Constrained fits with non-normal distributions

R. Frühwirth[1,2] and O. Cencic[3]

[1]Institute of High Energy Physics
Austrian Academy of Sciences

[2]Institute of Statistics and Mathematical Methods in Economics
Vienna University of Technology

[3]Institute of Water Quality, Resources and Waste Management
Vienna University of Technology

Connecting the Dots 2015

Berkeley, 9–11 February 2015

Outline

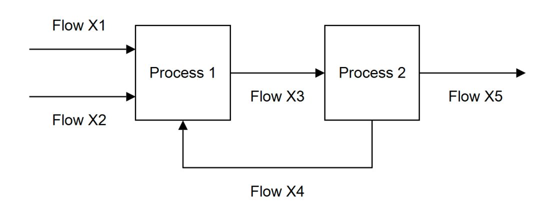
Outline

Motivation

Material flow

- Motivation for this work was **data reconciliation** in material flow
- Trace valuable materials in cycle of production, consumption, waste deposit, recycling,. . .
- Data are often just expert assessments and thus **not normally distributed**, but uniform, triangular, trapezoidal, etc.
- Improve quality by imposing constraints: **conservation of mass**

Example flow chart

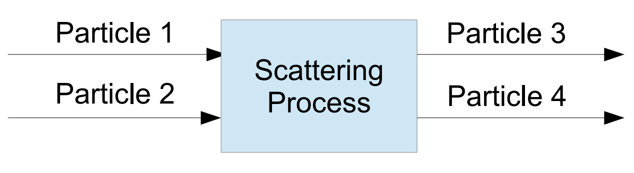


Motivation

Kinematic fit

- Very similar problem
- Track parameters are not always normally distributed, e.g. for electrons fitted with the Gaussian-sum filter
- Constraints are given by **conservation of momentum and energy**

Example process



Motivation

Vertex fit

- Again, track parameters are not always normally distributed
- Constraints are given by **common vertex**
- Constraints may contain **unobserved variables**

Combination of experiments

- Measurements are often not normally distributed, in particular if the measured parameter is non-negative
- Constraints are given by the fact that the **same quantity** is measured
- Feasible only if experiments publish **full marginal likelihood** (posterior density)

Motivation

Vertex fit

- Again, track parameters are not always normally distributed
- Constraints are given by **common vertex**
- Constraints may contain **unobserved variables**

Combination of experiments

- Measurements are often not normally distributed, in particular if the measured parameter is non-negative
- Constraints are given by the fact that the **same quantity** is measured
- Feasible only if experiments publish **full marginal likelihood** (posterior density)

Principle and details

Principle idea

- **Restrict** joint **prior** density of all observed and unobserved variables to the constraint manifold
- Unobserved variables are included by an **uninformative** or **weakly informative** prior
- **Renormalize** restricted **posterior** density to 1

Details

- A detailed description of the algorithm can be found in:

  O. Cencic and R. Frühwirth, A general framework for data reconciliation—Part I: Linear constraints

  Computers and Chemical Engineering, in press
- Available at:

  http://tinyurl.com/CencicFruhwirth

Principle and details

Principle idea

- **Restrict** joint **prior** density of all observed and unobserved variables to the constraint manifold
- Unobserved variables are included by an **uninformative** or **weakly informative** prior
- **Renormalize** restricted **posterior** density to 1

Details

- A detailed description of the algorithm can be found in:

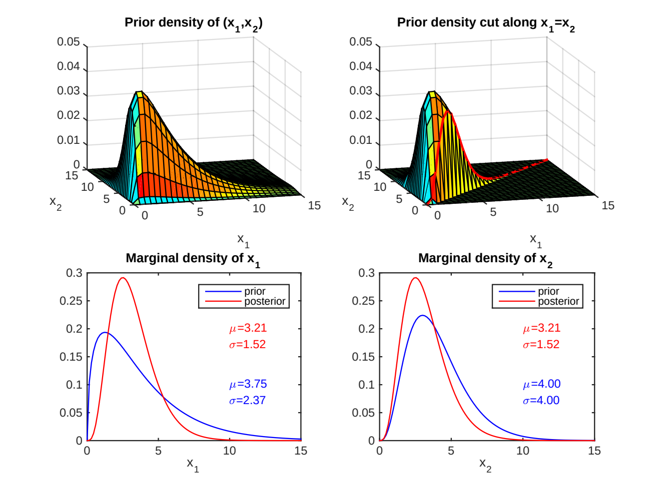
  O. Cencic and R. Frühwirth, A general framework for data reconciliation—Part I: Linear constraints

