Zuse Institute Berlin

Takustrasse 7
D-14195 Berlin-Dahlem
Germany

ILJA KLEBANOV, ALEXANDER SIKORSKI, CHRISTOF SCHÜTTE,
SUSANNA RÖBLITZ

Prior estimation and Bayesian inference from large cohort data sets[1]

[1]This research was carried out in the framework of MATHEON supported by Einstein Foundation Berlin.

Zuse Institute Berlin
Takustrasse 7
D-14195 Berlin-Dahlem

Telefon: 030-84185-0
Telefax: 030-84185-125

e-mail: bibliothek@zib.de
URL: http://www.zib.de

ZIB-Report (Print) ISSN 1438-0064
ZIB-Report (Internet) ISSN 2192-7782

Prior estimation and Bayesian inference from large cohort data sets

Ilja Klebanov, Alexander Sikorski, Christof Schütte, Susanna Röblitz

February 24, 2016

Abstract

One of the main goals of mathematical modeling in systems biology related to medical applications is to obtain patient-specific parameterizations and model predictions. In clinical practice, however, the number of available measurements for single patients is usually limited due to time and cost restrictions. This hampers the process of making patient-specific predictions about the outcome of a treatment. On the other hand, data are often available for many patients, in particular if extensive clinical studies have been performed. Using these population data, we propose an iterative algorithm for constructing an informative prior distribution, which then serves as the basis for computing patient-specific posteriors and obtaining individual predictions. We demonstrate the performance of our method by applying it to a low-dimensional parameter estimation problem in a toy model as well as to a high-dimensional ODE model of the human menstrual cycle, which represents a typical example from systems biology modeling.

Contents

1 Introduction

Bayesian inference is a powerful tool in the field of uncertainty quantification [5], which allows to determine the probability distribution (posterior distribution) of a quantity of interest (or set of parameters) $X \in \mathbb{R}^d$ after measuring certain characteristics, mathematically expressed as mappings $\phi\colon \mathbb{R}^d \to \mathbb{R}^n$ of this quantity. The measurements $Z \in \mathbb{R}^n$ are considered to be perturbed, therefore an error term E is added to the map ϕ:

$$Z = \phi(X) + E,$$

where $E \in \mathbb{R}^n$ is a random variable with the probability density function ρ_E.

The application of Bayesian inference requires the prior distribution of the quantity, reflecting our initial knowledge about it. However, in many cases no comprehensible prior can be assigned, which results in pure guessing and therefore different posterior distributions depending on who made the guess. This unsatisfactory lack of knowledge has lead to many controversial discussions about the reasonability and trustability of Bayesian inference.

This paper aims at estimating the prior distribution, which, naturally, will require further knowledge, in our case the measurements of the characteristics of *several* individuals or, mathematically, ρ_Z-distributed samples $z_1, \dots, z_M$, where ρ_Z denotes the probability density of the measurement Z. This knowledge exists in many applications.

One important application is the prediction of patient-specific treatment success rates based on clinical measurement data and a mathematical model describing the underlying physiological processes. One example are hormonal treatments of the human menstrual cycle, as they are frequently performed in reproductive medicine. In this case, clinical data are available as well as a robust mathematical model, which allows to simulate the cyclic behavior under varying external conditions [4].

Typically, predictions are required for a specific patient in the daily clinical practice, where the number of measurements is limited due to time and cost restrictions. On the other hand, data are often available for many (hundreds or even thousands of) patients, in particular if extensive clinical studies have been performed for the approval of a drug. Using these population data, we propose an iterative algorithm for constructing an informative prior distribution, which then serves as the basis for computing patient-specific posteriors and obtaining individual predictions.

We will set out the theory in Section 2, introduce the notation and derive the scheme for the numerical algorithm. Two different scenarios and their implementation are discussed in Section 3. Corresponding numerical results for a low-dimensional toy example are presented in Section 4. Finally, in Section 5 the algorithm is applied to a high-dimensional parameter estimation problem in an ODE model of the human menstrual cycle, which represents a typical example from systems biology modeling.

2 Theory

First, let us introduce the following

Notation 1.

(1) the letters ρ, π and p will denote probability density functions. π will be used for "prior" densities, p for posterior densities and ρ_X for the density of some (continuous) random variable $X \in \mathbb{R}^d$.

(2) For two continuous random variables $X \in \mathbb{R}^d$ and $Z \in \mathbb{R}^n$ with joint probability density $\rho_{(X,Z)}$,

$$\rho_Z(z \,|\, X = x) := \frac{\rho_{(X,Z)}(x, z)}{\rho_X(x)}$$

denotes the probability density of $(Z \,|\, X = x)$ *(*Z *conditioned on the event* $X = x$*), as long as* $\rho_X(x) \neq 0$.
In our case, where $Z = \phi(X) + E$ *is a perturbed measurement, this is just the likelihood function given by*

$$\rho_Z(z \,|\, X = x) := \rho_E\left(z - \phi(x)\right).$$

(3) It follows, that the probability density of Z *and the posterior density of* X *given the measurement* $Z = z$ *are given by*

$$\rho_Z(z) = \int \rho_{(X,Z)}(x, z)\,\mathrm{d}x = \int \rho_Z(z \,|\, X = x)\,\rho_X(x)\,\mathrm{d}x,$$
$$p^z(x) := \rho_X(x \,|\, Z = z) = \frac{\rho_{(X,Z)}(x, z)}{\rho_Z(z)} = \frac{\rho_Z(z \,|\, X = x)\,\rho_X(x)}{\int \rho_Z(z \,|\, X = \tilde{x})\,\rho_X(\tilde{x})\,\mathrm{d}\tilde{x}},$$

latter being the continuous version of Bayes' rule.

(4) If X *had the probability density* π *instead of* ρ_X*, the probability density of* Z *and the posterior density of* X *given a measurement* $Z = z$ *would have the form and will be denoted by*

$$\rho_{Z,\pi}(z) := \int \rho_Z(z \,|\, X = x)\,\pi(x)\,\mathrm{d}x,$$
$$p^z_\pi(x) := \frac{\rho_Z(z \,|\, X = x)\,\pi(x)}{\rho_{Z,\pi}(z)}.$$

Following this notation, the actual probability density of Z *is given by* $\rho_Z = \rho_{Z,\rho_X}$*. We will sometimes use the notation* $\rho_{Z,\pi}$ *in the more general case* $\pi \in L^1(\mathbb{R}^d)$.

