Constrained fits with non-Gaussian distributions

R. Frühwirth\textsuperscript{1} and O. Cencic\textsuperscript{2}

\textsuperscript{1} Institute of High Energy Physics, Austrian Academy of Sciences, Vienna, Austria
\textsuperscript{2} Institute of Water Quality, Resources and Waste Management, TU Wien, Vienna, Austria
E-mail: rudolf.fruehwirth@oeaw.ac.at

Abstract. Non-normally distributed data are ubiquitous in many areas of science, including high-energy physics. We present a general formalism for constrained fits, also called data reconciliation, with data that are not normally distributed. It is based on Bayesian reasoning and implemented via MCMC sampling. We show how systems of both linear and non-linear constraints can be efficiently treated. We also show how the fit can be made robust against outlying observations. The method is demonstrated on a couple of examples ranging from material flow analysis to the combination of non-normal measurements. Finally, we discuss possible applications in the field of event reconstruction, such as vertex fitting and kinematic fitting with non-normal track errors.

1. Introduction

Observed data are often in conflict with known conservation laws such as mass, momentum or energy balance. Also, observed data are often not distributed according to a Gaussian or normal distribution. In data reconciliation (DR) contradictions are resolved by statistically adjusting the data, assuming their uncertainty is described by a probability density function, not necessarily a normal one. The method outlined here is based on the following assumptions:

- There are $N$ measured or unmeasured variables that take values in a subset $D \subseteq \mathbb{R}^N$.
- The $I \leq N$ measured variables form an $I$-dimensional random variable with known joint density. The latter is called the prior density.
- The prior density can be either objective, i.e. the model of a measurement process, or subjective, i.e. the formalization of an expert opinion.
- The variables are subject to linear or nonlinear constraints that define a manifold $S \subset \mathbb{R}^N$ of dimension $P < N$.
- The posterior density is obtained by restricting the prior to $S$ and normalizing it to 1.

We illustrate the principle by an example with a simple linear constraint. Assume there are three non-normal observations $x_1$, $x_2$ and $x_3$ with the prior density $f(x_1, x_2, x_3)$. The constraint equation $x_3 = x_1 + x_2$ defines a plane in $\mathbb{R}^3$. The prior density is restricted to this plane and normalized to 1. This gives the posterior density $g(x_1, x_2, x_3)$ and its marginals $g_i(x_i)$, $i = 1, 2, 3$, shown in Fig. 1.
Prior density of \((x_1, x_2, x_3)\)

\[
\begin{array}{c}
\text{Marginal density of } x_1 \\
\text{Marginal density of } x_2 \\
\text{Marginal density of } x_3
\end{array}
\]

\[
\begin{array}{c}
\text{Prior density of } (x_1, x_2, x_3) \\
\text{Restricted density in the plane } x_3 = x_1 + x_2
\end{array}
\]

**Figure 1.** Top left: prior density \(f(x_1, x_2, x_3)\). Top right: prior density \(f(x_1, x_2, x_3)\) restricted to the plane \(x_3 = x_1 + x_2\). Bottom: measured (prior, dotted) and reconciled (posterior, continuous) densities of \(x_1, x_2, x_3\).

2. General linear constraints

2.1. The independence sampler

In the general case the linear constraints are solved explicitely for dependent variables \(u\) as a function of free variables \(w\):

\[
u = Dw + d
\]

For a joint prior \(f(u; w)\) the joint posterior of \(w\) is given by:

\[
\pi(w) \propto f(Dw + d; w)
\]

The explicit calculation of the normalizing constant can be avoided by drawing a random sample from the posterior density \(\pi(w)\) by Markov Chain Monte Carlo (MCMC). The sample is not independent, as each draw depends on the previous one. The corresponding sample of the dependent variables \(u\) is generated via the functional dependence \(u = Dw + d\). From this sample the joint posterior density of all variables can be estimated, along with the correlations, marginal densities, moments and quantiles. See [1] for an introduction into MCMC, including a comprehensive list of references.

The sampler best suited to the problem of DR is the Independence Sampler (IS). There is a natural proposal density, namely the prior density \(p(w)\) of \(w\). The sampler can be summarized as follows.
Independence sampler
1. Set \(i = 1\), choose the sample size \(L\) and the starting value \(w_1\).
2. Draw a proposal value \(\tilde{w}\) from the proposal density \(p(w)\).
3. Compute the acceptance probability \(\alpha\):
   \[
   \alpha(w_i, \tilde{w}) = \min \left(1, \frac{\pi(\tilde{w}) p(w_i)}{\pi(w_i) p(\tilde{w})} \right)
   \]
4. Draw a uniform random number \(u \in [0, 1]\).
5. If \(u \leq \alpha\), accept the proposal and set \(w_{i+1} = \tilde{w}\), otherwise set \(w_{i+1} = w_i\).
6. Increase \(i\) by 1. If \(i < L\), go to 2., otherwise go to 7.
7. Stop sampling.

For more details about the general linear case and additional references, see [2].

