

Parameter Uncertainty in CellML

Andrew Miller – ak.miller@auckland.ac.nz



**AUCKLAND
BIOENGINEERING INSTITUTE**

THE UNIVERSITY OF AUCKLAND
NEW ZEALAND

Te Whare Wānanga o Tāmaki Makaurau

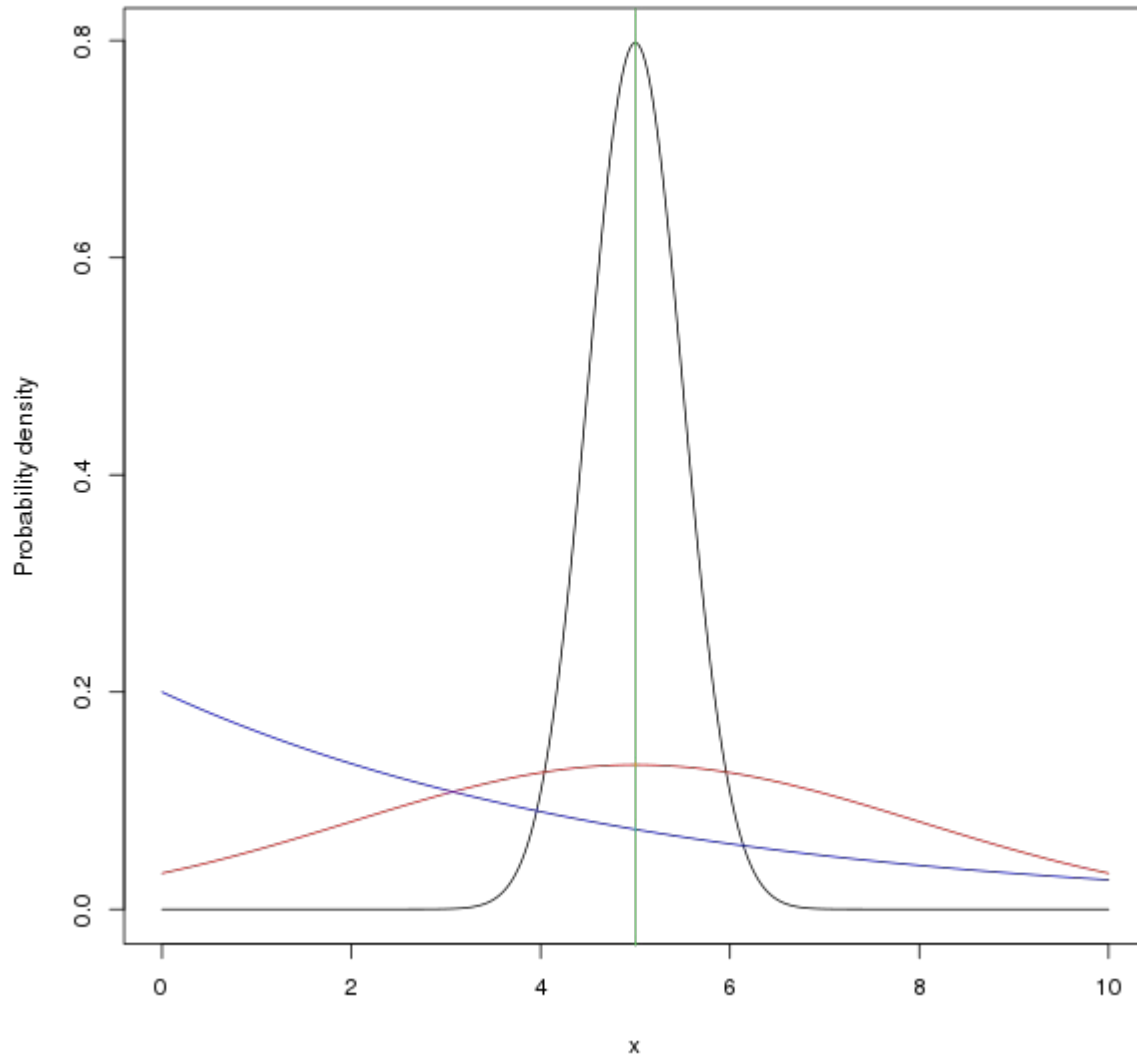
CellML models generally depend on a number of parameters and initial conditions.