In order to reproduce the probability density ρ_X of the parameters from the measurements $z_1, \ldots, z_M$, we will recursively apply the iteration

$$\pi_{n+1}(x) = (\Psi\pi_n)(x), \quad \text{where} \tag{2.1}$$
$$(\Psi\pi)(x) := \int p^z_\pi(x)\,\rho_Z(z)\,\mathrm{d}z = \pi(x) \int \frac{\rho_Z(z \,|\, X = x)}{\rho_{Z,\pi}(z)}\,\rho_Z(z)\,\mathrm{d}z. \tag{2.2}$$

This iteration is motivated by the observation that the correct probability density ρ_X of X is a fixed point of Ψ:

$$(\Psi\rho_X)(x) = \rho_X(x) \int \frac{\rho_Z(z \,|\, X = x)}{\rho_Z(z)}\,\rho_Z(z)\,\mathrm{d}z = \rho_X(x). \tag{2.3}$$

More precisely, the following statement holds:

Proposition 2. Let $\pi \in L^1(\mathbb{R}^d)$ be a globally supported probability density function. Then the following two statements are equivalent:

(i) $\rho_{Z,\pi} = \rho_Z$,
(ii) $\Psi\pi = \pi$.

Proof. The proof for (i)⇒(ii) goes analogously to (2.3). For (ii)⇒(i) define the following subspace of $L^1(\mathbb{R}^n)$

$$\mathcal{E} = \left\{\rho_{Z,f} \mid f \in L^1(\mathbb{R}^d)\right\} \subseteq L^1(\mathbb{R}^n)$$

with weighted L^2 inner products

$$\langle \rho_{Z,f_1}, \rho_{Z,f_2} \rangle_\pi := \int_{\mathbb{R}^n} \frac{\rho_{Z,f_1}(z)\,\rho_{Z,f_2}(z)}{\rho_{Z,\pi}(z)}\,\mathrm{d}z,$$

where $\pi \in L^1(\mathbb{R}^d)$ denotes a globally supported probability density function.

We can formulate the following chain of implications:

$$\begin{aligned}
\text{(ii)} &\implies \forall x: \quad \int \frac{\rho_Z(z)}{\rho_{Z,\pi}(z)}\, \rho_Z(z \,|\, X = x)\, \mathrm{d}z = 1 \\
&\implies \forall x: \quad \int \left(1 - \frac{\rho_Z(z)}{\rho_{Z,\pi}(z)}\right) \rho_Z(z \,|\, X = x)\, \mathrm{d}z = 0 \\
&\implies \forall x: \quad \int \frac{\rho_{Z,\pi}(z) - \rho_Z(z)}{\rho_{Z,\pi}(z)}\, \rho_Z(z \,|\, X = x)\, \mathrm{d}z = 0 \\
&\implies \int (\pi - \rho_X)(x) \int \frac{\rho_{Z,\pi}(z) - \rho_Z(z)}{\rho_{Z,\pi}(z)}\, \rho_Z(z \,|\, X = x)\, \mathrm{d}z\, \mathrm{d}x = 0 \\
&\implies \int \frac{\rho_{Z,\pi}(z) - \rho_Z(z)}{\rho_{Z,\pi}(z)}\, \rho_{Z,(\pi-\rho_X)}(z)\, \mathrm{d}z = 0 \\
&\implies \langle \rho_{Z,\pi} - \rho_Z, \rho_{Z,\pi} - \rho_Z \rangle_\pi = 0,
\end{aligned}$$

which implies (i) using the positive definiteness of the inner product. □

Therefore, (2.1) is a fixed point iteration and our numerical experiments show, that it truly converges to some π_∞ that fulfills $\rho_{Z,\pi_\infty} = \rho_Z$. In case of identifiability, see Section 2.1 below, we even have $\pi_\infty = \rho_X$. Further, this fixed point iteration is easy to approximate numerically in the case of a high number of measurements M, see Section 2.2.

2.1 Identifiability and convergence

Identifiability of a parameter $X \in \mathbb{R}^d$ is always an issue when it comes to convergence results of parameter estimators with a growing number of measurements, like convergence of the ML and MAP estimators or the Bernstein-von Mises theorem (see e.g. [1], [2]). If the model ϕ is non-injective (and the error-distribution of E is independent of the parameters as in our case), there is no chance to give precise statements about the parameter X from just measuring $Z = \phi(X) + E$ (no matter how many realizations), since different parameters can lead to the same distribution of the measurements!

Therefore, one either has to add further assumptions on the model ϕ, or, more generally, the conditional densities $\rho_Z(\bullet \,|\, X = x)$, or settle with weaker, less precise statements. We will follow both paths.

The proper identifiability assumption in our case differs from the usual one for the following reason:

The above mentioned results apply to the case of making *several* measurements for *one* individual with *one* true parameter $x^* \in \mathbb{R}^d$ and we hope for convergence of the estimators to the true parameter x^* (in some probabilistic sense).

In our case we have *one* measurement for each of *several* individuals with *varying*, in fact ρ_X-distributed parameters and we hope for convergence of the approximate probability density to the true density ρ_X (in L^1-sense).

The usual and weakest possible assumption for the identifiability [1] of x^*,

$$\|\rho_Z(\bullet \,|\, X = x) - \rho_Z(\bullet \,|\, X = x^*)\|_{L^1(\mathbb{R}^n)} = 0 \iff x = x^*, \tag{2.4}$$

will therefore be replaced by an assumption for the identifiability of ρ_X:

$$\|\rho_{Z,\pi} - \rho_{Z,\rho_X}\|_{L^1(\mathbb{R}^n)} = 0 \iff \pi = \rho_X. \tag{2.5}$$

[1] Often a slightly stronger assumption is used:

$$\|\rho_Z(\bullet \,|\, X = x_1) - \rho_Z(\bullet \,|\, X = x_2)\|_{L^1(\mathbb{R}^n)} = 0 \iff x_1 = x_2.$$

In fact, if the distribution of the measurements would not change when replacing the actual prior ρ_X by the prior π, we have no way of telling which of these two distributions the measurements stem from, making (2.5) the weakest possible identifiability assumption for ρ_X.

Remark 3. From a more abstract point of view, we can consider ρ_X as a parameter that has to be identified from the measurements $(z_1, \dots, z_M)$ and lies in the set $\mathcal{P}$ of all possible priors π. Viewing the parameter we seek as a random variable Π, the likelihood of each measurement is given by

$$\rho_Z(z \,|\, \Pi = \pi) = \int \rho_Z(z \,|\, X = x)\, \pi(x)\, \mathrm{d}x = \rho_{Z,\pi}(z).$$

Therefore, the identifiability assumption (2.5) has exactly the same form as assumption (2.4) for the proper choice of the unknown parameter, namely Π.

