LINEARIZATION OF LOCAL PROBABILISTIC SENSITIVITY VIA SAMPLE RE-WEIGHTING

R.M.Cooke[(1)], *D. Kurowicka*[(1)], *I. Meilijson* [(2)]

[(1)] Dep. of Math.,Delft University of Technology, Delft, The Netherlands
Email: R.M.Cooke@its.tudelft.nl , D.Kurowicka@its.tudelft.nl

[(2)] School of Mathematical Sciences, Tel Aviv University, Tel Aviv, Israel,
Email: isaco@math.tau.ac.il

ABSTRACT

Local probabilistic sensitivity of input variable *X* with respect to output variable *Z* is proportional to the derivative of the conditional expectation *E*(*X*|*z*). This paper reports on experience in computing this conditional expectation. Linearized estimates are found to give acceptable performance, but are not generally applicable. A new method of linearization based on re-weighting a Monte Carlo sample is introduced. Results are comparable to the linearized estimates, but this method is more widely applicable. Results generally improve by conditioning on a small window around *z*.

1. INTRODUCTION

Local probabilistic sensitivity measures (LPSM) were introduced in [3] to describe the importance of an input variable *X* to a given contour of an output variable *Z*:

$$LPSM(X) = \frac{\sigma_Z}{\sigma_X} \frac{\partial E(X|Z=z)}{\partial z}\Big|_{z=z_0} = \frac{\sigma_Z \partial E(X|z_0)}{\sigma_X \partial z_0} \qquad (1)$$

This measure is indicated when we are particularly interested certain values of the output variable. Thus when *Z* represents the 'strength - load' of a structure, we are particularly interested in the value *Z = 0* corresponding to failure of the structure. It was shown that if the regression *E*(*X*|*Z*) is linear, then *LPSM(X) = ρ(Z,X)* (see section (5)). In special cases, including the independent normal linear model, *LPSM*(*X*) can be computed analytically [3].

Problems in computing *LPSM* have motivated further study of its properties. It can be shown that in the case of independent linear normal models, the *LPSM* and the standard global measures are dual in a straightforward sense. The generalization of the standard global measure to non-linear models makes use of the correlation ratio. A similar generalization is conjectured for the *LPSM*. The duality relation suggests alternative ways of calculating the *LPSM* which appear to give acceptable performance.

Section (2) illustrates problems that can arise in computing the LPSM in Monte Carlo simulation. Section (3) reviews sensitivity measures in the linear model. Section (4) explores properties of the correlation ratio. Section (5) establishes the duality relationship for the independent linear normal case. This relationship suggests new ways of calculating the derivative of the conditional expectation. Analytical methods and linear approximations are discussed. Though not always applicable in practice, these

nonetheless provide a benchmark for the method introduced in section (6). This method `linearizes by re-weighting' a Monte Carlo sample, and can always be applied in Monte Carlo simulation. In section (7) the performance of these methods is compared. A final section gathers conclusions.

2. AN EXAMPLE

The obvious way to approximate *LPSM(X)* in Monte Carlo simulations is to compute

$$\frac{E(X\,|\,Z\in(z_0,z_0+\varepsilon))-E(X\,|\,Z\in(z_0-\varepsilon,z_0))}{E(Z\,|\,Z\in(z_0,z_0+\varepsilon))-E(Z\,|\,Z\in(z_0-\varepsilon,z_0))} \tag{2}$$

In some cases this is very unstable. Consider the following example, which was proposed by Ton Vrouwenvelder, where X and Y are independent standard normal:

$$Z = min(3-X,3-Y)$$

One can calculate that (see appendix):

$$\frac{\partial E(X|Z=z)}{\partial z}\Big|_{z=0} = -0.507\,.$$

On a Monte Carlo simulation with 5,000,000 samples and ε = 0.1 the above method yields the estimates

$$\frac{\partial}{\partial z}E(X|z=0)_{simulation} = -0.517,$$
$$\frac{\partial}{\partial z}E(Y|z=0)_{simulation} = -0.807.$$

