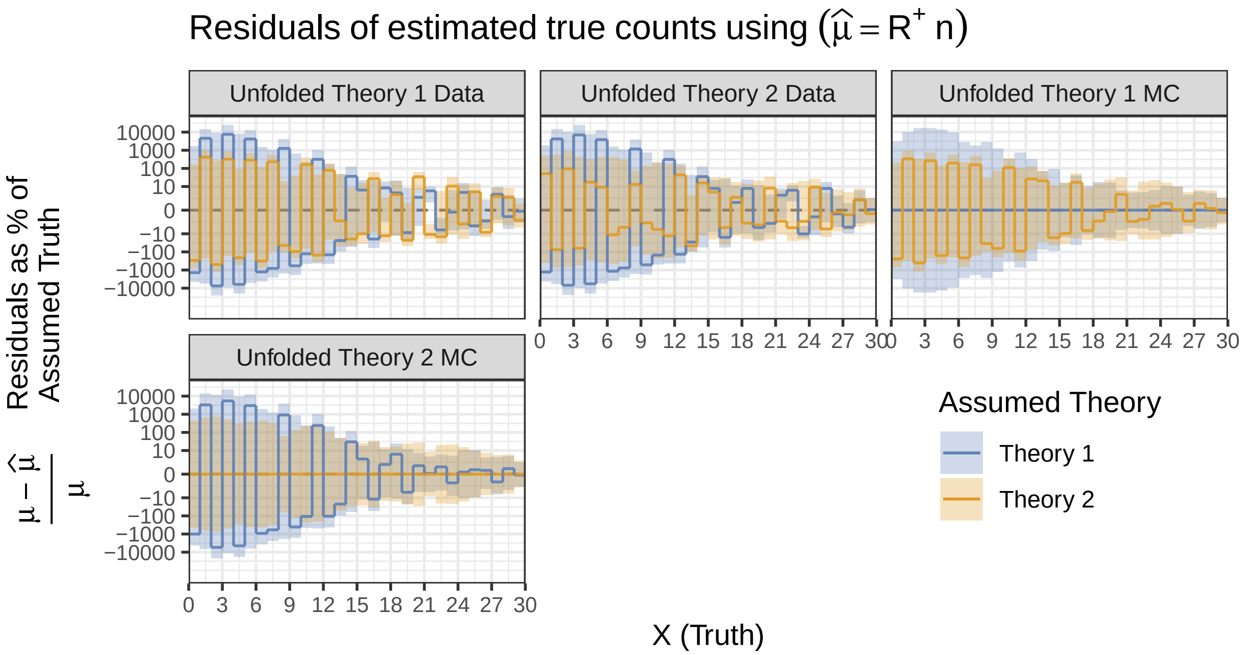
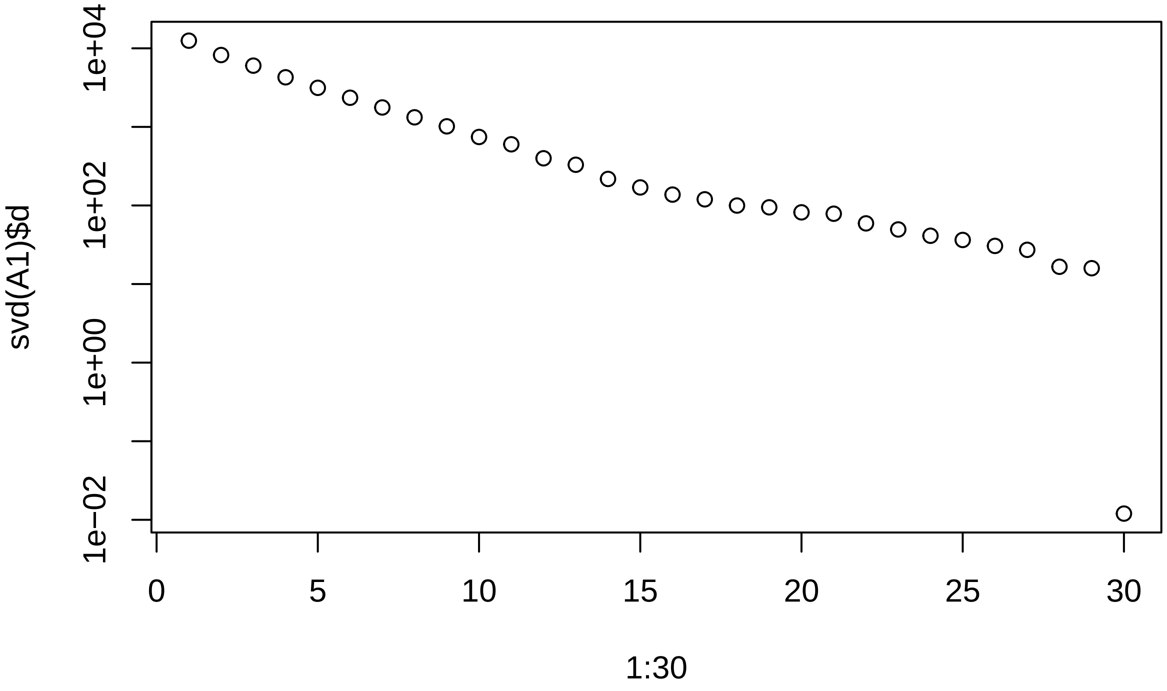