If assumption (2.5) is fulfilled, we observe a convergence of π_n to ρ_X in L^1:

$$\|\pi_n - \rho_X\|_{L^1(\mathbb{R}^d)} \xrightarrow{n\to\infty} 0.$$

However, we are still lacking a proof.

If assumption (2.5) is not fulfilled, we have to enforce it by restricting ourselves to equivalence classes of densities with respect to the equivalence relation

$$\pi \sim \pi' \iff \|\rho_{Z,\pi} - \rho_{Z,\pi'}\|_{L^1(\mathbb{R}^n)} = 0.$$

Note, that the set of equivalence classes $L^1(\mathbb{R}^d)/\!\sim$ is, in fact, the quotient space $L^1(\mathbb{R}^d)/\ker(\psi)$ emerging from the linear map

$$\psi\colon L^1(\mathbb{R}^d) \to L^1(\mathbb{R}^n), \qquad \pi \mapsto \rho_{Z,\pi} = \int \rho_Z(\bullet \,|\, X = x)\, \pi(x)\, \mathrm{d}x.$$

Therefore, $L^1(\mathbb{R}^d)/\!\sim$ inherits the L^1-norm and L^1-distance via

$$\|[\pi]\|_{L^1} = \inf_{\pi' \in [\pi]} \|\pi'\|_{L^1} \qquad \text{and} \qquad \|[\pi_1] - [\pi_2]\|_{L^1} = \inf_{\substack{\pi_1' \in [\pi_1] \\ \pi_2' \in [\pi_2]}} \|\pi_1' - \pi_2'\|_{L^1}$$

and we can choose from the following two definitions for the convergence of π_n to π_∞:

$$\pi_n \xrightarrow[Z]{n\to\infty} \pi_\infty \;:\iff\; \|[\pi_n] - [\pi_\infty]\|_{L^1} \xrightarrow{n\to\infty} 0$$

or

$$\pi_n \xrightarrow[Z]{n\to\infty} \pi_\infty \;:\iff\; \|\rho_{Z,\pi_n} - \rho_{Z,\pi_\infty}\|_{L^1} \xrightarrow{n\to\infty} 0.$$

Since, as already mentioned, we are lacking a proof of convergence, it is unclear which of these two definitions will turn out to be more meaningful.

However, we can already say that the second definition yields a weaker form of convergence as stated by the following lemma:

Lemma 4. Let $\pi_1, \pi_2 \in L^1(\mathbb{R}^d)$. Then, in the above notation,

$$\|\rho_{Z,\pi_1} - \rho_{Z,\pi_2}\|_{L^1} \leq \|[\pi_1] - [\pi_2]\|_{L^1}\ .$$

Proof. For each $\pi_1' \in [\pi_1]$ and $\pi_2' \in [\pi_2]$ we have

$$\begin{aligned}
\|\rho_{Z,\pi_1} - \rho_{Z,\pi_2}\|_{L^1} &= \left\|\rho_{Z,\pi_1'} - \rho_{Z,\pi_2'}\right\|_{L^1} \\
&= \int \left| \int \rho_Z(z \,|\, X = x)\, (\pi_1'(x) - \pi_2'(x))\, \mathrm{d}x \right| \mathrm{d}z \\
&\leq \int \int \rho_Z(z \,|\, X = x)\, |\pi_1'(x) - \pi_2'(x)|\, \mathrm{d}x\, \mathrm{d}z \\
&= \int |\pi_1'(x) - \pi_2'(x)| \underbrace{\int \rho_Z(z \,|\, X = x) \mathrm{d}z}_{=1}\, \mathrm{d}x \\
&= \|\pi_1' - \pi_2'\|_{L^1},
\end{aligned}$$

which proves the claim. □

2.2 Numerical realization

Since the measurements $z_1, \ldots, z_M$ are ρ_Z-distributed, the evaluation of the integral in (2.2) can be realized by a Monte Carlo approximation:

$$(\Psi\pi)(x) = \int p_\pi^z(x)\, \rho_Z(z)\, \mathrm{d}z \approx \frac{1}{M} \sum_{m=1}^{M} p_\pi^{z_m}(x). \tag{2.6}$$

In the first iteration step (going from π_0 to π_1) we will therefore produce a $p_{\pi_0}^{z_m}$-distributed sampling $(x_1^{(m)}, \ldots, x_L^{(m)})$ for each $m = 1, \ldots, M$ by the Metropolis-Hastings algorithm. Merging them will produce a π_1-distributed sampling

$$\mathcal{X} = \{x_1, \ldots, x_k\} = \bigcup_{m=1}^{M} \left\{x_1^{(m)}, \ldots, x_L^{(m)}\right\}.$$

For the Metropolis-Hastings algorithm it is necessary that π_0 is given in an explicit form, which we will assume from now on. π_1, however, is *not* given explicitly, but by the sampling $\mathcal{X}$. This form is often more suitable, e.g. for the evaluation of expectation values of certain quantities of interest $q \in L^1(\pi_1)$:

$$\mathbb{E}_{\pi_1}[q] = \int q(x)\, \pi_1(x)\, \mathrm{d}x \approx \sum_{k=1}^{K} w_k^{(1)}\, q(x_k)$$

with equal weights $w_k^{(1)} = 1/K$.

However, this forms hinders us to produce a Monte Carlo sampling of π_2 (and further iteration steps), since we cannot evaluate the density π_1. This can be overcome by either using kernel density estimation or similar in order to obtain approximations of the density values or by keeping the sampling $\mathcal{X}$ constant and adapting the weights $w^{(n)} = \left(w_k^{(n)}\right)_{k=1,\ldots,K}$. We will follow the latter approach, which is nothing else than an application of the importance sampling technique. Since

$$\pi_{n+1}(x) = (\Psi\pi_n)(x) \approx \frac{\pi_n(x)}{M} \sum_{m=1}^{M} \frac{\rho_Z(z_m \,|\, X = x)}{\int \rho_Z(z_m \,|\, X = \tilde{x})\, \pi_n(\tilde{x})\, \mathrm{d}\tilde{x}},$$

and using the abbreviation $L_{mk} := \rho_Z(z_m \,|\, X = x_k)$ for the likelihoods, the weights will be adapted in the following way:

$$\boxed{w^{(n+1)} := \overline{\Psi}\big(w^{(n)}\big) := \left(\frac{w_k^{(n)}}{M} \sum_{m=1}^{M} \frac{L_{mk}}{\sum_{j=1}^{K} w_j^{(n)} L_{mj}} \right)_{k=1,\ldots,K}}.$$