Of course, by symmetry these two derivatives must be equal. The number of samples used is unrealistically large, and still performance is poor. This is explained by a number of factors. First if high accuracy is desired, ε must be chosen small in (2). On the other hand the difference in conditional expectations must be large enough to be statistically significant. In the above example this difference was barely significant at the 5% level for *Y* and was not significant for *X*. In this case, the difference in conditional expectations in (2) is small, because, roughly speaking, *X* feels the effect of conditionalizing on *Z = 0* on only one half of the samples. Finally, conditionalizing on extreme values of *Z*, as in this case, can introduce strong correlations between the input variables. In this case the conditional correlations are negative. This means that sampling fluctuations in the estimates of the conditional expectations in (2) will be correlated. Indeed, it required an unrealistically large number simply to obtain estimates whose signs were both negative (see also the results in Table 4).

It is clear that alternative methods of calculating the LPSM are needed.

3. THE LINEAR MODEL

Let $Z = Z(\mathrm{X})$ be a function of vector $\mathrm{X}=(X_1,\ldots,X_n)$. Assuming that Z is analytic, it can be expanded in the neighbourhood of some point $x^* = (x_1^*,...,x_n^*)$ and neglecting higher order terms (HOT's):

$$Z(X) \sim Z(x^*) + \sum_{i=1}^{n} \partial_i Z(x^*)(X_i - x_i^*) \tag{3}$$

where ∂_i denotes $\partial / \partial_{X_i}$.

Let μ_i and σ_i denote mean and standard deviation of X_i respectively. We obtain

$$E(Z) \sim Z(x^*) + \sum_{i=1}^{n} (\mu_i - x_i^*)\partial_i Z(x^*),$$

$$Var(Z) \sim + \sum_{i,j=1}^{n} Cov(X_i, X_j)\partial_i Z(x^*)\partial_j Z(x^*).$$

If X_i are all uncorrelated then

$$Cov(Z, X_i) = \sigma_i^2 \partial_i Z = \rho(Z, X_i)\sigma_Z \sigma_{X_i}.$$

Hence, in the linear uncorrelated model, the rate of change of Z with respect to X_i may be expressed as

$$\partial_i Z = Cov(Z, X_i) / \sigma_i^2. \tag{4}$$

We note that the left hand side depends on the point x^* whereas the right hand side does not. This of course reflects the assumption of non-correlation and the neglect of HOT's. A familiar sensitivity measure involves a "sum square normalization":

$$\alpha_i = \rho(Z, X_i) = \frac{\partial_i Z(x^*)\sigma_i}{\sigma_Z}.$$

The factor α_i gives the influence of variable X_i on the standard deviation of Z. It depends on the slope of the tangent line of Z in the point z^*. For the linear model and when X_i's are uncorrelated,

$$R^2 = \sum_{i=1}^{n} \alpha_i^2 = 1. \tag{5}$$

This can be considered as a measure of the variance of Z explained by the linear model. If R^2 is less then one, this may be caused either by dependencies among X_i's or by the contribution of higher order terms neglected in (3).

When employing the Taylor expansion as above, it is common to introduce a transformation of the variable Z which enables us to capture as much of the behaviour of the transformed variable as possible in low order terms. Alternatively, one could transform the variables X_i toward the same end. These considerations lead to the correlation ratio, whose properties we study in the next section.

4. CORRELATION RATIO

The correlation ratio is one of the most important non-directional measures of uncertainty contribution [2].

Definition 4.1 *(Correlation ratio) Let G be a random variable, and X a random vector. The quantity* $\frac{\sigma^2_{E(G|X)}}{\sigma^2_G}$ *is called the correlation ratio of X to G and denoted CR(X,G).*

We consider a function $G = G(X,Y)$ of random vectors X and Y with $\sigma^2_G < \infty$. In analogy with non-linear regression methods, we may ask for which function $f(X)$ with $\sigma^2_{f(X)} < \infty$ is $\rho^2(G, f(X))$ maximal? The answer given in the following

Proposition 4.2 Let G = G(X,Y) with $\sigma^2_G < \infty$ then

(i) $Cov(G, E(G|X)) = \sigma^2_{E(G|X)}$

(ii) $max_{f;\sigma^2_{f(x)}<\infty} \rho^2(G, f(X)) = \rho^2(G, E(G|X)) = \frac{\sigma^2_{E(G|X)}}{\sigma^2_G} = CR(X, G)$

Proof:

(i) $Cov(G, E(G \mid X)) = E(E(GE(G \mid X)|X)) - EGE(E(G \mid X)) =$

$$= E(E^2(G \mid X)) - E^2(E(G \mid X)),$$

(ii): Let δ(X) be any function with finite variance.