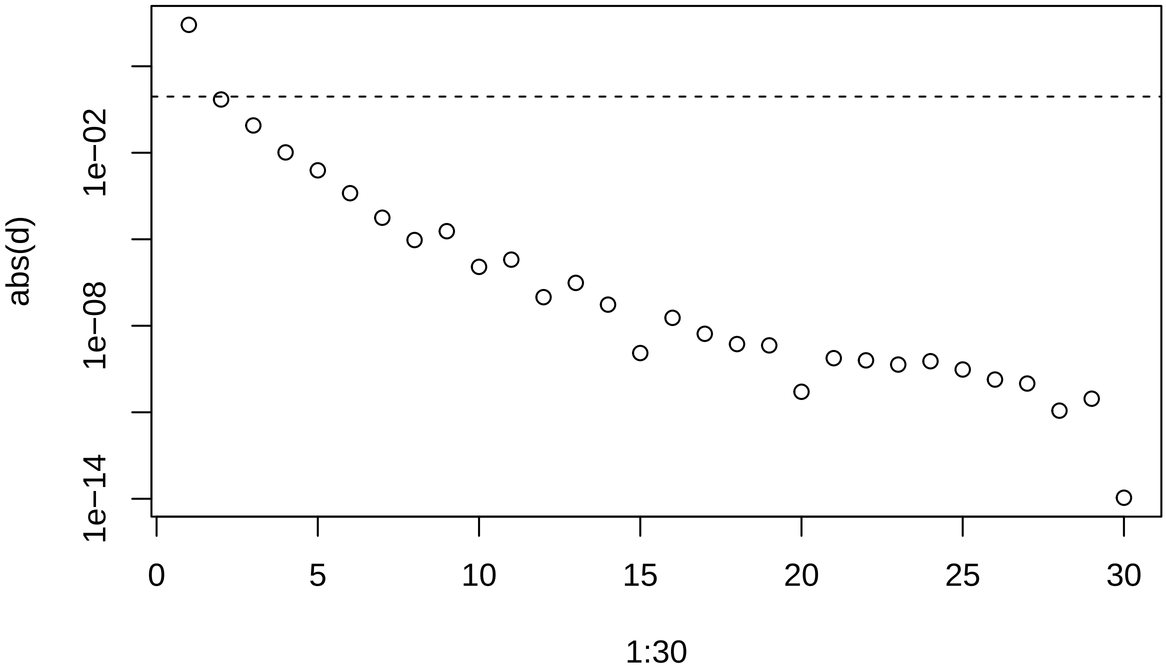
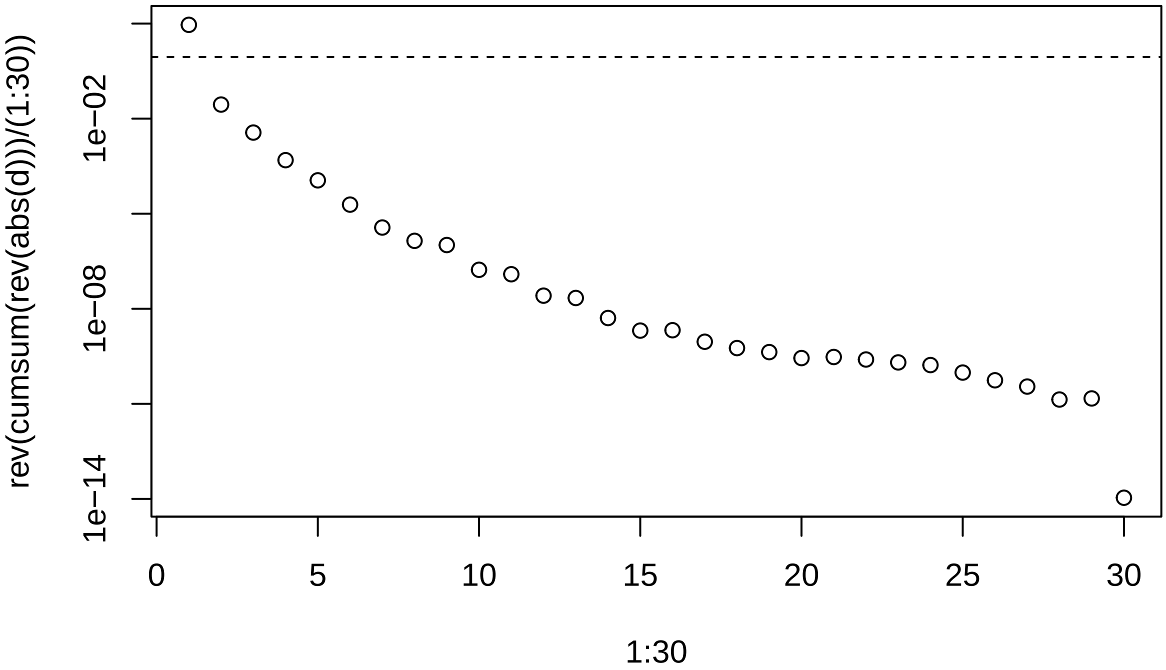