This choice relies on the following observation (this time $q \in L^1(\pi_{n+1})$):

$$\begin{aligned}
\mathbb{E}_{\pi_{n+1}}[q] &= \int q(x)\,\pi_{n+1}(x)\,\mathrm{d}x \\
&= \frac{1}{M}\sum_{m=1}^{M} \frac{\int q(x)\,\rho_Z(z_m \,|\, X=x)\,\pi_n(x)\,\mathrm{d}x}{\int \rho_Z(z_m \,|\, X=\tilde{x})\,\pi_n(\tilde{x})\,\mathrm{d}\tilde{x}} \\
&\approx \frac{1}{M}\sum_{m=1}^{M} \frac{\sum_{k=1}^{K} w_k^{(n)}\,q(x_k)\,L_{mk}}{\sum_{j=1}^{K} w_j^{(n)} L_{mj}} \\
&= \sum_{k=1}^{K} q(x_k)\,\frac{w_k^{(n)}}{M}\sum_{m=1}^{M} \frac{L_{mk}}{\sum_{j=1}^{K} w_j^{(n)} L_{mj}} \\
&= \sum_{k=1}^{K} w_k^{(n+1)}\,q(x_k).
\end{aligned} \tag{2.7}$$

Alternatively, from the importance sampling point of view, we have:

$$\mathbb{E}_{\pi_{n+1}}[q] = \int q(x)\,\frac{\pi_{n+1}}{\pi_n}(x)\,\pi_n(x)\,\mathrm{d}x \approx \sum_{k=1}^{K} w_k^{(n)}\,q(x_k)\,\frac{\pi_{n+1}}{\pi_n}(x_k) \qquad \text{and}$$

$$\frac{\pi_{n+1}}{\pi_n}(x_k) = \int \frac{\rho_Z(z \,|\, X=x_k)}{\int \rho_Z(z \,|\, X=\tilde{x})\,\pi_n(\tilde{x})\,\mathrm{d}\tilde{x}}\,\rho_Z(z)\,\mathrm{d}z \approx \frac{1}{M}\sum_{m=1}^{M} \frac{L_{m,k}}{\sum_{j=1}^{K} w_j^{(n)}\,L_{m,j}}\,.$$

Remark 5. The observed convergence of the fixed point iteration (2.1)–(2.2) is not inherited exactly by its Monte Carlo approximation (2.6) (for a fixed number of measurements M). After a large number of iterations n, the priors π_n become peaked at the preimages of the measurements. Therefore, in practice, the iteration has to be stopped after a reasonable number of iterations, depending on the number of measurements M. The larger M, the more iteration steps can be performed without encountering this problem.

3 Resulting algorithm

Given the data $\mathcal{Z} = \{z_1, \ldots, z_M\} \subseteq \mathbb{R}^n$ of M patients we will follow two scenarios:

(A) No diagnoses have been made.

(B) The patients have been diagnosed with diseases/sicknesses $s_1, \ldots, s_L$ (for simplicity, we will assume that each patient has exactly one disease):

$$\mathcal{Z} = \{z_1, \ldots, z_M\} = \bigsqcup_{l=1}^{L} \mathcal{Z}^{(l)}, \qquad \mathcal{Z}^{(l)} = \{z_1^{(l)}, \ldots, z_{M_l}^{(l)}\} \qquad (M = \sum_{l=1}^{L} M_l).$$

Approach for Scenario (A):

- Starting with a non-informative prior π_0 we construct an informative prior $\pi_\infty := \lim_{n\to\infty} \pi_n$ by the fixed point iteration discussed in Section 2.2

$$\pi_{n+1} = \frac{1}{M}\sum_{m=1}^{M} p_{\pi_n}^{z_m},$$

  given by the sampling $\mathcal{X} = \{x_1, \ldots, x_K\} \sim \pi_1$ and weights $w^{(\infty)} = \lim_{n\to\infty} w^{(n)}$.
- For each patient we compute his or her personal posterior $p_{\pi_\infty}^{z_m}$, $m = 1, \ldots, M$, with respect to this prior π_∞.

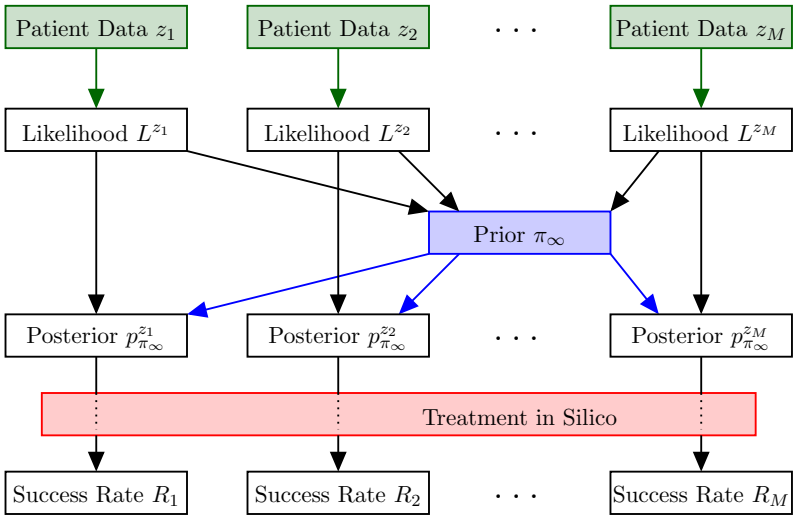


Figure 1: Algorithmic scheme for the computation of patient-specific parametrizations and predictions of individual treatment success rates.

- We can compute the success rate of a treatment for each patient or even a new patient with data z_* using the sampling $\mathcal{X}$ and adapted weights $v^* = (v_1^*, \dots, v_K^*)$ given by

$$v_k^* = \frac{w_k^{(\infty)} L_{*,k}}{\sum_{j=1}^K w_j^{(\infty)} L_{*,j}}\,.$$

  This reweighting is justified analogously to (2.7).
  The success rate R_* can now be approximated via

$$R_* = \int r(x)\, p^{z_*}_{\pi_\infty}(x)\,\mathrm{d}x \approx \frac{1}{K}\sum_{k=1}^K v_k^*\, r(x_k), \qquad \text{where}$$

$$r(x) = \begin{cases} 1 & \text{if the treatment, given the parameters } x, \text{ is successful,} \\ 0 & \text{otherwise.} \end{cases}$$

Approach for Scenario (B):
If the patients are diagnosed with diseases $s_1, \dots, s_L$ and the number of patients M_l is large for each disease s_l, then this extra information can be used by applying the procedure described in (A) to each subset $\mathcal{Z}^{(l)}$ separately in order to obtain more precise results.

4 Toy example

We buy two boxes (representing two diseases) of springs, the first containing 10000 springs with stiffness $K_1 = 15\,N/m$, the second containing 10000 springs with stiffness $K_2 = 30\,N/m$. The springs are of a low quality and their actual stiffness varies from the nominal value with a standard deviation of 15% (we assume a normal distribution for each box).

