# Uncertainty Propagation in Biomedical Models

Andrea Franco, Marco Correia, and Jorge Cruz(✉)

NOVA Laboratory for Computer Science and Informatics, DI/FCT/UNL, Lisboa, Portugal
{andreafrfranco,marco.v.correia}@gmail.com jcrc@fct.unl.pt

**Abstract.** Mathematical models are prevalent in modern medicine. However, reasoning with realistic biomedical models is computationally demanding as parameters are typically subject to nonlinear relations, dynamic behavior, and uncertainty. This paper addresses this problem by proposing a new framework based on constraint programming for a sound propagation of uncertainty from model parameters to results. We apply our approach to an important problem in the obesity research field, the estimation of free-living energy intake in humans. Complementary to alternative solutions, our approach is able to correctly characterize the provided estimates given the uncertainty inherent to the model parameters.

## 1 Introduction

Uncertainty and nonlinearity play a major role in modeling most real-world continuous systems. In this work we use a probabilistic constraint approach that combines a stochastic representation of uncertainty on the parameter values with a reliable constraint framework robust to nonlinearity. The approach computes conditional probability distributions of the model parameters, given the uncertainty and the constraints.

The potential of our approach to support clinical practice is illustrated in a real world problem from the obesity research field. The impact of obesity on health is widely documented and the main cause for the “obesity pandemics” is the energy unbalance caused by an increased calorie intake associated to a lower energy expenditure.

Many biomedical models use the energy balance approach to simulate individual body weight dynamics. Change of body weight over time is modeled as the rate of energy stored (or lost), which is a function of the energy intake (from food) and the energy expended. The inability to rigorously assess the energy intake hinders the success and adherence to individual weight control interventions. The correct evaluation of such interventions will be highly dependent on the precision of energy intake estimates and the assessment of the uncertainty inherent to those estimates. We show how the probabilistic constraint framework can be used in clinical practice to characterize such uncertainty given the uncertainty of the underlying biomedical model.

## 2 Energy Intake Problem

The mathematical models that predict weight change in humans are usually based on the energy balance equation, $R = I - E$, where $R$ is the energy stored or lost (kcal/d), $I$ is the energy intake (kcal/d) and $E$ is the energy expended (kcal/d). Several models have

J.H. Holmes et al. (Eds.): AIME 2015, LNAI 9105, pp. 166–171, 2015.
DOI: 10.1007/978-3-319-19551-3_21

been applied to provide estimates of individual energy intake [10]. Our paper focus on the EI model [6] that computes the energy intake based on the differential equation:

$$cf\,\frac{dF}{dt} + cl\,\frac{dFF}{dt} = I - (DIT + PA + RMR + SPA) \qquad (1)$$

The left hand side of eq. (1) represents the change in body's energy stores $(R)$ and is modeled through the weighted sum of the changes in Fat mass $(F)$ and Fat Free mass $(FF)$. Differently from other models, that express the relationship between $F$ and $FF$ using a logarithmic model $FF^{log}\,(F)$ [9], or linear model [16], the EI model uses a 4th-order polynomial $FF^{poly}\,(F, a, h)$ to estimate $FF$ as a function of $F$, the age of the subject $a$, and its height $h$. The rate of energy expended $(E)$ is the total energy spent in several physiological processes: Diet Induced Thermogenesis $(DIT)$; Physical Activity $(PA)$; Resting Metabolic Rate $(RMR)$; Spontaneous Physical Activity $(SPA)$.

## 3 Constraint Programming

Continuous constraint programming [13,7] has been widely used to model safe reasoning in applications where uncertainty on parameter values is modeled by intervals including all their possibilities. A Continuous Constraint Satisfaction Problem (CCSP) is a triple $\langle X, D, C\rangle$ where $X$ is a tuple of $n$ real variables $\langle x_1, \cdots, x_n\rangle$, $D$ is a Cartesian product of intervals $D(x_1) \times \cdots \times D(x_n)$ (a box), where each $D(x_i)$ is the domain of $x_i$ and $C$ is a set of numerical constraints (equations or inequalities) on subsets of the variables in $X$. A solution of the CCSP is a value assignment to all variables satisfying all the constraints in $C$. The feasible space $F$ is the set of all CCSP solutions within $D$.

Continuous constraint reasoning relies on branch-and-prune algorithms [12] to obtain sets of boxes that cover the feasible space $F$. These algorithms begin with an initial crude cover of the feasible space $(D)$ which is recursively refined by interleaving pruning and branching steps until a stopping criterion is satisfied. The branching step splits a box from the covering into sub-boxes (usually two). The pruning step either eliminates a box from the covering or reduces it into a smaller (or equal) box maintaining all the exact solutions. Pruning is achieved through an algorithm that combines constraint propagation and consistency techniques based on interval analysis methods [14].