2.2. Combination of non-normal observations
The following example shows how to combine non-normal observations by MCMC. Assume that there are three independent measurements \(x_1, x_2, x_3\) of a small nonnegative quantity, for instance a small cross section \(x\). We want to combine the measurements by imposing the constraints \(x_1 = x_2 = x_3\). The measurements are described by three experimental densities \(f_1(x_1), f_2(x_2), f_3(x_3)\) the support of which is restricted to the positive \(x\)-axis. The densities need not be given in closed form, but it must be possible to compute the densities at arbitrary points and to draw random numbers from at least one of them. The procedure is illustrated with the following example priors:

\[
\begin{align*}
  x_1 &\sim \text{Ex}(1.1) \quad \text{Exponential distribution with mean 1.1} \\
  x_2 &\sim \text{Ga}(2,0.5) \quad \text{Gamma distribution with mean 1} \\
  x_3 &\sim \text{TrNorm}(0,1.2,0,\infty) \quad \text{Half normal distribution with } \mu = 0, \sigma = 1.2 \text{ and mean 0.96}
\end{align*}
\]

The posterior marginals under the assumption of independent measurements are shown in Fig. 2. Computing a \(\chi^2\)-statistic does not make sense with non-normal observations, as its distribution is difficult to determine. Instead, a measure of goodness of fit can be obtained by computing the discrepancy between the prior densities and the posterior marginals. Possible measures are, among others, the Kolmogorov-Smirnov distance \(d_{KS}\) and the Hellinger distance \(d_H = \sqrt{1 - BC}\), where BC is the Bhattacharya coefficient. A small acceptance rate of the sampler also indicates a poor fit. Table 1 shows the discrepancies and some key figures of the posterior marginals.

![Figure 2](image-url)

**Figure 2.** Priors and posterior marginals under the assumption of independent measurements.
Table 1. Posterior discrepancies (left) and key figures of the posterior marginals (right) under the assumption of independent measurements.

|        | $X_1$ | $X_2$ | $X_3$ | Mean | Stdev | Median | 95%-Quantile |
|--------|-------|-------|-------|------|-------|--------|--------------|
| $d_{KS}$ | 0.272 | 0.295 | 0.280 | 0.572 | 0.379 | 0.491  | 1.308        |
| $d_H$   | 0.316 | 0.285 | 0.287 |      |       |        |              |

2.3. Correlated observations
Correlations can be introduced via a copula, making it possible to study the influence of correlations between the measurements on the resulting posterior. The posterior marginals with a Gaussian copula and a global correlation $\rho = 0.25$ are shown in Fig. 3. Table 2 shows the discrepancies and some key figures of the posterior marginals. By varying $\rho$ the influence of possible correlations between the measurements can be systematically studied.

![Figure 3](image)

Figure 3. Priors and posterior marginals under the assumption of correlated measurements with all correlations equal to $\rho = 0.25$.

Table 2. Posterior discrepancies (left) and key figures of the posterior marginals (right) under the assumption of correlated measurements with all correlations equal to $\rho = 0.25$.

|        | $X_1$ | $X_2$ | $X_3$ | Mean | Stdev | Median | 95%-Quantile |
|--------|-------|-------|-------|------|-------|--------|--------------|
| $d_{KS}$ | 0.272 | 0.295 | 0.280 | 0.432 | 0.415 | 0.310  | 1.276        |
| $d_H$   | 0.325 | 0.413 | 0.362 |      |       |        |              |

It can be seen that in this example the correlations have a significant influence on the shape and the key characteristics of the posterior.

3. Robust fitting
If there are outlying observations, the acceptance rate of the sampler gets smaller and can even be zero. In this case the fit has to be robustified. The well-known M-estimator [3] can be generalized to non-normal observations. To this end, the residuals

$$r_i = \frac{(E_{pr}[x_i] - E_{po}[x_i])}{\sigma_{pr}[x_i]}$$

are computed for all variables $x_i$, where $E_{pr}[x_i]$ is the prior expectation, $\sigma_{pr}[x_i]$ the prior standard deviation, and $E_{po}[x_i]$ the posterior expectation of $x_i$. From the residuals weights are determined
according to:

$$w(r_i) = \frac{\varphi(r_i) \varphi(0) + \varphi(c)}{\varphi(r_i) + \varphi(c) \varphi(0)}$$

where \( \varphi \) is a symmetric pdf and \( c \) a cut value, resulting in a redescending M-estimator [4]. The density \( \varphi \) should not decay too fast in the tails, in order to avoid the weights becoming numerically equal to 0. The weight as a function of the residual is shown in Fig. 4. In this particular case the cut value is \( c = 2.5 \), and \( \varphi \) is the density of the Student distribution with four degrees of freedom. Once the weights are available, the priors are stretched around their mode by the factor \( s_i = 1/\sqrt{w_i} \) and the estimation of the posterior is repeated. This is iterated until convergence. The iteration can be speeded up if the initial weights are computed in the normal approximation [5].

![Figure 4. The weight as a function of the absolute residual on a linear (left) and a log (right) scale.](image)

### 4. Applications in event reconstruction

Track errors are sometimes non-normal, for instance if tracks are fitted with the Gaussian-sum filter. In this case, vertex fitting and kinematic fitting can be extended to non-normal track errors as outlined above. Whereas momentum conservation gives linear constraints on the momentum vectors, energy conservation and vertex constraints are nonlinear. As usual, nonlinear constraints are linearized at a suitable expansion point. The dependent variables can then be computed via a Newton iteration. The sampler is basically the same as in the linear case. The nonlinear case will be discussed in detail in a forthcoming paper [5].

### 5. References

[1] L. Tierney, *Markov chains for exploring posterior distributions*, The Annals of Statistics 22 (1994) 1701–1762.

[2] O. Cencic and R. Frühwirth, *A general framework for data reconciliation—Part I: linear constraints*, Comp. Chem. Eng. 75 (2015) 196–208.

[3] P. Huber and E.M. Ronchetti, *Robust Statistics*, 2nd ed., John Wiley & Sons, Hoboken NJ, 2009.

[4] F.R. Hampel et al., *Robust Statistics*, John Wiley & Sons, New York, 1986.

[5] O. Cencic and R. Frühwirth, *Data reconciliation of nonnormal observations—Part II: nonlinear constraints, correlated observations and gross error detection by robust estimation*, in preparation.