Put, $A = \sigma^2_{E(G|X)}$; $B = Cov(E(G|X), \delta(X))$, $C = \sigma^2_G$; $D = \sigma^2_\delta$, Then

$$\rho^2(G, E(G|X) + \delta(X)) = \frac{(A+B)^2}{C(A+D+2B)}, \tag{6}$$

$$\frac{\sigma^2_{E(G|X)}}{\sigma^2_G} = \frac{A}{C}, \tag{7}$$

$$\frac{(A+B)^2}{C(A+D+2B)} \le \frac{A}{C} \Leftrightarrow B^2 \le AD. \quad \square \tag{8}$$

The latter inequality follows from the Cauchy Schwarz inequality. This is similar to a result in [5].

The correlation ratio of X to G may be taken as the general global, variance based sensitivity measure of G to X. This may be understood by recalling the simple relation:

$$Var(G) = Var(E(G|X)) + E(Var(G|X)).$$

Dividing both sides by Var(G), we may interpret *CR(X,G)* as the percentage of the variance of G which is explained by X.

Note that the correlation ratio is always positive, and hence gives no information regarding the direction of the influence. Note also that in general $CR(G,X) \neq CR(X,G)$

The following propositions explore some properties of the correlation ratio.

Proposition 4.*3* *Let G(X, Y) = f(X) + h(Y) where f and g are invertible functions with* $\sigma_f^2 < \infty, \sigma_h^2 < \infty$*, and X,Y are not both simultaneously constant (*$\sigma_G^2 > 0$*). If X and Y are independent then*

$$\rho^2(G, E(G|X)) + \rho^2(G, E(G|Y)) = 1.$$

Proof:

We have $E(G|X) = E(G|f(X))$, and $h(Y) \perp E(G|f(X))$, $f(X) \perp E(G|h(Y))$; therefore,

$$\begin{aligned}
\sigma_G^2 &= Cov(G,G) = Cov((G, f(X) + h(Y)) \\
&= Cov(G, f(X)) + Cov(G, h(Y)) \\
&= Cov(E(G|f(X)), f(X) + Cov(E(G|h(Y)), h(Y)) \\
&= Cov(E(G|f(X)) + E(G|h(Y)), f(X) + h(Y)) \\
&= Cov(E(G|f(X)) + E(G|h(Y))), G) \\
&= Cov(E(G|X) + E(G|Y), G) = \sigma^2_{E(G|X)} + \sigma^2_{E(G|Y)}.
\end{aligned}$$

The result now follows with Proposition (4.2). □

Proposition 4.*4* *Let G=G(X,Y), with Cov(E(G|X), E(G|Y)) = 0* then

$$\rho^2(G, E(G|X)) + \rho^2(G, E(G|Y)) \le 1.$$

Proof:

$$\rho(E(G|X), G - E(G|Y)) = \frac{Cov(E(G|X), G - E(G|Y)}{\sigma_{E(G|X)}\sqrt{\sigma_G^2 - \sigma^2_{E(G|Y)}}} = \frac{\sigma_{E(G|X)}}{\sqrt{\sigma_G^2 - \sigma^2_{E(G|Y)}}} \le 1,$$

$$\sigma^2_{E(G|X)} + \sigma^2_{E(G|Y)} \le \sigma_G^2. \quad \square$$

4.1.Computing the correlation ratio

The computations frequently use Monte Carlo methods. Efficiency in this context usually means on-the-fly. That is, we would like to perform all necessary calculations on a sample, then discard the sample and proceed to the next sample. A computation which involves retaining the entire sample is not efficient.