The direct application of classical constraint programming to biomedical models suffers from two major pitfalls: system dynamics modeled through differential equations cannot be represented and integrated within the constraint model; the interval representation of uncertainty is inadequate to distinguish between consistent scenarios.

**Differential Equations.** The behavior of many systems is naturally modeled by a system of Ordinary Differential Equations (ODEs). A parametric ODE system, with parameters $p$, represented as $y' = f(p, y, t)$, is a restriction on the sequence of values that $y$ can take over $t$. A solution within interval $T$, is any function that satisfies the equation for all values of $t \in T$. An Initial Value Problem (IVP) is characterized by an ODE system together with the initial condition $y(t_0) = y_0$ and its solution is the unique function that is a solution of the ODE system and satisfies the initial condition.

Several extensions to constraint programming [5,3,4] were proposed for handling differential equations based on interval methods for solving IVPs [14] which verify the existence of unique solutions and produce guaranteed error bounds for the solution trajectory along an interval $T$. They use interval arithmetic to compute safe enclosures for the trajectory, explicitly keeping the error term within safe bounds.

In this paper we use an approach similar to [5]. The idea is to consider an IVP as a function $\Phi$ where the first argument are the parameters $p$, the second argument is the initial condition to be verified at time point $t_0$ (third argument) and the last argument is a time point $t \in T$. A relation between the values at two time points $t_0$ and $t_1$ along the trajectory is represented by the equation $y(t_1) = \Phi(p, y(t_0), t_0, t_1)$. Using variables $x_0$ and $x_1$ to represent respectively $y(t_0)$ and $y(t_1)$, the equation is integrated into the CCSP as a constraint $x_1 = \Phi(p, x_0, t_0, t_1)$ with specialized constraint propagators to safely prune both variable domains based on a validated solver for IVPs [15].

**Probabilistic Constraint Programming.** An extension of the classical constraint programming paradigm is used to support probabilistic reasoning. Probabilistic constraint programming [2] associates a probabilistic space to the classical CCSP by defining an appropriate density function. A probabilistic constraint space is a pair $\langle\langle X, D, C\rangle, f\rangle$, where $\langle X, D, C\rangle$ is a CCSP and $f$ a p.d.f. defined in $\Omega \supseteq D$ such that: $\int_\Omega f(\mathbf{x})d\mathbf{x} = 1$. The constraints $C$ can be viewed as an event $\mathcal{H}$ whose probability can be computed by integrating $f$ over its feasible space. The probabilistic constraint framework relies on continuous constraint reasoning to get a box cover of the region of integration $\mathcal{H}$ and compute the overall integral by summing up the contributions of each box in the cover.

Monte Carlo methods [11] are used to estimate the integrals at each box. The success of this technique relies on the reduction of the sampling space where a pure Monte Carlo method is not only hard to tune but also impractical in small error settings.

## 4 Probabilistic Constraints for Solving the EI Problem

Let $t$ be the number of days since the beginning of treatment of a given subject, $F(t)$ the fat mass at time $t$, $w(t)$ the weight observed at time $t$, and $I$ the subject's energy intake, which is assumed to be a constant parameter between consecutive observations [6]. The energy balance equation and total body mass are related through the model:

$$F'(t) = g(I, F(t), t) \qquad\qquad w(t) = FF(a, h, F(t)) + F(t) \tag{2}$$

where $g$ is obtained by solving equation (1) with respect to $F'(t)$.

Let $i \in \{0, \ldots, n\}$ denote the $i$'th observation since beginning of treatment, occurred at time $t_i$, and let $F_i$ and $w_i$ be respectively the fat mass and the weight of the patient at time $t_i$ (with $t_0 = 0$). The EI model may be formalized as a CCSP $\left\langle X, \mathbb{IR}^{2n+1}, C\right\rangle$ with a set of variables $X = \{F_0\} \bigcup_{i=1}^{n} \{F_i, I_i\}$ representing the fat mass $F_i$ at each observation and the energy intake $I_i$ between consecutive observations (at $t_{i-1}$ and $t_i$), and a set of constraints $C = \{b_0\} \bigcup_{i=1}^{n} \{a_i, b_i\}$ enforcing eqs. (2):

$$a_i \equiv [F_i = \Phi(I_i, F_{i-1}, t_{i-1}, t_i)] \qquad b_i \equiv \left[w_i = FF^M(a, h, F_i) + \epsilon_i + F_i\right]$$

where uncertainty inherent to $FF$ estimation is integrated by considering that the true value of $FF$ is the model given $FF^M$ plus an error term $\epsilon_i \sim \mathcal{N}(\mu = 0, \sigma_\epsilon)$. Additionally, bounding constraints are considered for each observation: $3\sigma_\epsilon \leq \epsilon_i \leq 3\sigma_\epsilon$.