Figure 2: *Plots featuring the ratio of the residuals to the assumed theory's expected truth for the five sets of reconstructed counts when unfolded using response matrices derived from MC simulations for both Theory 1 and Theory 2. Note the log10 scale of the vertical axes, the large uncertainties, and negative estimated counts. The MC results are exact when unfolded under the correct theory for tautological reasons.*





abs(d)
1e−02
1e−08
1e−14
0
5
10
15
20
25
30
1:30


rev(cumsum(rev(abs(d)))/(1:30))
1e−02
1e−08
1e−14
0
5
10
15
20
25
30
1:30

A Supplementary Mathematical Definitions and Derivations

A.1 Hilbert Spaces

Hilbert spaces are a prominent feature in the field of functional analysis. They see significant application in partial differential equations, quantum mechanics, and signal processing, where they are commonly implemented in the performance of Fourier analysis. Mathematically they represent an extension beyond the real and complex geometric-like vector spaces developed by earlier generalizations of Euclidean spaces in the 19th century. Developments in real analysis at the beginning of the 20th century lead the spaces of functions and sequences to being conceptualized as linear spaces in their own right.

As extensions of previously understood spaces they necessarily exist at the intersection of several other important spaces that aught to be understood beforehand. With that said, the following definitions come from Rudin in [11]. To start, a **vector space**, as defined here, consists of a set X of vectors for which addition and scalar multiplication are defined such that for all $x, y, z \in X$ and any complex number $\alpha \in \mathbb{C}$

1. there exists a vector in X such that
   (a) addition is commutative: $x + y = y + x$,
   (b) addition is associative: $x + (y + z) = (x + y) + z$,
2. αx exists in X such that $1x = x$, $0x = 0$ (the zero vector), and multiplication is distributive:
   (a) $\alpha(\beta x) = (\alpha\beta x)$,
   (b) $\alpha(x + y) = \alpha x + \alpha y$, and
   (c) $(\alpha + \beta)x = \alpha x + \beta x$.