Computing the correlation ratio may be difficult in some cases. However, if we can sample *Y'* from the conditional distribution *(Y|X)* independently of *Y*, and if the evaluation of G is not too expensive, then the following simple algorithm may be applied (Ishigami and Homma [4]) :

1. Sample *(x,y)* from *(X,Y),*
2. Compute *G(x,y),*
3. Sample *y'* from *(Y|X = x)* independent of *Y = y,*
4. Compute *G' = G(x, y')*
5. Store *Z = G * G'*
6. Repeat

The average value of Z will approximate $E(E^2(G|X))$, from which the correlation ratio may be computed as

$$\frac{E(E^2(G|X)) - E^2(G)}{\sigma_G^2}.$$

Of course, if *Y* and *X* are independent, then this algorithm poses no problems. If *Y* and *X* are not independent, then it may be difficult to sample from *(Y|X).* In this case there is no alternative to the "pedestrian" method: save a large sample, compute $E(G \mid X = x_i \pm \varepsilon)$ for suitable $x_{1,\ldots,} x_n$, and compute the variance of these conditional expectations. To do this for a large number of variables can be slow.

The notion of the correlation ratio can be generalized by introducing the following definition

Definition 4.5 [Generalized correlation ratio] *Correlation ratio of* $X_{i_1},....,X_{i_s}$ *to G is*

$$CR\left(G,\left\{X_{i_1},....,X_{i_s}\right\}\right) = \frac{Var\left(E\left(G\middle|\left\{X_{i_1},....,X_{i_s}\right\}\right)\right)}{Var(G)}.$$

5. LOCAL PROBABILISTIC SENSITIVITY MEASURES

The local sensitivity measure (1) is intended to measure the rate of change with respect to *Z* of "some function" of *X|Z* at a given point. For the uncorrelated linear model, "global" and "local" are equivalent, hence the global and local measures should coincide. This motivates choosing "some function" as a normalized conditional expectation in (1). In fact, local probabilistic and global sensitivity measures may be be seen as dual, in the following sense. Apply the Taylor expansion to *E(X|Z)*:

$$\begin{aligned} Cov(Z,X) &= Cov(Z - E(Z), E(X|Z)) \\ &\sim Cov\left(Z - E(Z), E(X|z_0) + (Z - z_0)\frac{\partial E(X|z_0)}{\partial z_0}\right) = \sigma_Z^2 \frac{\partial E(X|z_0)}{\partial z_0}. \end{aligned}$$

Thus, if the regression of *X* on *Z* is linear, then higher order terms vanish and

$$\frac{\partial E(X|z_0)}{\partial z_0} = \frac{Cov(Z,X)}{\sigma_Z^2}. \tag{9}$$

which may be compared with (4). If the roles of *Z* and *X* were reversed in the linear uncorrelated model, then (9) would express the rate of change of *X* with respect to *Z*. Of course, these roles cannot be reversed, as *Z* is correlated with $X_1,\ldots,X_n$. However, the regression *E(X|Z)* can be linear, indeed this arises for linear normal, mixed normal and elliptical models (correlated as well as uncorrelated) ([1]). Hence in the uncorrelated linear models with linear regression of *X* on *Z*, we have

$$\frac{\partial Z}{\partial x_0} = \frac{Cov(Z,X)}{\sigma_X^2}, \tag{10}$$

$$\frac{\partial E(X|z_0)}{\partial z_0} = \frac{Cov(Z,X)}{\sigma_Z^2}. \tag{11}$$

Note that the quantities on the right hand side are global, whereas those on the left are local. As seen above, the correlation ratio of X to Z is the maximal squared correlation attainable between Z and some function of $f(X)$ of X with finite variance. In the same vein, we could ask, `which function $f(X)$ of X maximizes

$$\left(\frac{\sigma_Z}{\sigma_{f(X)}} \frac{\partial E(f(X)|z_0)}{\partial_{Z_0}} \right)^2 ?$$

We conjecture that the maximum is attained for $f(X) = E(Z|X)$.

We first discuss methods of computing and approximating $\frac{\partial E(X|z_0)}{\partial z_0}$. These methods cannot always be applied in practice, but serve as a benchmark. In the following section we develop a method based on the above duality.