Once we arrive at home, we realize that the boxes are not labeled and that we already forgot the values K_1 and K_2 as well as the standard deviation.

We will discuss two scenarios:

(A) We mix up the springs by putting all of them into one big box (no diagnosis for each spring).

(B) We keep them in the two separate boxes (the springs are diagnosed with diseases s_1 or s_2, depending on the box they come from).

In order to determine the stiffness of a single spring, we perform the following experiment (see Figure 2):

- We fix one end of the spring and put a mass $m = 700\text{g}$ to the other end.
- After compressing it by 10cm, we let it swing. Applying Hooke's law this results in the following ODE

$$x''(t) = -\frac{K}{m}x(t) \qquad , \qquad x(0) = -10\text{cm}.$$

- We measure its amplitude at times $t_1 = 1\text{s}$ and $t_2 = 1.7\text{s}$. Therefore the model $\phi : \mathbb{R} \to \mathbb{R}^2$ and the measurement $Z \in \mathbb{R}^2$ are given by

$$\phi(K) = (x(t_1), x(t_2))^\intercal \qquad , \qquad Z = \phi(K) + (E_1, E_2)^\intercal,$$

where the measurement errors $E_1, E_2 \in \mathbb{R}$ are assumed to be standard normal distributed with mean 0 and a standard deviation of 5cm. Denoting the probability density of the error $E = (E_1, E_2)$ by $\rho_E \colon \mathbb{R}^2 \to \mathbb{R}$, the likelihoods are given by $\rho_Z(z \,|\, X = K) = \rho_E(z - \phi(K))$.

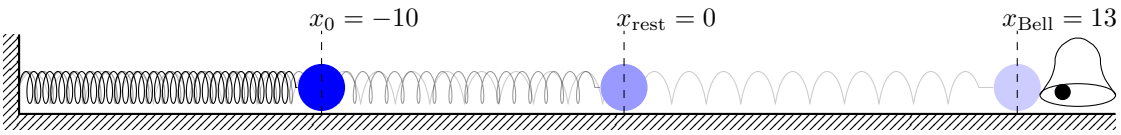


Figure 2: Experimental arrangement for the toy example described above.

We implemented the fixed point iteration described in Section 2, starting with a "non-informative" prior π_0, which we chose as the Beta(2, 4)-distribution on $[0, 110]$. The result is shown in Figure 3. Though π_∞ appears to be a better approximation to the true prior ρ_X, the approximation is far from perfect. The lack of convergence $\pi_n \nrightarrow \rho_X$ arises from the non-identifiability of the system (as defined in Section 2.1). However, the approximated evidence $\rho_{Z,\pi}$ converges to the true evidence ρ_Z, as can be seen in Figures 4 and 5. As argued in Section 2.1, this is the best we can hope to achieve.

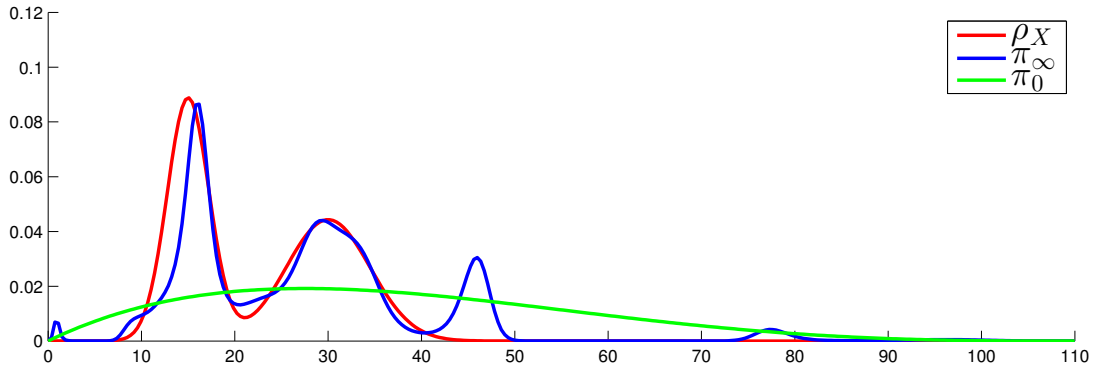


Figure 3: The true prior ρ_X, the non-informative prior π_0 and the approximation π_∞ of the true prior (after 100 iteration steps) plotted over the stiffness K.

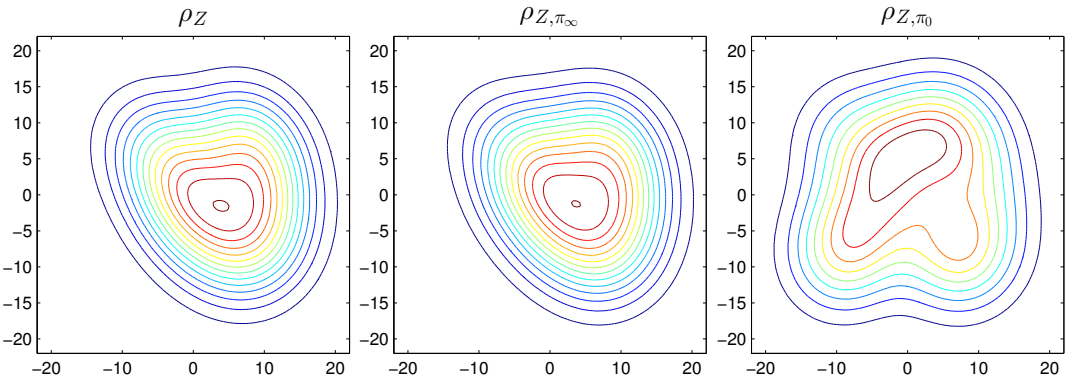


Figure 4: The evidence $\rho_{Z,\pi}$ induced by the three different priors ρ_X, π_∞ and π_0 plotted over the measurements $(z_1, z_2) \in \mathbb{R}^2$. Even though π_∞ differs substantially from the true prior ρ_X, as can be seen in Figure 3, the evidences they induce are very similar.