  Computers and Chemical Engineering, in press
- Available at:

  http://tinyurl.com/CencicFruhwirth

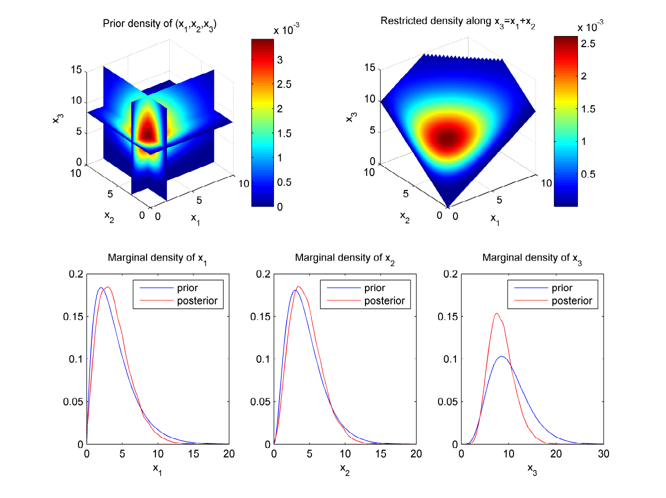
Graphical illustration

A problem in 2D, $x_1 = x_2$



Graphical illustration

A problem in 3D, $x_3 = x_1 + x_2$



Outline

Combination of measurements

Observables

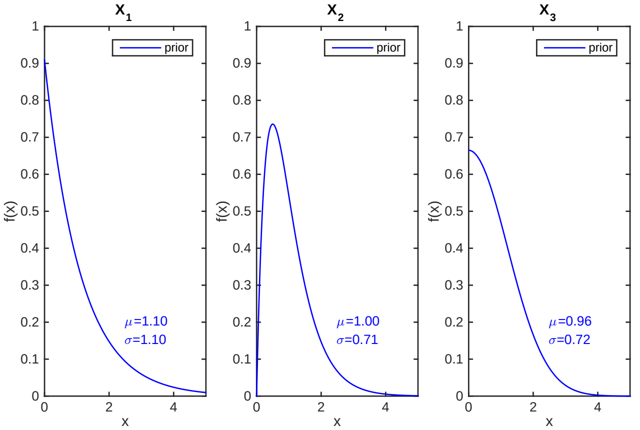
- Three measurements X_1, X_2, X_3 of a small cross section x
- Three **experimental densities** $f_1(x_1), f_2(x_2), f_3(x_3)$, support restricted to the positive axis
- Considered as **prior** densities in this context
- Densities need **not** be given in closed form
- Must be possible to **compute** the densities and to **draw random numbers** from them
- In this simple example:

$X_1 \sim \mathrm{Ex}(1.1)$	Exponential
$X_2 \sim \mathrm{Ga}(2,0.5)$	Gamma
$X_3 \sim \mathrm{TrNorm}(0,1.2,0,\infty)$	Half Normal

- Assume **independence** at the moment, will allow correlations later

Combination of measurements

Prior densities



Combination of measurements

Constraints

- Want to **combine** the measurements by imposing $X_1 = X_2 = X_3$
- If the posteriors are normal densities, the combined measurement is the weighted mean
- If not, we compute the **joint density** of of (X_1, X_2, X_3) **under the constraints** $X_1 = X_2$ and $X_1 = X_3$
- The **constraint manifold** is the line $\boldsymbol{x} = \lambda(1, 1, 1)^{\mathrm{T}}$
- There is one **free variable**, which we choose to be $\boldsymbol{y} = x_1$
- The **dependent variables** $\boldsymbol{z} = (x_2, x_3)^{\mathrm{T}}$ are functions of $\boldsymbol{y}$:

$$\boldsymbol{z} = -\boldsymbol{D}\boldsymbol{y} - \boldsymbol{d} \quad \text{or} \quad \boldsymbol{I}\boldsymbol{z} + \boldsymbol{D}\boldsymbol{y} + \boldsymbol{d} = \boldsymbol{0}$$

with:

$$\boldsymbol{z} = \begin{pmatrix} x_2 \\ x_3 \end{pmatrix}, \quad \boldsymbol{y} = x_1, \quad \boldsymbol{D} = \begin{pmatrix} -1 \\ -1 \end{pmatrix}, \quad \boldsymbol{d} = \begin{pmatrix} 0 \\ 0 \end{pmatrix}$$

Combination of measurements

Derivation of the posterior

- The **joint prior density** of the measurements is given by:

$$f(\boldsymbol{x}) = f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3) = f_\mathrm{f}(\boldsymbol{y}) \cdot f_\mathrm{d}(\boldsymbol{z})$$