5.1.Computing $\frac{\partial E(X|z_0)}{\partial z_0}$

We discuss methods for computing the derivative of a conditional expectation. In general, if the rightmost integral converges absolutely for all z_o;

$$\frac{\partial E(X|z_0)}{\partial z_0} = \frac{\partial}{\partial z_0} \int x f(x|z_0) dx = \int \left(\frac{\partial x}{\partial z_0} f(x|z_0) + x \frac{\partial f(x|z_0)}{\partial z_0} \right) dx.$$

Alternatively, we could compute the conditional expectation $E(X|Z)$ directly and take its derivative. Assume for example that X,Y are independent and uniformly distributed on [0, 1], and let $Z=Z(X,Y)$ be sufficiently differentiable in both arguments. To compute the expectation of X given $Z=z_o$, we define a density along the contour $Z=z_o$ which is proportional to arc length. If the contour is simple we may parametrize arc length in terms of x and write $z_o=Z(x,y(x))$ The arc length element, ds and conditional expectation are given by

$$ds = \sqrt{dx^2 + dy^2} = dx\sqrt{1 + (dy/dx)^2}$$

$$E(X|z_0) = \int x f(X|z_0) dx = \frac{\int x\sqrt{1 + (dy/dx)^2}\, dx}{\int \sqrt{1 + (dy/dx)^2}\, dx}$$

The reader may verify the following examples:

Example 5.1

$$Z = 2X + Y;\ f(x|z) = 2/z;\ 0 < x < z/2,$$

$$Z = XY;\ f(x|z) = \frac{\sqrt{1 + z^2/x^4}}{\int_z^1 \sqrt{1 + z^2/x^4}\, dx};\ 0 \le z \le 1,\ z < x < 1,$$

$$Z = X^2 Y;\ f(x|z) = \frac{\sqrt{1 + 4z^2/x^6}}{\int_{\sqrt{z}}^1 \sqrt{1 + z^2/x^6}\, dx};\ 0 \le z \le 1,\ \sqrt{z} < x < 1,$$

$$Z = X^2 + Y^2;\ f(x|z) = \frac{\sqrt{1 + x^2/(z^2 + y^2)}}{\int_0^{\sqrt{z}} \sqrt{1 + x^2/(z^2 + y^2)}\,dx};\ 0 \le z \le 1,\ 0 < x < \sqrt{z}\,.$$

5.2. Linear approximations

Since (9) does not depend on z_o, it does not provide a good basis for linear approximations. For random variables *X, Y* let *Z=Z(X,Y)* and suppose for some analytic function G we can write *X=G(Z,Y).* The Taylor expansion gives:

$$X = G(z_0, y_0) + (Z - z_0)\partial_{z_0} G + (Y - y_o)\partial_{y_0} G + (Z - z_0)(Y - y_o)\partial_{z_0}\partial_{y_0} G + \frac{(Y - y_o)^2}{2}\partial^2_{y_0} G + \frac{(Z - z_o)^2}{2}\partial^2_{z_0} G + HOT$$

Take $y_o = E(Y|z_o)$ and take conditional expectations on both sides with respect to z_o. The first order terms, the cross term and the second order term in *Z* all vanish. We find

$$E(X \mid z_0) \sim G(z_0, E(Y \mid z_0)) + \tfrac{1}{2}\partial^2_{y_0}(GVar(Y \mid z_0)). \tag{12}$$

We now take derivatives on both sides with respect to zo. Retaining only the first term yields *estimate 1*:

$$\frac{\partial E(X \mid z_0)}{\partial z_0} \sim \frac{\partial}{\partial z_0} G(z_0, E(Y \mid z_0))\,. \tag{13}$$

Retaining both terms yields *estimate 2*:

$$\frac{\partial E(X \mid z_0)}{\partial z_0} \sim \frac{\partial}{\partial z_0}\left(G(z_0, y_0) + \tfrac{1}{2}\partial^2_{y_0}(GVar(Y \mid z_0))\right). \tag{14}$$

Note that both these estimates depend on z_o.

