

Creating a simple **emulator** case study from scratch: a cookbook

Robin K. S. Hankin

March 7, 2007

1 Introduction

Package **emulator** of bundle **BACCO** performs Bayesian emulation of computer models. This document constructs a minimal working example of a simple problem, step by step. Datasets and functions have a `.vig` suffix, representing “vignette”.

This document is not a substitute for KOH or KOHa or Hankin 2005 or the online help files in **BACCO**. It is not intended to stand alone: for example, the notation used here is that of KOH, and the user is expected to consult the online help in the **BACCO** package when appropriate.

This document is primarily didactic, although it is informal.

Nevertheless, many of the points raised here are duplicated in the **BACCO** helpfiles.

The author would be delighted to know of any improvements or suggestions. Email me at `r.hankin@noc.soton.ac.uk`.

2 List of objects that the user needs to supply

The user needs to supply three objects:

- A design matrix, here `val.vig` (rows of this show where the code has been evaluated)
- Basis functions. Here `basis.vig`. This shows the basis functions used for fitting the prior
- Data, here `z.vig`. This shows the data obtained from evaluating the various levels of code at the points given by the design matrix and the subsets object.

Each of these is discussed in a separate subsection below.

But the first thing we need to do is install the library:

```
> library(emulator)
```

2.1 Design matrix: USER TO SUPPLY

In these sections I show the objects that the user needs to supply, under a heading like the one above. In the case of the **emulator** we need a design matrix and a vector of outputs.

The first thing needed is the design matrix `val.vig`, ie the points in parameter space at which the lowest-level code is executed. The example here has just two parameters, `a` and `b`:

```
> head(val.vig)
```

```
      [,1]      [,2]
[1,] 0.15000000 0.3833333
[2,] 0.05000000 0.4500000
[3,] 0.85000000 0.7833333
[4,] 0.91666667 0.4833333
```

```
[5,] 0.08333333 0.9833333
[6,] 0.38333333 0.2833333
```

```
> nrow(val.vig)
```

```
[1] 30
```

Notes

- Each row is a point in parameter space, here two dimensional.
- The parameters are labelled `a` and `b`

2.2 Basis functions: USER TO SUPPLY

Now we need to choose a basis function. Do this by copying `basis.toy()` but fiddling with it:

```
> basis.vig <- function(x) {
+   out <- c(1, x, x[1] * x[2])
+   names(out) <- c("const", LETTERS[1:2], "interaction")
+   return(out)
+ }
```

Notes

- This is shamelessly ripped off from `basis.toy()`, except that I've changed the basis to be `c(1,a,b,ab)`.

2.3 Data: USER TO SUPPLY

The data we have for the `.vig` example is a vector whose elements are the output of the code at the points specified in `val.vig`:

```
> head(z.vig)
```

```
[1] 4.384898 4.075449 8.099786 6.499868 4.771578 4.888417
```

```
> summary(z.vig)
```

Min.	1st Qu.	Median	Mean	3rd Qu.	Max.
2.446	4.432	5.460	5.468	6.495	8.100

3 Data analysis

The previous section showed what data and functions the user needs to supply. These all have a `.vig` suffix. This section shows the data being analyzed.

First we will estimate the scales to use:

```
> os <- optimal.scales(val = val.vig, scales.start = c(10, 10),
+   d = z.vig, func = basis.vig)
> os
```

```
[1] 2.519524 6.193261
```

So we can estimate the coefficients. But first we have to calculate the variance matrix and invert it:

```
> A.os <- corr.matrix(xold = val.vig, scales = REAL.SCALES)
> Ainv.os <- solve(A)
```

Given this, use `betahat.fun()` to get the coeffs:

```
> betahat.fun(xold = val.vig, d = z.vig, Ainv = solve(A), func = basis.vig)
```

const	A	B interaction	
1.519356	2.396322	3.015938	3.985749

The central function is interpolant:

```
> interpolant(x = c(0.5, 0.5), d = z.vig, Ainv = Ainv.os, scales = os,
+   xold = val.vig, func = basis.vig, give.full.list = TRUE)
```

```
$betahat
      const      A      B interaction
1.519356 2.396322 3.015938 3.985749
```

```
$prior
      [,1]
[1,] 5.221923
```

```
$beta.var
      const      A      B interaction
const  0.2815960 -0.2488106 -0.2906636  0.2786592
A      -0.2488106  0.6254137  0.2720926 -0.7198991
B      -0.2906636  0.2720926  0.6118967 -0.5628236
interaction 0.2786592 -0.7198991 -0.5628236  1.3742204
```

```
$beta.marginal.sd
      const      A      B interaction
0.5306562 0.7908310 0.7822383 1.1722715
```

```
$sigmahat.square
[1] 0.4433475
```

```
$mstar.star
      [,1]
[1,] 5.517794
```

```
$cstar
[1] -0.01356517
```

```
$cstar.star
[1] -0.01241458
```

```
$Z
[1] 0.07418877
```

And that's it, really.

A Data generation

The data used in this study were created by directly sampling from the appropriate multivariate Gaussian:

```
> REAL.BETA <- 1:4
> REAL.SCALES <- c(3, 6)
> REAL.SIGMASQUARED <- 0.3
> A <- corr.matrix(xold = val.vig, scales = REAL.SCALES)
> z.vig <- as.vector(rmvnorm(n = 1, mean = crossprod(REAL.BETA,
+   apply(val.vig, 1, basis.vig)), sigma = A * REAL.SIGMASQUARED))
```

Robin K. S. Hankin
National Oceanography Centre, Southampton
European Way
Southampton SO14 3ZH
United Kingdom
E-mail: r.hankin@noc.soton.ac.uk URL: <http://www.noc.soton.ac.uk>