- We compute the **posterior density** of $\boldsymbol{x}$ conditional on the constraints
- The posterior is the prior restricted to the constraint manifold, renormalized to 1
- It is easy to show that the posterior of $\boldsymbol{y}$ is given by:

$$\pi(\boldsymbol{y}) = \frac{f_\mathrm{d}(-\boldsymbol{D}\boldsymbol{y} - \boldsymbol{d}) \cdot f_\mathrm{f}(\boldsymbol{y})}{\int f_\mathrm{d}(-\boldsymbol{D}\boldsymbol{y} - \boldsymbol{d}) \cdot f_\mathrm{f}(\boldsymbol{y})\,\mathrm{d}\boldsymbol{y}}$$

- $\pi(\boldsymbol{y})$ is also called the **target density**

Combination of measurements

Normalization of the posterior

- In this example the integral can be computed by numerical integration
- In more complex cases this may get rather tedious, in particular if the dimension of $\boldsymbol{y}$ is large
- Explicit calculation of the integral **can be avoided** by drawing a random sample from the posterior $\pi(\boldsymbol{y})$ by **Markov chain Monte Carlo** (MCMC)
- We use the **Metropolis-Hastings** algorithm for sampling
- Need a **proposal density** $p(\boldsymbol{y})$ to generate values of the **free variables**
- No need to draw from the dependent variables
- **Independence sampler** most suitable in this context

Combination of measurements

Generating the Markov chain

1. Set $i = 1$, choose the sample size L and draw the starting value $\boldsymbol{y}_1$ from $p(\boldsymbol{y})$
2. Draw a proposal value $\hat{\boldsymbol{y}}$ from $p(\boldsymbol{y})$
3. Compute the acceptance probability α by

$$\alpha(\boldsymbol{y}_i, \hat{\boldsymbol{y}}) = \min\left(1, \frac{\pi(\hat{\boldsymbol{y}})\, p(\boldsymbol{y}_i)}{\pi(\boldsymbol{y}_i)\, p(\hat{\boldsymbol{y}})}\right)$$

4. Draw a uniform random number $u \in [0, 1]$
5. If $u \leq \alpha$, accept the proposal and set $\boldsymbol{y}_{i+1} = \hat{\boldsymbol{y}}$, otherwise set $\boldsymbol{y}_{i+1} = \boldsymbol{y}_i$
6. Increase i by 1. If $i < L$, go to 2, otherwise stop sampling

Combination of measurements

Advantages of the independence sampler

- There is a **natural proposal density**:

$$p(\boldsymbol{y}) = f_{\mathrm{f}}(\boldsymbol{y})$$

- There is need for "burn-in"
- If the observations are independent, the acceptance probability has a **very simple form**:

$$\alpha(\boldsymbol{y}, \hat{\boldsymbol{y}}) = \min\left(1, \frac{f_{\mathrm{d}}(-\boldsymbol{D}\hat{\boldsymbol{y}} - \boldsymbol{d})}{f_{\mathrm{d}}(-\boldsymbol{D}\boldsymbol{y} - \boldsymbol{d})}\right)$$

- Sampler can be **SIMDized** by precomputing proposal values and their pdf values
- Sampler can be **parallelized** by generating several independent Markov chains on different cores and combining them afterwards