6. LINEARIZATION VIA RE-WEIGHTED MONTE CARLO SIMULATION

The methods of the previous section are not generally useful in practice. Indeed, *estimate 1* will typically require $\frac{\partial}{\partial z_0} E(Y \mid X)$ to estimate $\frac{\partial}{\partial z_0} E(X|Z)$ which is just as hard to calculate as quantity being estimated. *Estimate 2* requires $\frac{\partial}{\partial z_0} Var(Y|z_0)$ which is more difficult to estimate than $\frac{\partial}{\partial z_0} E(X|z_0)$.

A new method of calculating $\frac{\partial E(X|z_0)}{\partial z_0}$ suggested by Meilijson is currently being developed. The idea is the to make the duality relation (9) approximately true by re-weighting the sample emerging from a Monte Carlo simulation. Since *E(X|Z)* can be expanded around z_o as

$$E(X \mid Z) = E(X \mid Z = z_0) + (Z - z_0)\frac{\partial E(X|z_0)}{\partial z_0} + \frac{1}{2}(Z - z_0)^2 \frac{\partial^2 E(X|z_0)}{\partial z_0^2} + HOT$$

then

$$Cov(X,Z) = Cov(E(X \mid Z), Z) =$$

$$\sim \frac{\partial E(X \mid z_0)}{\partial z_0} VarZ + \frac{1}{2}\frac{\partial^2 E(X \mid z_0)}{\partial z_0^2}\left\{E(Z - z_0)^3 + (z_0 - EZ)(E(Z - z_0)^2)\right\}.$$

So if we assign a "local distribution" to Z such that the terms between curly brackets are equal to zero then

$$\frac{\partial \overline{E}(X \mid z_0)}{\partial z} = \frac{\overline{Cov}(X,Z)}{\overline{Var}(Z)}.$$

To achieve this the local distribution should be chosen so that

$$\overline{E}Z = z_o$$

and

$$\overline{E}(Z - z_o)^3 = z_o$$

where $\overline{Z}$ means Z with a local distribution. We want this distribution to be as close as possible to the distribution of Z. In our case we take the distribution which minimizes the relative information with respect to the original distribution of Z.

7. RESULTS

The first table presents the theoretical results for the functions given in example (5.1). X,Y are independent and uniform on [0, 1]. The theoretical values have been computed with MAPLE. Note that for $Z=2X+Y$ the estimates are exact, as the regression is piece-wise linear: $E(X|z)$=z/4; $0 < z < 1$; $2 < z < 3$; but $E(X|z) = z/2$; $1 < z < 2$.

Z	z_o	$\frac{\partial E(X \mid z_0)}{\partial z_0}$	*est1*	*est2*
$2X+Y$	0.25	0.25	0.25	0.25
	0.5	0.25	0.25	0.25
	1.5	0.5	0.5	0.5
	2.5	0.25	0.25	0.25
X^2+Y^2	0.1	1.0066	1.2189	1.2024
	0.5	0.4488	0.5451	0.5063
	0.9	0.3355	0.4056	0.3538
XY	0.1	1.0724	1.8046	1.7764
	0.5	0.6342	0.7768	0.7095
	0.9	0.5180	0.5361	0.5271
X^2Y	0.1	1.2698	1.8013	1.5733
	0.5	0.4454	0.5099	0.4501
	0.9	0.2746	0.2816	0.2746

Table 1: Comparison of theoretical values and linearized estimates.

Table 2 shows the re-weighted estimates for the same models as in Table 1. To show the sampling fluctuations, the results have been computed on five runs, each run using 10,000 samples. The weights defining the local distribution for *Z* have been computed with MOSEK. We see that the results are reasonably stable and are generally between those of *estimate 1* and *estimate 2* in Table 1. An exception occurs for $Z = 2X + Y$; in this case the second derivative of the regression function $E(X|z)$ does not exist for $z=1$. This suggests that better results could be obtained by first defining a window around the value z_o and applying the re-weighting method within this window.