The range of α above describes a complex vector space. If α is restricted to the reals $\mathbb{R}$, then X is considered a real vector space. Note that vector spaces include more than just traditional coordinate-style vectors, but also include function spaces such as the vector space of all polynomials with degree of at most n, which has the basis $\{1, x, x^2, \ldots, x^{n-1}, x^n\}$.

Typically associated in applications, metric spaces form a another relevant set of spaces that has some significant overlap with the vector spaces. A space X is said to be a **metric space** if for all $x, y \in X$ there exists an operator $d(x, y)$ that maps them to a nonnegative real number that defines their distance from each other within X. The properties of this operator are

1. $0 \leq d(x, y) < \infty$ for all x and $y \in X$,
2. $d(x, y) = 0$ iff $x = y$,
3. $d(x, y) = d(y, x)$ for all x and $y \in X$,
4. $d(x, z) \leq d(x, y) + d(y, z)$ for all x, y, $z \in X$.

For a metric space X, the distance operator d is referred to as the metric on X. The intersection of the vector and metric spaces form the set of normed spaces. As an extension of the conditions thus far, a space X is a **normed space** if $\forall x \in X$ there exists a nonnegative real number $||x||$, called the **norm** of x such that

1. $||x + y|| \leq ||x|| + ||y|| \; \forall x, y \in X$,
2. $||\alpha x|| = |\alpha| \, ||x||$ if $x \in X$ and α is a scalar,
3. $||x|| > 0$ if $x \neq 0$.

Such a set is said to be **complete** if every **Cauchy sequence** in X converges to a point in X. A Cauchy sequence in a metric space X is any sequence $\{x_n\}$ that $\forall \varepsilon > 0$ there exists an integer N such that $d(x_m, x_n) < \varepsilon$ when $m > N$ and $n > N$. A quick example of this is the sequence defined by $x_n = \sqrt{n}$. For some starting x_m and x_n where $m - n = \delta$, we have

$$\begin{aligned}
d(x_m, x_n) &= \sqrt{m} - \sqrt{n} \\
&= \sqrt{n + \delta} - \sqrt{n} \\
&= (\sqrt{n + \delta} - \sqrt{n}) \frac{\sqrt{n + \delta} + \sqrt{n}}{\sqrt{n + \delta} + \sqrt{n}} \\
&= \frac{n + \delta - n}{\sqrt{n + \delta} + \sqrt{n}} \\
&= \frac{\delta}{\sqrt{n}(\sqrt{1 + \delta/n} + \sqrt{1})} \\
&< \frac{1}{\sqrt{n}} \left(\frac{\delta}{2} \right) < \varepsilon \\
\implies n &> \left(\frac{\delta}{2\varepsilon} \right)^2 .
\end{aligned}$$

Noting that for constant δ the limit of $\frac{1}{\sqrt{n}} \left(\frac{\delta}{2} \right)$ as $n \longrightarrow \infty$ is the zero vector (the point of convergence) would also be sufficient to show that $x_n = \sqrt{n}$ is a Cauchy sequence.

Incidentally, a normed vector space that is complete as defined here meets the definition of a **Banach space**. An additional subset of the normed vector spaces consists of those spaces in which for all $x, y \in X$ there exists a real or complex number $\langle x, y\rangle$ defined by an operator called the **inner product**. For all $x, y, z \in X$ this operation must satisfy

1. $\langle x, y\rangle = \langle y, x\rangle^*$ (where the * represents the complex conjugate),
2. $\langle x + y, z\rangle = \langle x, z\rangle + \langle y, z\rangle$,
3. $\langle \alpha x, y\rangle = \alpha\langle x, y\rangle$ (for $\alpha \in \mathbb{C}$),
4. $\langle x, x\rangle \geq 0$, and
5. $\langle x, x\rangle = 0$ iff $x = 0$.