The true values of these parameters in the individual being modelled are often unknown.

Uncertainty in inputs can lead to uncertainty in outputs – so it is therefore important to record and exchange information about uncertainty in inputs.

This is easiest to think about in the Bayesian framework – given my prior beliefs, updated based on the available evidence, what is the value of this parameter?

Posterior probability density for value of parameter x



In the ideal case, we have a closed form for the probability density function (p.d.f.), or the probability mass function (p.m.f.) of the posterior distribution.

Sometimes, there is no closed form, or we want to use data obtained from a source that doesn't give us the closed form (example: MCMC simulations, such as from WinBUGS, or experimental measurement). Therefore, we want to be able to represent a distribution based on a sufficiently large number of values sampled from that distribution (these are called realisations).

Parameters are not necessarily independent. Therefore, describing two different parameters that are not independent separately and composing the descriptions will not give accurate results.

CellML 1.0 and 1.1 only have scalar real values, so describing multivariate p.m.f.s and p.d.f.s is a challenge. However, it is often possible to split a multivariate distribution into marginal and conditional distributions. Consider \mathbf{X} , a vector of three components. We can't represent $P(\mathbf{X})$, so we instead represent $P(X_1)$, $P(X_2 | X_1)$, $P(X_3 | X_1 \cup X_2)$.

UncertML is an XML language specifically for describing uncertainty. It supports realisations, and also samples from a controlled vocabulary of distributions, with constant parameters to the distributions. It is being considered for use in an SBML proposal.

I haven't used UncertML for my work because it is quite different from how maths is represented in CellML (it doesn't use MathML) – which means you can't easily define conditional distributions or computed distribution parameters, and also because it doesn't support defining your own p.d.f.s or p.m.f.s.

However, I have produced a program that can convert from UncertML to the MathML used in CellML (even for the multivariate case) and can reverse it.

https://github.com/A1kmm/uncertml_to_physiome

Content MathML 2 does not include any predefined way to represent uncertainty.

We added support for uncertainty by creating new operators that are referenced using the `csymbol` element.

There are two types of operators:

- Several operators to construct expressions representing distributions.

- One operator to say link a parameter to the distribution, making the statement that the parameter is sampled from the distribution.

All these `csymbols` are prefixed by:

<http://www.cellml.org/uncertainty-1#>

This operator is used to make a statement that a parameter is sampled from a distribution.

Because CellML is declarative, it is not an instruction to sample, but a declaration of the relationship that holds between the variable and the distribution.

It is therefore used directly within the top level of the MathML, like an equality or inequality is, and not as a subexpression of another operator like equals.

It takes two arguments – a variable reference (ci element), and a description of a distribution, as defined on the following slides.

This operator is used to describe a value using realisations.

It takes a single parameter, which is either a vector of values, giving the different samples of a single variable, or a vector of vectors, giving a series of samples, each from a series of non-independent variables.

When a vector of vectors is used with this operator, the `uncertainParameterWithDistribution` operator will describe a vector of variables to assign to (of the same size the vector for each realisation), rather than a single variable.

This operator is used to describe a value using p.d.f.