Z	z_o	$\frac{\partial E(X\|z_0)}{\partial z_0}$	*1*	*2*	*3*	*4*	*5*
2X+Y	0.25	0.25	0.2577	0.2387	0.2447	0.2500	0.2436
	0.5	0.25	0.2646	0.2496	0.2603	0.2514	0.2560
	1.5	0.5	0.4021	0.4007	0.3966	0.4022	0.4002
	2.5	0.25	0.2484	0.2440	0.2475	0.2540	0.2602
X^2+Y^2	0.1	1.0066	1.1626	1.2117	1.1473	1.1391	1.1568
	0.5	0.4502	0.5309	0.5163	0.5428	0.5238	0.5317
	0.9	0.3355	0.4388	0.4441	0.4440	0.4411	0.4475
XY	0.1	1.0724	1.4854	1.5025	1.5200	1.4916	1.6087
	0.5	0.6342	0.6939	0.7018	0.7031	0.6955	0.6860
	0.9	0.5180	0.4666	0.5908	0.4641	0.5108	0.5182
X^2Y	0.1	1.2698	2.3313	2.3890	2.3878	2.3453	2.3634
	0.5	0.4454	0.5784	0.5839	0.5822	0.5808	0.5782
	0.9	0.2746	0.2706	0.2974	0.2498	0.2636	0.2717

Table 2: Comparison of theoretical values and re-weighting method, five runs of 10,000 samples.

The results in Table 3 are obtained by drawing 100,000 samples, and conditionalizing on the window $Z \in (z_o - 0.2, z_o + 0.2)$. Two runs are shown; Ñ indicates number of samples in the conditional distribution on each run. We note that for $Z = X^2Y = 0.1$ the results are poor, despite the fairly large number of samples falling in the window. The function $Y(x) = 0.1/x^2$; $x \in [0.1,1]$ is highly non-linear; the derivative ranges over 4 orders of magnitude. Reducing the window size to 0.05 returns results comparable to estimate 2.

Z	*zo*	$\frac{\partial E(X\|z_0)}{\partial z_0}$	1	$\tilde{N}_1$	2	$\tilde{N}_2$
2X+Y	0.25	0.25	0.2530	5021	0.2464	4832
	0.5	0.25	0.2573	10079	0.2608	9874
	1.5	0.5	0.4943	19867	0.4894	20015
	2.5	0.25	0.2570	10047	0.2501	9949
X^2+Y^2	0.1	1.0066	1.1564	23774	1.1655	23593
	0.5	0.4502	0.4610	31320	0.4509	31300
	0.9	0.3355	0.4141	29323	0.4180	29347
XY	0.1	1.0724	1.5140	66131	1.5260	66072
	0.5	0.6342	0.6438	28813	0.6464	28713
	0.9	0.5180	0.5167	5020	0.5363	5038

X^2Y	0.1	1.2698	2.3759	79690	2.3580	79403
	0.5	0.4454	0.4564	17792	0.4469	17855
	0.9	0.2746	0.2689	2638	0.2871	2658

Table 3: Comparison of theoretical values with re-weighting method results with window.

Table 4 shows results for *X, Y* independent standard normal. There are five runs with 10,000 samples per run and no window (the use of a window did not improve results). The results for *Z = min(3-X,3-Y)=0* are quite bad and quite unstable. This presumably reflects the small number of samples in the region *z = 0*. The other values are quite acceptable.

Z	z_o	$\frac{\partial E(X\|z_0)}{\partial z_0}$	1	2	3	4	5
X+Y	0	0.5	0.5058	0.4964	0.5000	0.4921	0.4965
	2	0.5	0.5387	0.4831	0.4983	0.5030	0.5086
	3	0.5	0.4292	0.4415	0.4840	0.4681	0.5826
2X+Y	0	0.4	0.4005	0.4037	0.4012	0.3967	0.3939
	1	0.4	0.4006	0.3992	0.3497	0.3996	0.3967
	3	0.4	0.4033	0.3996	0.3927	0.4027	0.4027
Min{3-X,3-Y}	0	-0.5067	-0.8779	0.7641	0.3178	1.1178	-0.6585
	1	-0.5568	-0.6478	-0.5839	-0.5445	-0.7650	-0.4516
	2	-0.6852	-0.6925	-0.6686	-0.7204	-0.6676	-0.6792
	3	-0.8183	-0.8061	-0.8031	-0.8002	-0.7979	-0.8048

Table 4: Comparison of theoretical and re-weighting method results for normals.