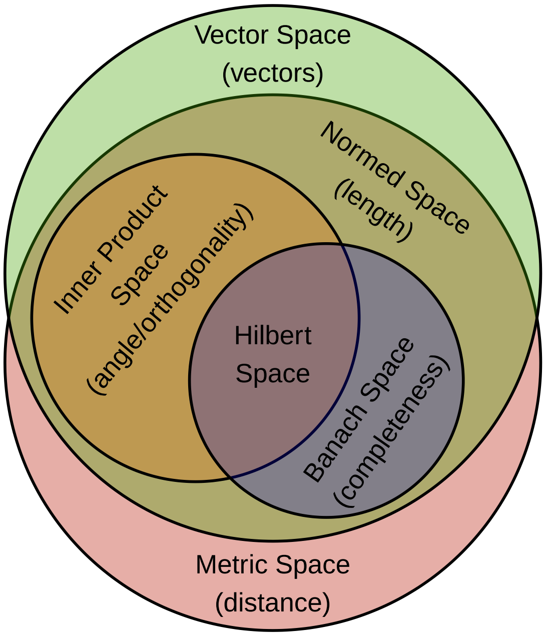


Figure 3: *A Venn diagram representing the intersection and nesting of the spaces described in Appendix A.1.*

A space that satisfies these requirements forms an **inner product space**, and the inner product defined in such a space relates to the form of its norm, such that $||x|| = \langle x, x\rangle^{1/2}$. Finally, at the intersection of Banach spaces and inner product spaces are the Hilbert spaces. I.e. a **Hilbert space** is a complete vector space with an inner product defined by its norm.

A commonly presented example is the L^2 function space, which consists of functions that are square integrable, i.e. if $f(x) \in L^2 \implies ||f(x)||^2 = \int_\chi |f(x)|^2 dx < \infty$, where χ is the domain of x. The subset $L^2[-\pi, \pi]$, where $\chi = [-\pi, \pi]$, has the well known Fourier series as a basis, which is commonly written such that for $f(x) \in L^2[-\pi, \pi]$

$$f(x) = \frac{a_0}{2} + \sum_{n=1}^{\infty} \left[a_n \cos(nx) + b_n \sin(nx)\right],$$

where

$$a_n = \tfrac{1}{\pi} \int_{-\pi}^{\pi} f(x) \cos(nx) dx$$

and

$$b_n = \tfrac{1}{\pi} \int_{-\pi}^{\pi} f(x) \sin(nx) dx.$$

Verification that this basis meets all the requirements laid out so far is beyond the scope of this paper.

B Description of simulations

The Cauchy processes specifically are of the form

$$\begin{aligned} X_{1,i} &\sim \text{Cauchy}(11, 4) \\ X_{2,i} &\sim \text{Cauchy}(18, 4) \\ X_{3,i} &\sim \text{Cauchy}(14, 5) \end{aligned}$$

The probabilities under Theory 1 (in which only the first two processes take place) is $\boldsymbol{p} = \{0.3, 0.7\}$. The probabilities governing Theory 2 are generated from $\boldsymbol{p}$ by

$$\boldsymbol{p'} = \boldsymbol{\Lambda p} = \begin{pmatrix} 0.75 & 0 \\ 0 & 0.75 \\ 0.6 & 0.1 \end{pmatrix} \begin{pmatrix} 0.3 \\ 0.7 \end{pmatrix} = \begin{pmatrix} 0.225 \\ 0.525 \\ 0.25 \end{pmatrix}$$

the effects of detector smearing is represented by i.i.d random variables generated by the conditional Gaussian process

$$\varepsilon_i \sim N\left(\mu(X_i), \sigma(X_i)^2\right),$$

the mean and variance of which are functions defined by

$$\begin{aligned} \mu(X_i = x) &= -x^{1/4} \quad \text{and} \\ \sigma(X_i = x) &= \log\left(\frac{x + 10}{4}\right). \end{aligned}$$

