Strategic questions in simulation

Timo Tiihonen

2016

Strategic questions in simulation

How should we operate when we want more than a single result from a single simulation model.

- How does the result depend from the input data?
- How does the result depend on the structural variants of the model?
- How does the uncertainty of parameters affect the reliability of results?
- Which variant or which combination of parameters gives the best results?

How can we answer such questions (apart from making ad hoc trials and inspecting summaries and graphs)?

Metamodelling

In metamodelling

- We fix a simple class of models to describe the phenomenon of interest.
- We fit an auxiliary model to the data obtained by simulation.
- We study if the auxiliary model can explain the observed results.
- We check that all variables in the auxiliary model are really needed.
- As a result we have a "small" model that explains the results.

Formal metamodelling

The real system can be considered as a function

$$y = f_0(V_1, V_2, \ldots),$$

where, f_0 is unknown and only part of the variables V_i are known.
The simulation model can be described as

$$y = f_1(z_1, \ldots, z_k, R_0),$$

where function f_1 is implicitly known (as the simulation code), z_i:s are the parameters of the code and R_0 stands for the seed value for random numbers.

Formal metamodelling

We often look for a metamodel in the form

$$y = \sum_{i=0}^{q-1} \beta_i x_i + e,$$

where x_i:s are known functions of z_j:s (powers, logarithms, etc). (If needed also y can be a transformation of the "real" output).
Such a model is called as regression model, x_i:s are regression variables and β_i:s are regression coefficients. e is a stochastic error term (with zero expectation).

Regression analysis

In regression analysis we make n ($n \geq q$) experiments with different combinations of z_k values. Then also the values of x_j:s vary.
Let us denote $X_{ij} = x_j$ for test i, $i = 1, \ldots, n$, $j = 0, \ldots, q-1$.
($X_{i0} = 1$, $\forall i$). $X = \{X_{ij}\}_{ij}$. The results of the tests are denoted by the vector $Y = \{y_i\}_i$, $i = 1, \ldots, n$.
The goal is to define values for the coefficients $\beta = \{\beta_l\}_l$,
$l = 0, \ldots, q-1$.
The equation $Y = X\beta$ is not solvable in general (if $n > q$). We look for so called LS (least squares) solution β, that minimizes the error $\|Y - X\beta\|^2$,

$$\beta = (X^t X)^{-1} X^t Y.$$

As Y is a random variate, also β is a random variate.

Regression analysis

What kind of model is good and how is it determined:

- Are all x_i:s needed (or is some $\beta_i \approx 0$)?
- Do the x_i:s explain the results (is $Y - X\beta$ small)?
- How reliable are the values of β:s?
- How β:s can be determined efficiently and reliably?

We analyse the situation when the model is accurate. That is, for given X,

$$y_i = (X\beta)_i + e_i,$$

for some β, with e_i:s being independent $N(0, \sigma^2)$ variables. That is, the tests are independent and $Var(y) = \sigma^2$ in all tested configurations.

Regression analysis

Then $\hat{\beta} = (X^t X)^{-1} X^t Y$ is an unbiased estimate for β. The components of β are mutually correlated,

$$\begin{aligned} Cov(\beta) &= Cov(\beta_i, \beta_j)_{ij} \\ &= (X^t X)^{-1} X^t Cov(Y)[(X^t X)^{-1} X^t]^t \\ &= (X^t X)^{-1} \sigma^2 \end{aligned}$$

as $Cov(Y) = \sigma^2 I$.
The variance σ^2 at a single test point can be estimated either by making several replications for each test of by making $n >> q$ tests, and estimating

$$\hat{\sigma}^2 = \sum_{i=1}^{n} \frac{(y_i - (X\hat{\beta})_i)^2}{n - q}.$$

Regression analysis

How to recognize unneccessary variables x?

- First form all possible regression variables (z_j:s, z_j^2:s, $z_i z_j$:s, etc).
- Make n tests ($n > 2q$).
- Discard those x_i:s, for which $\beta_i \approx 0$.
- Compute new values for β:s (using the previous y values but with smaller X matrix).

Assumptions

Is it realistic to assume that $y = X\beta + e$, with e:s independent and $N(0, \sigma^2)$?

- If the test points are properly chosen and rich enough set of regression variables x is chosen, the basic assumption $y = X\beta$ is OK.
- Independence can be guaranteed by choosing independent random number streams (if this is desired).
- Normality of observations is usually valid if simulations are long enough.
- It is not realistic to assume the variance to be constant accross test points.