If we assume that the $FF$ model errors over the $n + 1$ distinct observations are independent, then each solution has an associated probability density value given by the joint p.d.f. $\prod_{i=0}^{n} f_i(\epsilon_i)$ where $f_i$ is the normal distribution associated with the error $\epsilon_i$. A more realistic alternative to errors independence, explicitly represents the deviation between error $\epsilon_i$ and the previous error $\epsilon_{i-1}$ as a normally distributed random variable $\delta_i \sim \mathcal{N}(\mu = 0, \sigma_\delta)$, resulting in the joint p.d.f. $\prod_{i=0}^{n} f_i(\epsilon_i) \prod_{i=1}^{n} h_i(\epsilon_i - \epsilon_{i-1})$ where $f_i$ and $h_i$ are the normal distributions associated with the errors $\epsilon_i$ and $\delta_i$ respectively.

**Method.** We developed an incremental method to efficiently solve the problem. It starts by computing the probability distribution of $F_0$ given the initial weight $w_0$ subject to the constraint $b_0$ and the bounding constraints for $\epsilon_0$. This distribution, $P^{\boxplus}(F_0)$, is discretized on a grid over $D(F_0)$ computed through probabilistic constraint programming. Given a sampled point $\dot{F}_0$, value $\dot{\epsilon}_0$ is determined by the constraint $b_0$, and its p.d.f. value is $f(\dot{F}_0) = f_0(\dot{\epsilon}_0)$. Similarly, the joint probability $P^{\boxplus}(F_1, I_1)$, is computed through probabilistic constraint programming by considering the constraints associated with observation 1, the observed weight $w_1$, and $P^{\boxplus}(F_0)$. Given a sampled point $(\dot{F}_1, \dot{I}_1)$, the values $\dot{F}_0$ and $\dot{\epsilon}_1$ are determined by constraints $a_1$ and $b_1$, and assuming errors independence, its p.d.f. is $f_0(\dot{\epsilon}_0) f_1(\dot{\epsilon}_1)$. However, we replace the computation of $f_0(\dot{\epsilon}_0)$ with the value of the probability $P^{\boxplus}(\dot{F}_0)$ computed in the previous step providing an approximation that converges to the correct value when the number of grid subdivisions goes to infinity: $f(\dot{F}_1, \dot{I}_1) \approx P^{\boxplus}(\dot{F}_0) f_1(\dot{\epsilon}_1)$. If the alternative p.d.f. is used, the approximation is: $f(\dot{F}_1, \dot{I}_1) \approx P^{\boxplus}(\dot{F}_0) f_1(\dot{\epsilon}_1) h_1(\epsilon_i - \epsilon_{i-1})$. Finally, the $P^{\boxplus}(F_1, I_1)$ is marginalized to obtain $P^{\boxplus}(F_1)$, and the process is iterated for the remaining observations.

## 5 Experimental Results

This section demonstrates how the approach may be applied to complement EI model predictions with measures of confidence. The algorithm was implemented in C++ and used for obtaining the probability distribution approximations $P^{\boxplus}(F_i, I_i)$ of a 45 years old woman over the course of the 24-week trial (CALERIE Study phase I [8]). The runtime was about 2 minutes per observation on an Intel Core i7 @ 2.4 GHz.

Fat Free mass is estimated using two distinct models: $FF^{poly}$[6], and $FF^{log}$[9]. Both models were initially fit to a set of 7278 North American women resulting in the standard deviation of the error, $\sigma_\epsilon^{poly} = 3.35$ and $\sigma_\epsilon^{log} = 5.04$. This data set was collected during NHANES surveys (1994 to 2004) and is available online [1]. We considered both assumptions regarding independence of the error. Due to space reasons we only show the results of the $FF^{poly}$ model assuming a correlated error with $\sigma_\delta = 0.5$.