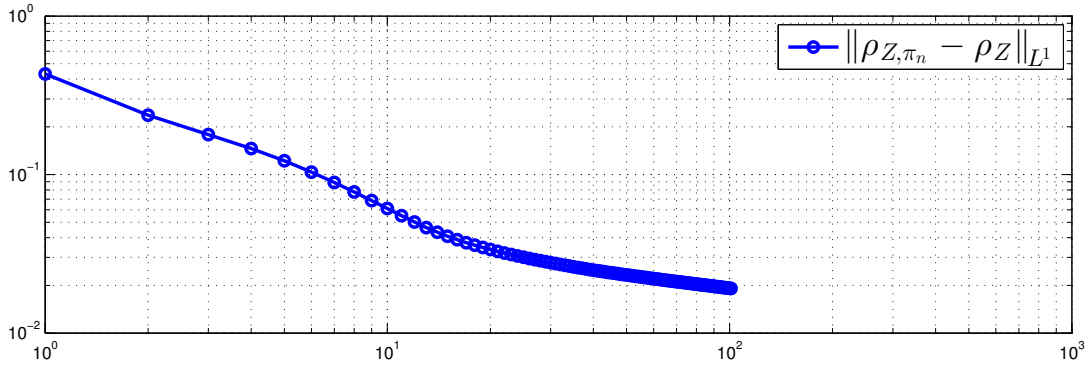


Figure 5: Convergence of the approximated evidence ρ_{Z,π_n} to the true evidence $\rho_Z = \rho_{Z,\rho_X}$. The L^1-error is plotted over the number of iterations n.

The "treatment procedure" will be modeled by hitting the mass in positive x-direction with several pulses at certain times given by the force $F(t)$ plotted in Figure 6, which results in the following perturbed ODE:

$$x''(t) = -\frac{K}{m}x(t) + F(t) \qquad , \qquad x(0) = -10\text{cm}.$$

The treatment will be considered successful, if the mass hits the bell located at $x_{\text{Bell}} = 13\text{cm}$ within ten seconds.

We are interested in the success rate of this treatment for a each spring separately computed only from their performance in the first experiment!

We implemented the procedure described in Section 3 and computed the success rate R_* for many randomly chosen springs with measurements z_* using the following priors:

- the true prior ρ_X,
- our approximation π_∞ to the true prior,
- the non-informative initial guess $\pi_0 = \text{Beta}(2, 4)$.

In Figure 7, R_* is plotted over z_* for all three priors. One can clearly see the improvement in approximating the true success rate (stemming from the true prior ρ_X) when we switch from π_0 to π_∞. But again the non-identifiability of the system results in a lack of convergence to the true success rate.

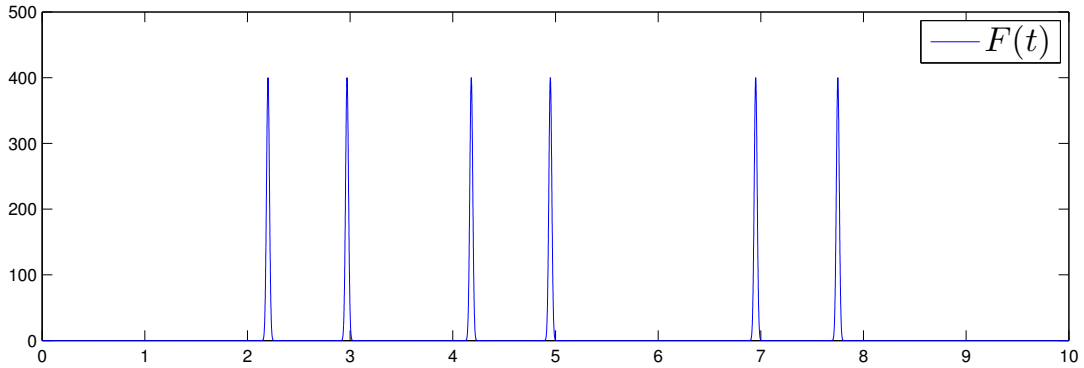


Figure 6: The force F used for modeling the treatment procedure plotted over time t.

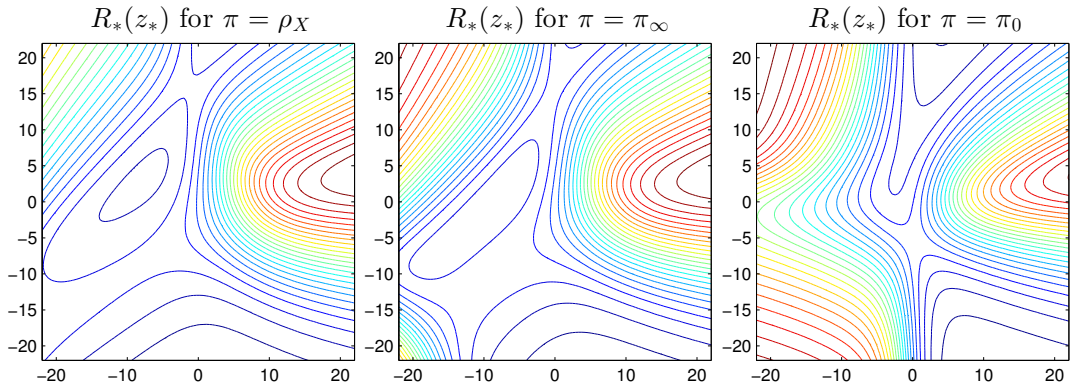


Figure 7: The resulting success rate R_* for three different prior distributions plotted over the measurement $z_* \in \mathbb{R}^2$.

5 Parameter estimation in a large ODE model

As model system, we consider a model for the human menstrual cycle, named GynCycle [4]. This model consists of 33 ordinary differential equations (ODEs) and 114 parameters. It had been calibrated previously with time-series data of blood concentrations for four hormones from 12 patients during the unperturbed cycle and during treatment (dose-response experiments). Using deterministic, local optimization (an error-oriented Gauss-Newton method), only 63 out of 114 parameters could be identified from the given data. The remaining parameters kept their values from previous versions of the model. In the following, we will denote these parameter values as nominal values. The model is currently used to make patient-specific predictions about the outcome of treatment strategies in reproductive medicine[2]. Hence, quantification of uncertainty in these predictions is of utmost importance.

We got access to additional measurement values of 39 woman for the four hormones LH, FSH, E2 and P4 during normal cycles[3]. These data are sparse or incomplete in the sense that measurements were not taken on all cycle days, resulting in about 15 measurement time points per patient and hormone. Our approach, however, is flexible enough to handle such a data situation.

Based on these data, our aim is to estimate the prior distribution for 82 out of the 114 parameters Θ (Hill exponents have been fixed), denoted by π_0^Θ, as well

[2]EU project PAEON-Model Driven Computation of Treatments for Infertility Related Endocrinological Diseases, project number 600773.

[3]Courtesy of Dorothea Wunder, University Hospital of Lausanne.

as for the initial values Y_0 of the 33 model species, $\pi_0^{Y_0}$, resulting in a total of 115 dimensions: $X = (Y_0, \Theta) \in \mathbb{R}^{115}$.