With regard to the example *Z=min{3-X,3-Y}* the results are better than those given in section (2), but not overwhelming. With 5,000,000 samples and window $z \in [-0.1,0.1]$ we find

$$\frac{\partial E(X \mid Z = z)}{\partial z}\Big|_{z=0} = -0.5029,$$
$$\frac{\partial E(Y \mid Z = z)}{\partial z}\Big|_{z=0} = -0.5038.$$

Needless to say, this number of samples is not realistic in practice. With only 10,000 samples the results were not acceptable. Note that in this case the linearized estimates are not defined, as the function *G* satisfying *X = G(Z,Y)* does not exist.

8. CONCLUSIONS

In the linear model (3), with $X_1,\ldots,X_n$ independent normal, we have observed following relations:

$$\frac{\partial Z}{\partial X} = \frac{Cov(X_i, Z)}{\sigma^2_{X_i}} = \frac{\rho(Z, X_i)\sigma_Z}{\sigma_{X_i}},$$
$$\frac{\partial E(X / Z = z)}{\partial z} = \frac{Cov(X_i, Z)}{\sigma^2_Z} = \frac{\rho(Z, X_i)\sigma_{X_i}}{\sigma_Z}.$$

When these assumptions do not apply, one can still use these relations by way of crude estimation. Thus one can estimate the rate of change of Z with respect to X_i as, $\frac{\rho(Z,X_i)\sigma_Z}{\sigma_{X_i}}$ and one can estimate the rate of change of $E(X|z)$ with respect to z as $\frac{\rho(Z,X_i)\sigma_{X_i}}{\sigma_Z}$. Better estimates can be obtained by the linearization techniques introduced in sections (5.2, 6). In particular the re-weighting approach to linearization gives acceptable results in most of the benchmark problems and is applicable quite generally. None the less, there is room for improvement. We have tried adding additional constraints to the re-weighting algorithm, but did not find any constraints which produced better results for all of the benchmark functions. Reducing the window size generally leads to better results, but of course this drives up the number of samples required. For difficult problems, such as that discussed in section (2) the re-weighting method returns good results only after using a small window with a very large number of samples. It seems likely that the re-weighting method of linearization can still be further improved.

Acknowledgements. We gratefully acknowledge assistance of M. Kallen, M. Poelman.

9. APPENDIX

Let $Z=min(3-X,3-Y)$ with X,Y independent standard normal.

$$\begin{aligned} E(X \mid Z=z) = &\; E(X \mid Z=z, X<Y)P(X<Y) + E(X \mid Z=z, Y \le X)P(Y \le X) \\ = &\; (E(X \mid Y=3-z, X<Y) + E(X \mid X=3-z, X \ge Y))/2 \\ = &\; (E(X \mid X<3-z) + E(X \mid X=3-z, Y<3-z))/2 \\ = &\; E(X \mid X=3-z) + 3-z)/2 \end{aligned}$$

where

$$E(X \mid X<3-z) = \frac{\int_{-\infty}^{3-z} x\phi(x)dx}{\int_{-\infty}^{3-z} \phi(x)dx}$$

and ϕ is the standard normal density, with cumulative distribution function Φ. The partial derivative of the right hand side at $z = 0$ is

$$\frac{-3\phi(3)\Phi(3) + \phi(3)\int_{-\infty}^{3} x\phi(x)dx}{2\Phi(3)^2} - 0.5 = -0.507\,.$$

REFERENCES

[1] R.M. Cooke D.Kurowicka. Conditional, partial and rank correlation for elliptical copula; dependence modeling in uncertainty analysis. Proc. of ESREL 2001, 2001.

[2] M.G. Kendall and A. Stuart. The Advanced Theory of Statistics, vol. 2. Griffin, London, 1967.

[3] J. van Noortwijk R.M. Cooke. Local probabilistic sensitivity measures for comparing form and monte carlo calculations illustrated with dike ring reliability calculations. Computer Physics Communications, 117:86-98, 1998.

[4] T. Homma T. Ishigami. An importance quanti_cation technique in uncertainty analysis for computer models. In Proceedings of the ISUMA '90 First International Symposium on Uncertainty Modeling and Analysis, University of Maryland, pages 398-403, 1990.

[5] Whittle. Probability via Expectation. Springer Verlag, New York, 1992.