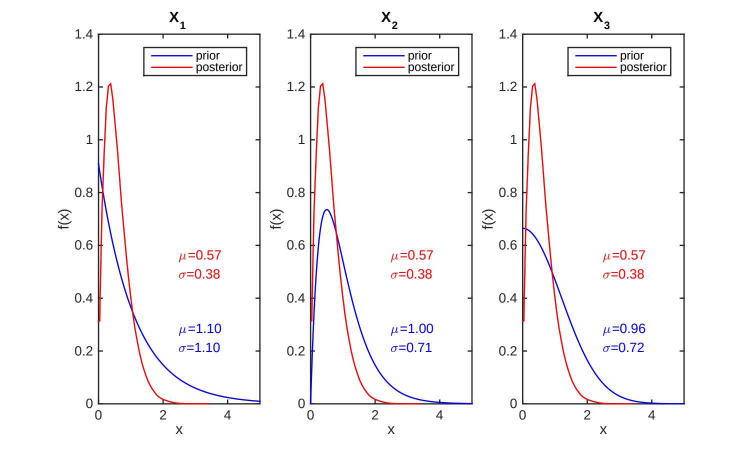
Combination of measurements

Posterior analysis

- The generated chain is a **non-independent** random sample $\boldsymbol{Y}$ from the posterior $\pi(\boldsymbol{y})$ of the free variables
- The corresponding sample $\boldsymbol{Z}$ of the dependent variables $\boldsymbol{z}$ is calculated by $\boldsymbol{z}_i = -\boldsymbol{D}\boldsymbol{y}_i - \boldsymbol{d}, i = 1, \ldots, L$
- Posterior means, variances, correlations, quantiles and credible intervals are estimated from the complete sample $\boldsymbol{X} = (\boldsymbol{Z}; \boldsymbol{Y})$
- Marginal densities are smoothed before graphical representation
- A **measure of goodness** can be obtained by computing the **discrepancy** between the prior densities and the posterior marginals
- Examples:
  - Kolmogorov-Smirnov distance d_{KS}
  - Hellinger distance $d_{\mathrm{H}} = \sqrt{1 - \mathrm{BC}}$ (BC=Bhattacharya coefficient)
- Small acceptance rate also indicates poor fit

Combination of measurements

Posterior density



Combination of measurements

Posterior discrepancies

	X_1	X_2	X_3
d_{KS}	0.272	0.295	0.280
d_{H}	0.316	0.285	0.287

Key properties of posterior

Mean	STD	Median	95%-Quantile
0.5721	0.3794	0.4912	1.3081

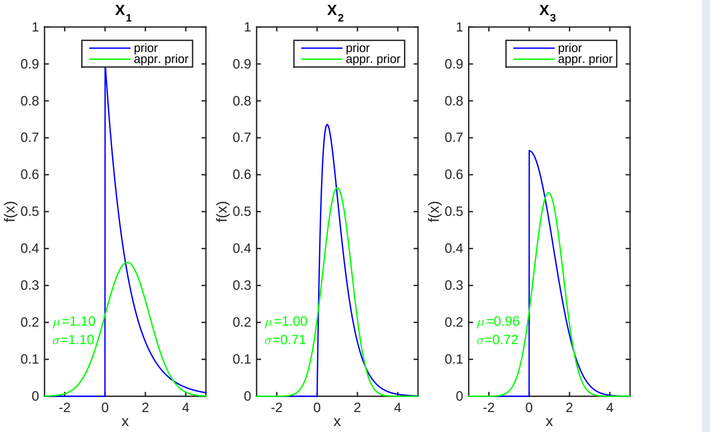
Combination of measurements

Gaussian approximation

- Approximate each prior by a Gaussian with the same mean and variance
- Posterior is then Gaussian too, can be computed explicitely by weighted mean ☺
- Useful cross-check of sampler ☺
- Posterior Gaussian extends into negative axis ☹

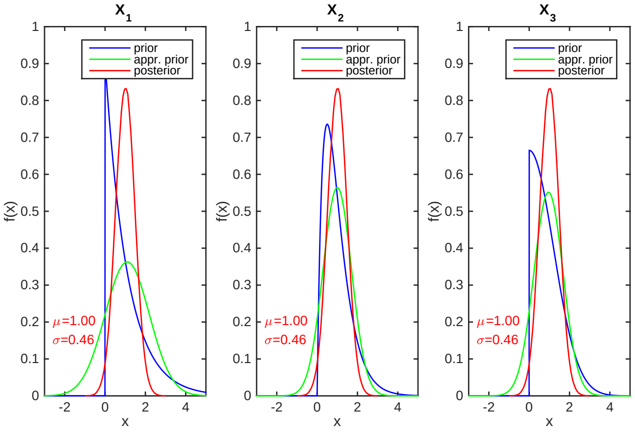
Combination of measurements

Approximating densities



Combination of measurements

Posterior density



Combination of measurements

Posterior discrepancies

	X_1	X_2	X_3
d_{KS}	0.231	0.104	0.126
d_{H}	0.396	0.209	0.220

Key properties of approximate posterior

Mean	STD	Median	95%-Quantile
1.0027	0.4604	1.0027	1.7577

Key properties of exact posterior

Mean	STD	Median	95%-Quantile
0.5721	0.3794	0.4912	1.3081

Combination of measurements

Posterior discrepancies

	X_1	X_2	X_3
d_{KS}	0.231	0.104	0.126
d_{H}	0.396	0.209	0.220

Key properties of approximate posterior

Mean	STD	Median	95%-Quantile
1.0027	0.4604	1.0027	1.7577

Key properties of exact posterior

Mean	STD	Median	95%-Quantile
0.5721	0.3794	0.4912	1.3081

Combination of measurements

Correlated measurements

- The sampler also works with **non-independent** measurements
- Construct a **joint prior with correlations** from the given marginals via a copula distribution
- Have chosen a **Gaussian copula** $g(u_1, u_2, u_3)$ with $\rho = 0.25$
- Joint prior:

$$h(x_1, x_2, x_3) = g(F_1(x_1), F_2(x_2), F_3(x_3)) \cdot f_1(x_1) \cdot f_2(x_2) \cdot f_3(x_3)$$