**Joint Probability Distributions.** Figure 1(left) plots the results regarding the first observation showing the correlation between the uncertainty on $F$ and $I$. Experiments with the error independence assumption clarify its negative repercussion on the predicted distribution of $I$. Experiments with the $FF^{log}$ model revealed that the improved accuracy of $FF^{poly}$ model ($\sigma_\epsilon^{poly} < \sigma_\epsilon^{log}$) does not seem to impact the estimation of $I$.

**Marginal Probability Distributions with Confidence Intervals.** Figure 1(right) shows the estimated $I_i$ over time. Each box depicts the most probable value (marked in the center

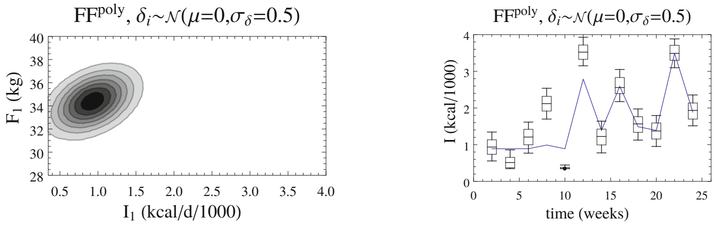

**Fig. 1.** Joint probabilities on the 1st observation (left). Confidence intervals for $I$ over time (right)

of the box), the union hull of the 50% most probable values (rectangle), and the union hull of the 82% most probable values (whiskers). Additionally, the plot overlays the estimates obtained from the algorithm published by the author of the EI model.

## 6 Conclusions

The standard practice for characterizing confidence on the predictions resulting from a complex model is to perform controlled experiments to assess its fitness statistically. However, controlled experiments are not always practical or have associated high costs. Contrary to the empirical, black-box approach, this paper proposes to characterize the uncertainty on the model estimates by propagating the errors stemming from each of its parts. The approach extends constraint programming to integrate probabilistic reasoning and dynamic behavior, offering a mathematically sound and efficient alternative.

The application field of the presented approach is quite broad: it targets models which are themselves composed of other (sub)models, for which there is a known characterization of the error. The selected EI model is a fairly complex model including dynamic behavior and nonlinear relations, and integrates various (sub)models with associated uncertainty. The experimental section illustrated how different choices for one of these (sub)models, the $FF$ model, impacts the error of the complete EI model, providing valuable information that can be integrated in a decision making support tool.

## References

1. National health and nutrition examination survey, http://www.cdc.gov/nchs/nhanes.htm
2. Carvalho, E.: Probabilistic Constraint Reasoning. PhD thesis, FCT/UNL (2012)
3. Cruz, J.: Constraint Reasoning for Differential Models. IOS Press (2005)
4. Cruz, J., Barahona, P.: Constraint reasoning in deep biomedical models. Artificial Intelligence in Medicine 34(1), 77–88 (2005)
5. Goldsztejn, A., Mullier, O., Eveillard, D., Hosobe, H.: Including ordinary differential equations based constraints in the standard CP framework. In: Cohen, D. (ed.) CP 2010. LNCS, vol. 6308, pp. 221–235. Springer, Heidelberg (2010)
6. Thomas, D., et al.: A computational model to determine energy intake during weight loss. Am. J. Clin. Nutr. 92(6), 1326–1331 (2010)
7. Benhamou, F., et al.: CLP(intervals) revisited. In: ISLP, pp. 124–138. MIT Press (1994)
8. Redman, L., et al.: Effect of calorie restriction with or without exercise on body composition and fat distribution. J. Clin. Endocrinol. Metab. 92(3), 865–872 (2007)

9. Forbes, G.: Lean body mass-body fat interrelationships in humans. Nut. R. 45, 225–231 (1987)
10. Hall, K.D., Chow, C.C.: Estimating changes in free–living energy intake and its confidence interval. Am. J. Clin. Nutr. 94, 66–74 (2011)
11. Hammersley, J., Handscomb, D.: Monte Carlo Methods. Methuen London (1964)
12. Van Hentenryck, P., Mcallester, D., Kapur, D.: Solving polynomial systems using a branch and prune approach. SIAM J. Num. Analysis 34, 797–827 (1997)
13. Lhomme, O.: Consistency techniques for numeric CSPs. In: IJCAI, pp. 232–238 (1993)
14. Moore, R.: Interval Analysis. Prentice-Hall, Englewood Cliffs (1966)
15. Nedialkov, N.: Vnode-lp a validated solver for initial value problems in ordinary differential equations. Technical report, McMaster Univ., Hamilton, Canada (2006)
16. Thomas, D., Ciesla, A., Levine, J., Stevens, J., Martin, C.: A mathematical model of weight change with adaptation. Math. Biosci. Eng. 6(4), 873–887 (2009)