The efficiency is similarly conditional on X_i, and is modeled here as a Bernoulli process with i.i.d random variables $\epsilon_i \sim \text{Bernoulli}\Big(p(X_i)\Big)$, where the average detection rate (when $\epsilon_i = 1$) is a function of the form

$$p(X_i = x) = 1 - e^{-\sqrt{x}/4}.$$

C Bin-by-bin

In this approach a multiplicative **correction factor** C_i is applied to the observed number of signal events n_i for each bin to produce the estimator of μ_i [6],

$$\hat{\mu}_i = C_i n_i. \tag{19}$$

The correction factors are determined by taking the respective ratios of a bin's MC simulated truth signal event counts μ_i^{MC} to its MC simulated reconstructed signal event counts ν_i^{MC},

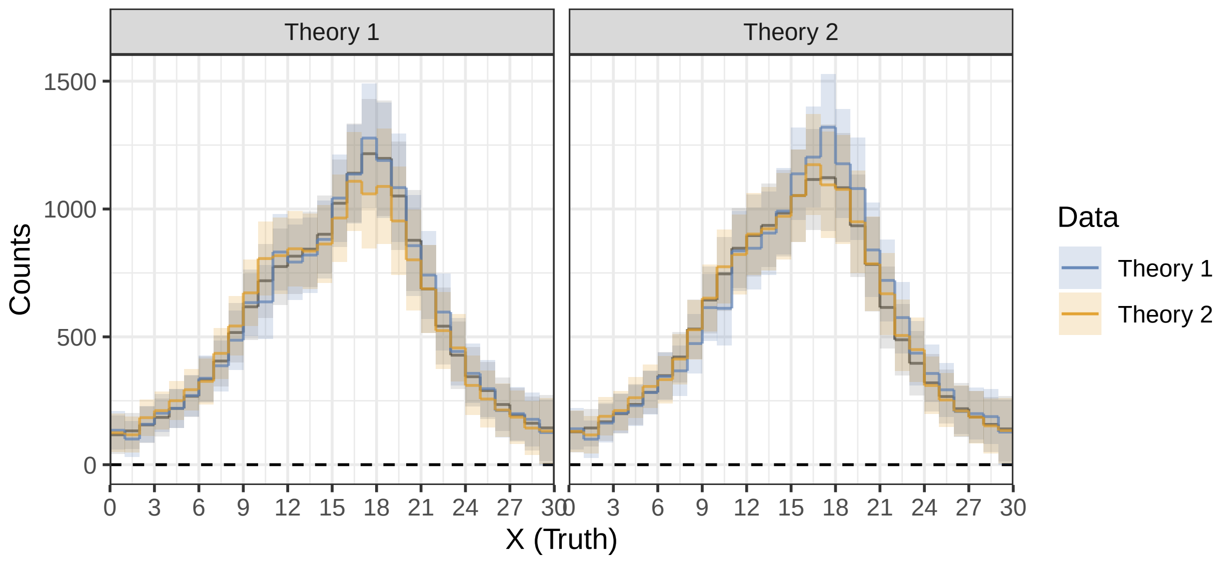
$$C_i = \frac{\mu_i^{\text{MC}}}{\nu_i^{\text{MC}}}. \tag{20}$$