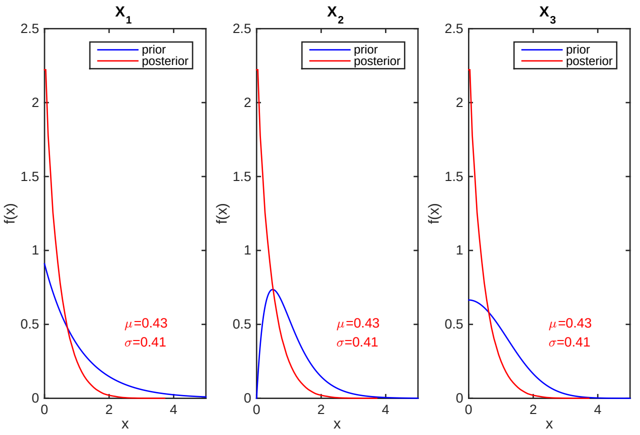
- Target density:

$$\pi(x_1) \propto h(x_1, x_1, x_1)$$

- Proposal density is still $f_1(x_1)$, but computation of α now according to the general formula

Combination of measurements

Posterior densities, $\rho = 0.25$



Combination of measurements

Key properties of posterior with $\rho = 0.25$

Mean	STD	Median	95%-Quantile
0.4320	0.4148	0.3096	1.2759

Key properties of posterior with $\rho = 0$

Mean	STD	Median	95%-Quantile
0.5721	0.3794	0.4912	1.3081

Combination of measurements

Key properties of posterior with $\rho = 0.25$

Mean	STD	Median	95%-Quantile
0.4320	0.4148	0.3096	1.2759

Key properties of posterior with $\rho = 0$

Mean	STD	Median	95%-Quantile
0.5721	0.3794	0.4912	1.3081

Outline

1. Introduction
2. A Simple Example
3. Other Applications
4. Summary and Outlook

Other applications

Extensions of the algorithm

- **Unobserved variables** get an uninformative or weakly informative prior (see paper)
- **Inequality constraints** are reduced to equality constraints by introducing slack variables (see paper)
- **Non-linear constraints** are Taylor-expanded to linear ones (work in progress)

Vertex fit

- Track parameters can have **non-normal track errors**, for instance electrons fitted with Gaussian-sum filter
- Vertex constraints are non-linear
- Vertex position enters with or without prior information
- Vertex can be additionally constrained to a line (e.g. beam line), a plane (e.g. target foil) or a volume (e.g. interaction region)

Other applications

Extensions of the algorithm

- **Unobserved variables** get an uninformative or weakly informative prior (see paper)
- **Inequality constraints** are reduced to equality constraints by introducing slack variables (see paper)
- **Non-linear constraints** are Taylor-expanded to linear ones (work in progress)

Vertex fit

- Track parameters can have **non-normal track errors**, for instance electrons fitted with Gaussian-sum filter
- Vertex constraints are non-linear
- Vertex position enters with or without prior information
- Vertex can be additionally constrained to a line (e.g. beam line), a plane (e.g. target foil) or a volume (e.g. interaction region)

Generalizations

Kinematic fit

- Momentum conservation gives linear constraints, if momentum suitably parameterized
- Energy conservation gives non-linear constraints
- Missing energy can be given an uninformative or an informative prior, and can be restricted to positive values

Outline

Summary and Outlook

Summary

- **Have developed method to impose linear or non-linear constraints on non-normal observations**
- **Constraints may include unobserved variables**
- **Inequality constraints can be dealt with**
- **Independence sampler easy to vectorize and to parallelize**

Outlook

- **Will study non-linear constraints in more detail, in particular performance penalty**
- **Will study gross error detection and robustification in case of poor fit or zero acceptance rate**
- **Intend to apply the method to electrons fitted with GSF: vertex fit, kinematic fit**

Summary and Outlook

Summary

- **Have developed method to impose linear or non-linear constraints on non-normal observations**
- **Constraints may include unobserved variables**
- **Inequality constraints can be dealt with**
- **Independence sampler easy to vectorize and to parallelize**

Outlook

- **Will study non-linear constraints in more detail, in particular performance penalty**
- **Will study gross error detection and robustification in case of poor fit or zero acceptance rate**
- **Intend to apply the method to electrons fitted with GSF: vertex fit, kinematic fit**

Summary and Outlook