As initial guess for the prior π_0^Θ of model parameters, we have chosen uniform distributions on the intervals between zero and five times the nominal parameter values as non-informative approach. For $\pi_0^{Y_0}$ we chose a mixture distribution of independent normals around daily values of a reference cycle computed with the nominal parameters, i.e.

$$\pi_0^{Y_0} = \frac{1}{31} \sum_{t=0}^{30} G\left[y_{\text{ref}}(t), \Sigma\right],$$

where $G[m, C]$ denotes the Gaussian function with mean m and covariance matrix C, $y_{\text{ref}} : \mathbb{R} \to \mathbb{R}^{33}$ is the reference solution over one menstrual cycle and Σ is a diagonal matrix consisting of the squared standard deviations of each species, respectively,

$$\Sigma = \text{diag}(\sigma_1^1, \ldots, \sigma_{33}^2), \qquad \sigma_j^2 = \frac{1}{30} \sum_{t=0}^{30} |y_{\text{ref},j}(t) - \overline{y}_{\text{ref},j}|^2.$$

The total prior $\pi_0 \colon \mathbb{R}^{115} \to \mathbb{R}$ is build up from $\pi_0^{Y_0}$ and π_0^Θ under the assumption that Y_0 and Θ are independent,

$$\pi_0(y_0, \theta) := \pi_0^{Y_0}(y_0)\, \pi_0^\Theta(\theta).$$

The likelihood for specific measurements $z \in \mathbb{R}^{4\times 31}$ is modeled via relative Gaussian errors with a (relative) standard deviation $\sigma = 10\%$,

$$\rho_Z(z \,|\, X = x) \propto \exp\left(-\frac{d(\Phi(x), z)^2}{2\sigma^2}\right),$$

where

$$d(u, v)^2 = \sum_{j=1}^{4} \sum_{t=0}^{30} \left|\frac{u_j(t) - v_j(t)}{v_j(t)}\right|^2$$

is the relative distance between simulated and measured data,

$$\Phi(x) = \Phi(y_0, \theta) = \left(\text{proj}_4(y(t))\right)_{t=1,\ldots,30},$$

proj_4 denotes the projection onto the four measurable components, and $y(t)$ is the solution of the GynCycle model with initial values y_0 and parameters θ.

Remark 6. As mentioned above, the measurements for most women were not taken daily, resulting in incomplete data. In this case, Φ and d have to be chosen separately for each woman, restricting them to measured components. This does not influence the applicability of our algorithm.

We sampled the posterior using the adaptive mixture Metropolis algorithm by Roberts and Rosenthal [3], which basically is a multivariate Metropolis Hastings algorithm tuning its Gaussian proposal density for the current sample based on the covariance of the former ones. As the computation is independent for each person this problem is well scalable in the number of persons and we were thus able to compute 10 million samples for each person. The Raftery-Lewis diagnostic suggests around 7 million samples for convergence and the Gelman and Rubin criterion confirms this in our case with potential scale reduction factors smaller than 1.05.

As an example, we briefly present the results for one specific parameter, namely the blood volume. Figure 8 shows the sampled posterior $p_{\pi_0}^{z_m}$ in form of the marginal densities of the individual patients as well as the resulting first prior estimation π_1

in comparison to the flat uninformative prior π_0. Figure 9 illustrates the next 20 iterations of the prior estimation. Note that for illustrative purposes we scaled ρ in the likelihood for the iterations by a factor of five, decreasing the variation of the iterations and thus slowing down the peaking process, to obtain a clearer view of the proceedings. As it can be seen in the figure, the last prior estimate clearly favors a blood volume of about 5 liters, which is the blood volume of a typical adult.

Our next step will be to compute individual success rates for treatments frequently used in reproductive medicine and to compare the results with clinical outcomes.

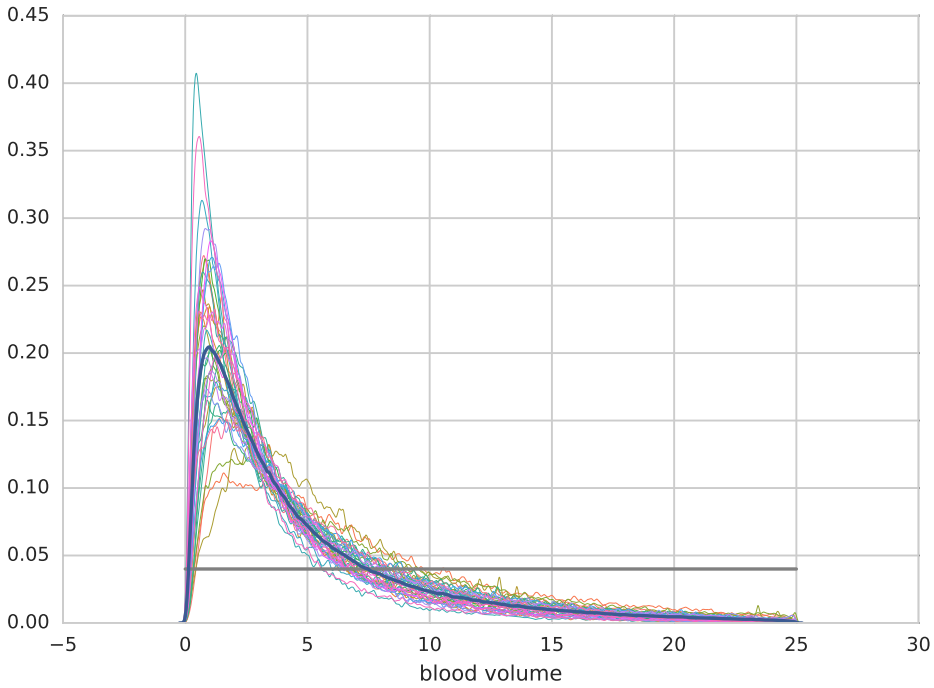


Figure 8: Marginal blood volume densities of the prior π_0 (gray), the individual posteriors $p_{\pi_0}^{z_m}$ (colored) for each patient, and the first prior estimation π_1 (blue).