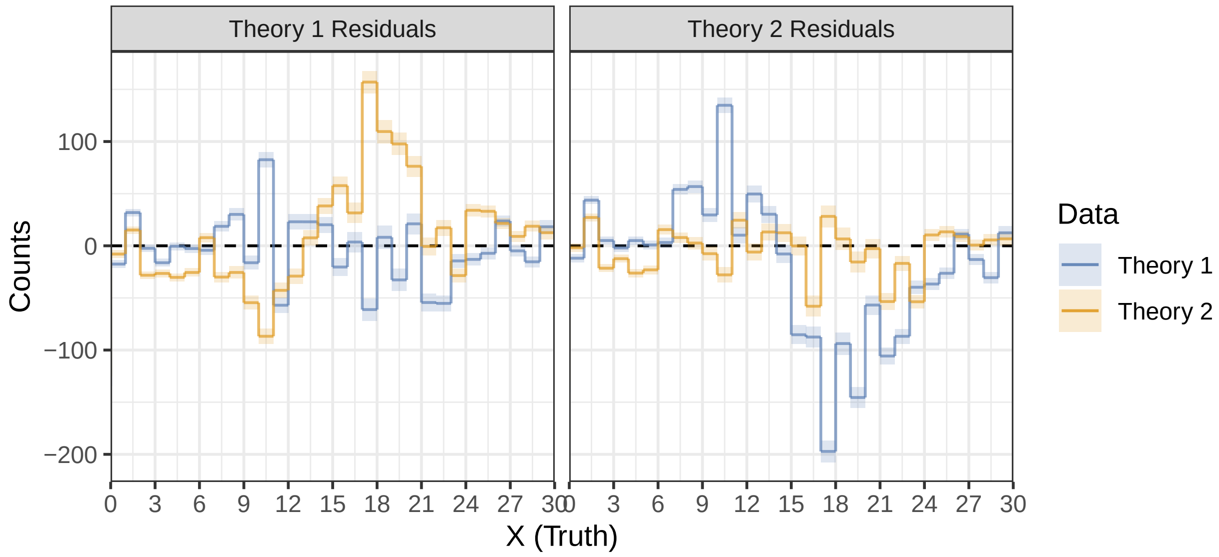
The covariance matrix $\mathbf{\Sigma}_\mu$ of this estimator derives naturally from Equations (14) and (19), with components

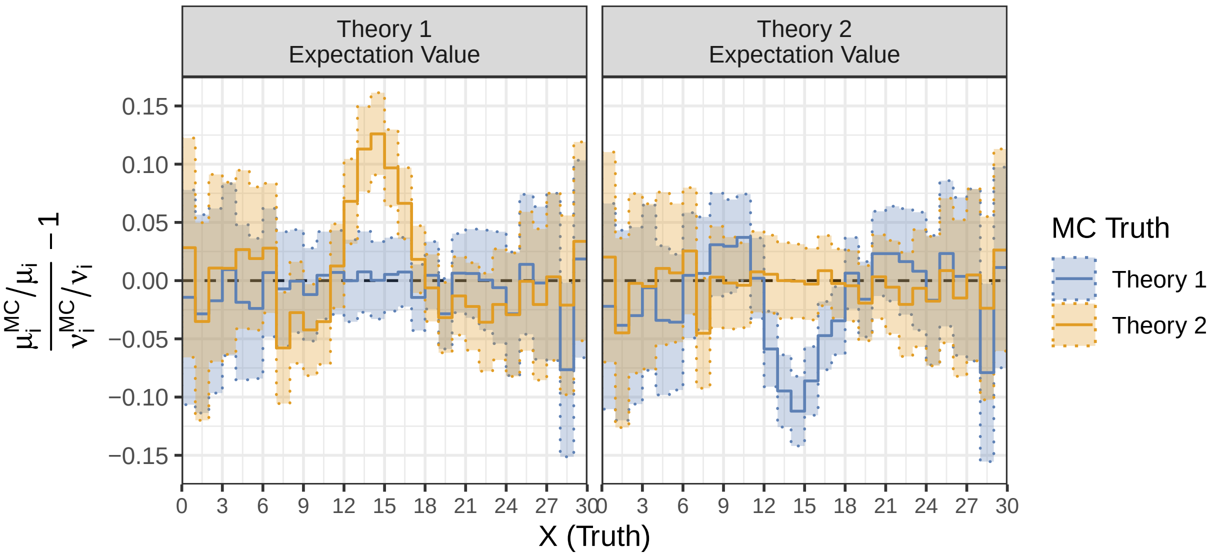
$$\begin{aligned}
\Sigma_{ij}^{\mu} &= \text{Cov}[\hat{\mu}_i, \hat{\mu}_j] \\
&= C_i C_j \text{Cov}[n_i, n_j] \\
&= C_i^2 \delta_{ij} \nu_i \\
&= \left(\frac{\mu_i^{\text{MC}}}{\nu_i^{\text{MC}}}\right)^2 \delta_{ij} \nu_i.
\end{aligned} \tag{21}$$