To estimate the varying variance each test point has to be replicated several times.

Dependent tests

If the tests are dependent (same random numbers), also the covariance has to be estimated. If we make m replications for each test, we get the estimate

$$\hat{\sigma}_{ij} = \frac{1}{m(m-1)} \sum_{l=1}^{m} (y_{il} - \bar{y}_i)(y_{jl} - \bar{y}_j).$$

The regression coefficients can be determined from the generalized LS-problem

$$\min_{\beta} (y - X\beta)^t (Cov(y))^{-1} (y - X\beta).$$

The covariance matrix for β is $Cov(\beta) = (X^t (Cov(y))^{-1} X)^{-1}$

Dependent tests

What is the effect of using common random streams? We consider the most simple regression model $y = \beta_0 + \beta_1 x$. We can assume (for notational simplicity) that $\bar{x} = 0$ and $\bar{y} = 0$.
We assume constant variance (σ^2) for each test point. Then the least square values for coefficients are

$$\hat{\beta}_1 = \frac{\sum_i (x_i - \bar{x})(y_i - \bar{y})}{\sum_i (x_i - \bar{x})^2} = \frac{\sum_i x_i y_i}{\sum_i x_i^2}$$

and $\hat{\beta}_0 = \bar{y} - \hat{\beta}_1 \bar{x}$.

Dependent tests

Denote $a_i = \frac{x_i}{\sum_j x_j^2}$ to get $\hat{\beta}_1 = \sum_i a_i y_i$.
If tests are independent, $Var(\hat{\beta}_1) = \sum_i a_i^2 Var(y_i) = \sigma^2 \sum_i a_i^2$.
If tests are dependent, $Var(A^t Y) = ACov(Y)A^t$,

$$Var(\hat{\beta}_1) = \sigma^2 \sum_i a_i^2 + \sum_{i \neq j} a_i a_j Cov(y_i, y_j)$$

Now $\sum_{i \neq j} a_i a_j < 0$. If $Cov(y_i, y_j)$ is non-negative $\forall i, j$, the variance for $\hat{\beta}_1$ is smaller than in the case of independent tests.
Positive correlation may be obtained a.o. by using common random numbers.
Correspondingly, $Var(\hat{\beta}_0)$ will be bigger if tests are positively correlated.

Model validation

Construction and validation of a regression model proceed as follows:

- Define the general form of the model ($y = X\beta + e$),
- Determine the model coefficients ($\hat{\beta} = (X^t X)^{-1} X^t y$).
- Test the prediction power of the model by making a new test at point x_{n+1} and comparing the result y_{n+1} to the prediction $x_{n+1}^t \hat{\beta}$.

The test variable

$$z_{n+1} = \frac{y_{n+1} - x_{n+1}^t \hat{\beta}}{\sqrt{Var(y_{n+1}) + Var(\hat{\beta} x_{n+1})}}$$

should obey $N(0, 1)$ distribution. If z_{n+1} has an exceptional value, we reject the regression model.
Typically extreme values for input variables are used to determine the β:s. Then values in the middle of domain are used for testing. This reveals possible quadratic dependencies.

Model validation

Another common method is cross validation. Each test in its turn is dropped from the calibration set and used as a reference for testing. This gives n different validation tests. If sufficiently many tests pass, the model is accepted.
When model is validated, we can test which joint effects are significant (corresponding β significantly different from zero). By removing the rest of the joint effects we get a smaller model that explains the observations.
If model is not valid we can either augment it (by higher order terms and joint effects) or change it by modifying the regression variables.

Model validation

Example: we want to model the queuing time w as a function of service time s and interarrival time a. The model $w = \beta_0 + \beta_1 s + \beta_2 a$ is not natural as in reality the expectation of waiting time grows without limit as s approaches a.
More appropriate could be for example $w = (\beta_0) + s\beta_1 + \beta_2 \frac{as}{a-s}$.
Here w grows to infinity as s approaches to a.

Experiment design

Experiment design aims to efficient and reliable construction of regression models to explain real or simulated systems.
Historically it has been developed to control long and demanding experiment series that depend on external conditions (a.o. in agriculture).
In the context of simulation, experimental design is easy in principle.