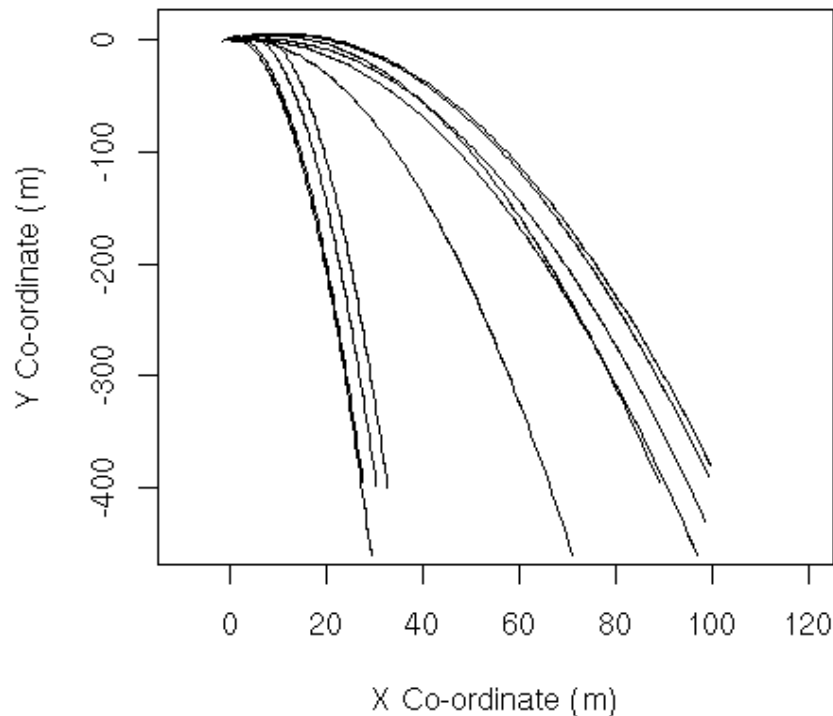
It takes a single parameter, which is a MathML lambda function, giving the probability density function. Because the p.d.f. is specified using MathML, the user is free to make it depend on other variables, including other uncertain parameters (to give a conditional distribution).

To describe multivariate distributions in this way, you have to use the marginal and conditional univariate distribution approach discussed – this is possible for nearly all commonly used multivariate distributions.

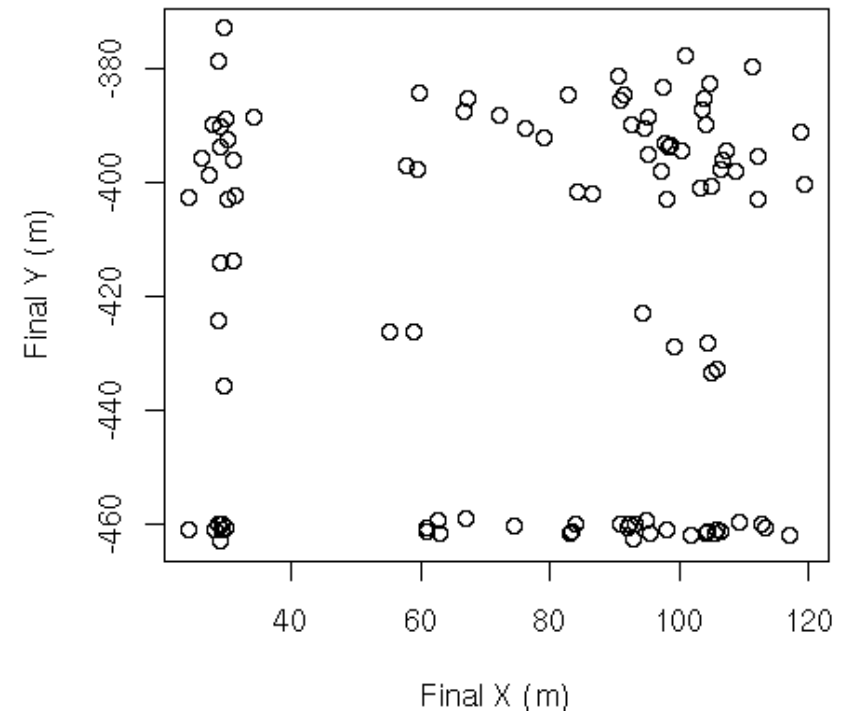
A very similar operator, `distributionFromMass`, is available to define uncertainty in discrete variables.

This approach to uncertainty has been implemented in the CellML API. The data in the graphs below were generated by the API, using a model of parabolic motion where the initial position and velocity are uncertain.

Sampled outcomes of parabolic motion model with uncertain initial position and velocity



Sensitivity analysis of final position in parabolic motion with uncertain initial position and velocity



Questions?

Comments / Discussion?

Suggestions?

Criticism?