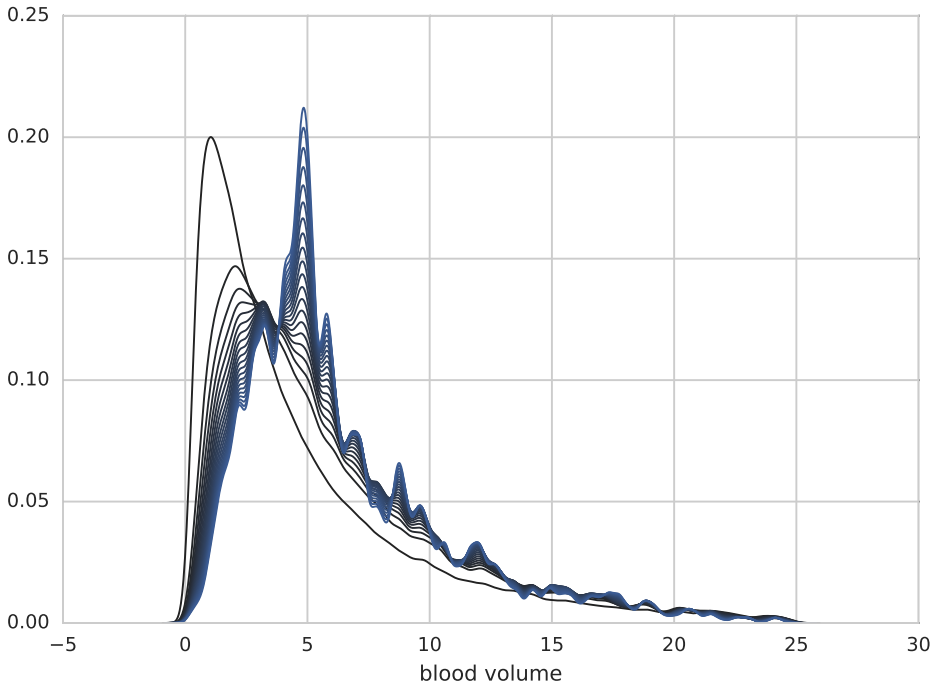


Figure 9: Marginal blood volume densities of the iterative prior estimations π_1 (black) to π_{21} (blue).

6 Conclusion

We have introduced a method that estimates the prior distribution in the setup of Bayesian statistics in the case when measurements for a large number of individuals are available. We have argued in which case a unique prior can, in principle, be deduced from measurements and have formulated a clear criterion for this case (the "identifiability assumption", see Section 2.1). We have also discussed what happens in the non-identifiable case.

A detailed scheme for the numerical realization of the method has been elaborated, see Sections 2.2 and 3. The numerical approximation to the fixed point iteration has to be applied with caution, since its convergence properties differ from the ones of the exact iteration, see Remark 5.

The method has been applied to a toy example in low dimensions to confirm our theoretical results, see Section 4. The work on a high-dimensional real life problem is in progress. The problem as well as a detailed line of action, how the method will be applied in this case, have been formulated (Section 5). First results are already available and have been presented.

As a byproduct, the method can be applied to deconvolve blurred images, as has been discussed in Appendix A.

A Deconvolution of blurred images

The probability densities π_n turn out to be blurred versions of ρ_X, whereby ρ_X is smoothed by a convolution kernel G_n:

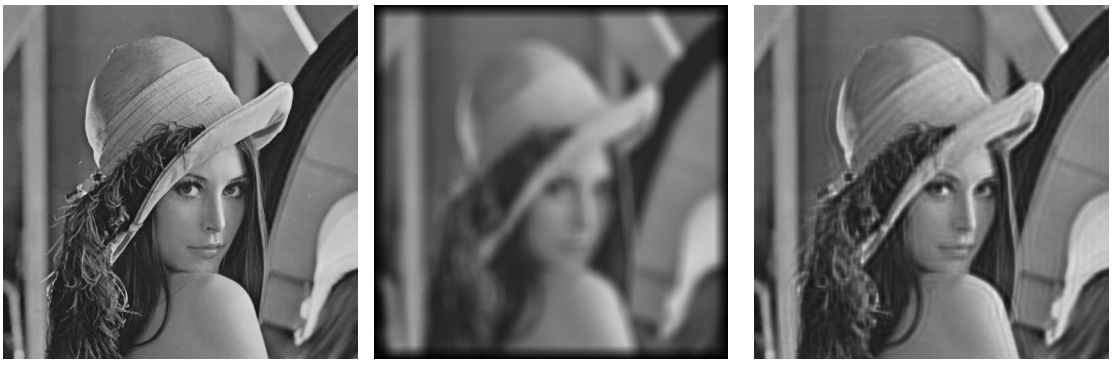
$$\pi_{n+1}(x) = \Psi\pi_n(x) = \int p^z_{\pi_n}(x)\,\rho_Z(z)\,\mathrm{d}z = \int \rho_X(\tilde{x}) \underbrace{\int p^z_{\pi_n}(x)\,\rho_Z(z\,|\,X=\tilde{x})\,\mathrm{d}z}_{=:G_{n+1}(x,\tilde{x})}\,\mathrm{d}\tilde{x}.$$

With growing number n the iterates appear less and less smoothed, converging to ρ_X (in the sense described in Section 2.1). Therefore, the fixed point iteration (2.1)–(2.2) results in a deconvolution process of π_0 to the original prior ρ_X.

This observation can readily be applied to the ("non-blind") deconvolution of blurred images. Setting $\phi = \mathrm{Id}$ and choosing the error density ρ_E as the point spread function, we can view $\rho_X : \mathbb{R}^2 \to \mathbb{R}_{\geq 0}$ as the original image (without loss of generality, we can assume that it is normalized and given by gray scale values) and the evidence

$$\rho_Z(z) = \int \rho_X(x)\,\rho_E(z-\phi(x))\,\mathrm{d}x = (\rho_X * \rho_E)(z)$$

as the blurred image. In this setup, our algorithm provides a method for the restoration of the original image from the blurred image, as demonstrated in Figure 10.



(a) original image (b) blurred image (c) deconvolved, 50 iter.

(d) deconvolved, 1500 iter. (e) deconvolved, 10000 iter. (f) deconvolved, 50000 iter.

Figure 10: Deconvolution of an artificially blurred image (b) using the fixed point iteration (2.1)–(2.2).

References

[1] Joseph L Doob. Application of the theory of martingales. *Le calcul des probabilites et ses applications*, pages 23–27, 1949.

[2] Lucien Le Cam and Grace Lo Yang. *Asymptotics in statistics: some basic concepts*. Springer Science & Business Media, 2012.

[3] Gareth O Roberts and Jeffrey S Rosenthal. Examples of adaptive MCMC. *Journal of Computational and Graphical Statistics*, 18(2):349–367, 2009.

[4] S. Röblitz, C. Stötzel, P. Deuflhard, H.M. Jones, D.-O. Azulay, P. van der Graaf, and S.W. Martin. A mathematical model of the human menstrual cycle for the administration of GnRH analogues. *J. Theoretical Biology*, 321:8–27, 2013.

[5] A. M. Stuart. Inverse problems: A Bayesian perspective. *Acta Numerica*, 19:451–559, 2010.