- All parameters can be fully controlled
- Each test can be easily repeated/replicated
- Time needed for a single experiment is typically short.

Experiment design

The goal is to fit the model $y = \sum_{i=1}^{q} \beta_i x_i$ using n ($n \geq q$) experiments.
Minimal requirement is that the $q \times q$ matrix $X^t X$ is invertible.
Otherwise the experiment points can be chosen freely.
If regression variables are functions of actual simulation parameters (powers, products, etc), all x values can not be varied independently from the others. By **factor** we refer to variable that can be independently varied during the series of experiments. The possible values of factors (within the experiment series) is called as **level**.
In case of a single factor the analysis of linear dependency requires two levels, quadratic dependency three levels etc.

Experiment design

If we have k factors, i:th factor L_i levels, the full experiment would require at least $L_1 L_2 \ldots L_k$ test points (each replicated to estimate variances) which is too much if there are many factors.
With experiment design we aim to

- Determine chosen effects in a reliable way.
- Minimize the number of experiments by ignoring certain joint effects from consideration.
- Ensure that joint effects do not disturb the evaluation of main effects.
- Determine the effects with high confidence (small covariance for the coefficient matrix).

Experiment design

Two extreme approaches are: one factor at time (total of $1 + \sum_i (L_i - 1)$ tests), which gives the main effects but no control to the joint effects - and the full design (all factor-level combinations which gives all needed information but is very expensive).
Experiment design operates between these extremes.

Example

Consider a simple case where all factors have only two levels. So we only consider linear effects and their interactions.
It is custom to denote the levels by "+" and "-" ($+1$ and -1 although the actual values may be real, integer, logical, ..).
Consider a model with three factors, $y = \beta_0 + \sum_{i=1}^{3} \beta_i x_i$.

Example

One factor at time design could be for example

Test	x_1	x_2	x_3
1	$-$	$-$	$-$
2	$+$	$-$	$-$
3	$-$	$+$	$-$
4	$-$	$-$	$+$

The effect of factor j can be estimated by computing
$\hat{\beta}_j = (y_{j+1} - y_1)/2$ (assuming the levels $+1$ and -1).
If linear model is valid (no joint effects), the estimate is unbiased

$$E(\hat{\beta}_j) = \frac{1}{2}(E(y_{j+1}) - E(y_1)) = \beta_j.$$

Example

If tests are independent and variance of y_i is σ^2, $Var(\hat{\beta}_j) = \sigma^2/2$.
In case of joint effects, i.e. real model being
$y = \beta_0 + \sum_i \beta_i x_i + \sum_{i \neq j} \beta_{ij} x_i x_j + \ldots$, the estimate is biased:

Test	x_1	x_2	x_3	x_1x_2	x_2x_3	x_3x_1	$x_1x_2x_3$
1	$-$	$-$	$-$	$+$	$+$	$+$	$-$
2	$+$	$-$	$-$	$-$	$+$	$-$	$+$
3	$-$	$+$	$-$	$-$	$-$	$+$	$+$
4	$-$	$-$	$+$	$+$	$-$	$-$	$+$

$$E(\hat{\beta}_1) = \beta_1 - \beta_{12} - \beta_{13} + \beta_{123}$$

Example

If we do all factor-level combinations

Test	x_1	x_2	x_3	x_1x_2	x_2x_3	x_3x_1	$x_1x_2x_3$
1	$-$	$-$	$-$	$+$	$+$	$+$	$-$
2	$+$	$-$	$-$	$-$	$+$	$-$	$+$
3	$-$	$+$	$-$	$-$	$-$	$+$	$+$
4	$+$	$+$	$-$	$+$	$-$	$-$	$-$
5	$-$	$-$	$+$	$+$	$-$	$-$	$+$
6	$+$	$-$	$+$	$-$	$-$	$+$	$-$
7	$-$	$+$	$+$	$-$	$+$	$-$	$-$
8	$+$	$+$	$+$	$+$	$+$	$+$	$+$

we do double work compared to factor at time approach.