The expectation value of the estimate can be calculated easily enough as well, and with it the bias

$$\begin{aligned}
\text{Bias}[\hat{\mu}_i] &= E_i[\hat{\mu}_i] - \mu_i \\
&= C_i E[n_i] - \mu_i \\
&= \frac{\mu_i^{\text{MC}}}{\nu_i^{\text{MC}}} \nu_i - \mu_i \\
&= \left(\frac{\mu_i^{\text{MC}}}{\nu_i^{\text{MC}}} - \frac{\mu_i}{\nu_i}\right) \nu_i.
\end{aligned} \tag{22}$$


Theory 1
Theory 2
Counts
1500
1000
500
0
0 3 6 9 12 15 18 21 24 27 30
X (Truth)
Data
Theory 1
Theory 2


Theory 1 Residuals
Theory 2 Residuals
Counts
100
0
−100
−200
0 3 6 9 12 15 18 21 24 27 30
X (Truth)
Data
Theory 1
Theory 2



D R Code

References

[1] Tim Adye. “Unfolding algorithms and tests using RooUnfold”. In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 313–318. DOI: 10.5170/CERN-2011-006.313. eprint: 1105.1160.

[2] Volker Blobel. “Unfolding”. In: *Data analysis in high energy physics: A practical guide to statistical methods*. Ed. by Olaf Behnke et al. Weinheim, Germany: Wiley-VCH, 2013. Chap. 6, pp. 187–226.

[3] Volker Blobel. “Unfolding Methods in Particle Physics”. In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 240–251. DOI: 10.5170/CERN-2011-006.252. eprint: 1105.1160.

[4] Mary L. Boas. *Mathematical Methods in the Physical Sciences*. Third. Wiley, 2005. ISBN: 9780471198260.

[5] G. Casella and R.L. Berger. *Statistical Inference*. Second. Cengage Learning, 2001. ISBN: 9780534243128.

[6] G. Cowan. *Statistical Data Analysis*. Oxford University Press, USA, 1998. ISBN: 9780198501558.

[7] G. D’Agostini. “A Multidimensional unfolding method based on Bayes’ theorem”. In: *Nucl. Instrum. Meth. A* 362 (1995), pp. 487–498. DOI: 10.1016/0168-9002(95)00274-X.

[8] R. Johnson and D.W. Wichern. *Applied Multivariate Statistical Analysis*. Sixth. Pearson, 2007. ISBN: 9780131877153.

[9] Alexander Meister. *Deconvolution Problems in Nonparametric Statistics*. Vol. Lecture Notes in Statistics. Springer, 2009. ISBN: 9783540875567.

[10] Victor M. Panaretos. "A Statistician's View on Deconvolution and Unfolding". In: *PHYSTAT 2011*. Geneva: CERN, 2011, pp. 252–259. DOI: 10.5170/CERN-2011-006.252. eprint: 1105.1160.

[11] W. Rudin. *Functional Analysis*. International Series in Pure and Applied Mathematics. McGraw-Hill, 1991. ISBN: 9780070542365.

[12] Eric W. Weisstein. *Convolution. From MathWorld—A Wolfram Web Resource*. Last visited on 15/2/2022. URL: https://mathworld.wolfram.com/Convolution.html.

[13] Anatoly G. Yagola. "Ill-Posed Problems and Methods for Their Numerical Solution". In: *Optimization and Regularization for Computational Inverse Problems and Applications*. Ed. by Yanfei Wang, A.G. Yagola, and Changchun Yang. Springer, Berlin, Heidelberg, 2011. Chap. 2, pp. 17–34. ISBN: 9783642137419. DOI: https://doi.org/10.1007/978-3-642-13742-6.