Example

To make full use of collected information we can estimate

$$\hat{\beta}_j = \sum_i x_{ij} y_j / n$$

Because matrix X has the property

$$\sum_i x_{ij} = 0 \ \forall j, \qquad \sum_i x_{ij} x_{il} = 0, \ j \neq l$$

we get

$$\begin{aligned} E(\hat{\beta}_j) &= \frac{1}{n} \sum_i x_{ij} E(y_i) \\ &= \frac{1}{n} \sum_i x_{ij} (\beta_0 + \sum_l \beta_l x_{il}) = \beta_j \end{aligned}$$

Example

Correspondingly the variance is

$$Var(\hat{\beta}_j) = \frac{1}{n^2} \sum_i x_{ij}^2 Var(y_i) = \sigma^2/n$$

In the full test the variance was reduced by factor four by doubling the work. Thus the full test is more efficient than one factor at time. More over the joint effects can be resolved and they are not causing bias to the main effects.

Example

Can the same efficiency be obtained with smaller experiment?
We consider a so called partial (2^{3-1}) experiment.
We select from the full experiment those tests for which $x_1x_2x_3 = +1$. That is, we make the tests

Test	x_1	x_2	x_3	x_1x_2	x_2x_3	x_3x_1	$x_1x_2x_3$
2	+	−	−	−	+	−	+
3	−	+	−	−	−	+	+
5	−	−	+	+	−	−	+
8	+	+	+	+	+	+	+

Example

For the main effects we get the estimates as for the full experiment

$$\hat{\beta}_j = \sum_i x_{ij} y_j / n$$

The variance $Var(\hat{\beta}) = \sigma^2/4$ is smaller than for the factor at time experiment although the work is the same.
If joint effects are important, they cause bias to the main effects, for example.

$$E(\hat{\beta}_1) = \beta_1 + \beta_{23}.$$

2^{k-p}- designs

Above we chose four $(q+1)$ experiments from eight (2^q) possible. In general case (for bigger q) this can be done in may ways. How can we control the bias remanining in the results.
In the example we chose experiments with $x_1x_2x_3 = 1$. The the joint effect of three variables can not be distinguished from the constant factor x_0 and it causes a bias to β_0.
Likewise x_1 and the product x_2x_3 are aliased (as any main effect with the joint effect of the other two).

2^{k-p}- designs

If there are more factors, by chosing those tests where the product of all factors is constant we half the amount of experiments and alias the main effects to the joint effect of all the remaining factors (and not to for example to joint effect of two factors).
Half of full test is generally too much. So we have to reduce the experiments even more.
A common aproach is to construct so called 2^{k-p} experiments. This is done by selecting p joint effects that are aliased to the constant factor x_0. (Above $p = 1$ and the joint effect of all factors was aliased). The aliased joint effects are called the generators of the experiment design.

2^{k-p}- designs

For example if $q = 5$, we need at minimum $n = q + 1 = 6 < 2^{q-2}$ experiments. So we can contruct a design that resolves the main effects and aliases two joint effects to the constant. These can be chosen in several ways. Chosing, for example $1 = x_1x_2x_3 = x_1x_4x_5$, it follows that we alias also $x_1 = x_2x_3 = x_4x_5$, $x_2 = x_1x_3$, etc.
If we have reasons to believe that certain joint effects are small and/or we want certain main effects un biased, we can select the generators taking these aspects in mind.

2^{k-p}- designs

If we want to avoid two factor joint effects from aliasing with main effects we need more than $q+1$ experiments. In case of $q=5$ we can create 2^{5-1} design by aliasing $x_1x_2x_3x_4x_5$ to a constant. Then the main effects are aliased to joint effects of four other factors and two factor joint effects to three factor joint effects. Hence both main factor and their simple joint effects can be obtained relatively unbiased. The price to pay is 16 experiments for just 6 coefficients. In a general case 2^{k-p} designs are not always the most efficient ones.

2^{k-p}- designs

General designs are classified according to their resolution as follows:

- Resolution III, if no main effects are aliased
- Resolution IV, if main effects are not aliased to two factor joint effects
- Resolution V, if no two factor joint effects are aliased to each other

The names of the classes come from the lengths of the generators of 2^{k-p} designs having the above properties.

2^{k-p}- designs

R-III design has at least $q + 1$ experiments, (like factor at time). Optimal variance can be reached if the number of experiment is a multiple of four ($n = 4m$). If $4m = 2^s$, we get the design with 2^{k-p} technique. For other values there is no general algorithm but design tables can be found in the litterature.
R-IV desing can be obtained from a R-III design by repeating the experiments with all factor values changed to the other level (joint effects do not change sign but main effects do).
R-V design requires minimum of $1 + q + (q - 1)q/2$ experiments.