Sobol' indices and Shapley value[*]

Art B. Owen

April 2014; Original: September 2013

Abstract

Global sensitivity analysis measures the importance of some input variables to a function f by looking at the impact on f of making large random perturbations to subsets of those variables. Using measures like those of Sobol' we can attribute importance to input variables based on the extent to which they help predict the target function f. There is a longstanding literature in economics and game theory that considers how to attribute the value of a team effort to individual members of that team. The primary result, known as Shapley value, is the unique method satisfying some intuitively necessary criteria. In this paper we find the Shapley value of individual variables when we take 'variance explained' as their combined value. The result does not match either of the usual Sobol' indices. It is instead bracketed between them, for variance explained or indeed any totally monotone game. Because those indices are comparatively easy to compute, Sobol' indices provide effectively computable bounds for the Shapley value.

1 Introduction

An important task in uncertainty quantification is to measure the relative importance of the various inputs to a black box function f. Importance can be quantified via the effects of changing those inputs at random. This leads to a global sensitivity analysis (Saltelli et al., 2008) in which statistical methods based on an analysis of variance (ANOVA) decomposition measure variable importance. The most commonly used measures are the Sobol' indices (Sobol', 1990, 1993).

A very similar problem has been studied for a long time in the economics and game theory literatures. The motivation there is to find a fair way to attribute the value created in a team effort to the individual members of that team. They have studied the setting where one can measure the value that any subset of the team would have created. In that case, the Shapley value (Shapley, 1953) is a uniquely compelling choice; the only way to satisfy four very desirable criteria.

[*] First published in Journal of Uncertainty Quantification, Volume 2, Issue TBD, 2014. Published by the Society of Industrial and Applied Mathematics (SIAM);

The goal of this paper is to draw comparisons between the Sobol' index approach and the theory of Shapley value. For any subset of input variables we define their value to be a measure of how well we can approximate f using only those variables. That value can be quantified via variance which will lead us to Sobol' indices. There are two main versions of Sobol' indices. We will see that neither of them agrees with the Shapley value. But Sobol' indices are easier to compute than the Shapley value. We find that one Sobol' index serves as an upper bound to the Shapley value, while the other is a lower bound. Thus we can use Sobol' indices to bracket the Shapley value. Conversely, Shapley value is a very reasonable midpoint between the two Sobol' indices.

An outline of this paper is as follows. Section 2 gives motivation for variance-based sensitivity measures, and introduces the ANOVA decomposition and Sobol' indices. Section 3 defines the Shapley value and gives expressions for it, when the value of a set of variables is the variance explained by those variables. Neither of Sobol's two variance importance measures for singletons coincide with Shapley value. One omits interaction effects while the other overcounts them relative to Shapley. This is why the Shapley value is bracketed between the two Sobol' indices. Section 4 shows that Sobol' indices satisfy three of the four Shapley criteria. They can be expressed as Shapley values for some alternative variance measures exhibited there. Section 5 has some final remarks.

The Shapley value has been used before for variable importance. Lipovetsky and Conklin (2001) measure variable importance in linear regression with multicollinear predictor variables. They take R^2 as the value for a set of predictors and find Shapley value for each individual predictor. Lindeman et al. (1980) and Kruskal (1987) averaged the increase in R^2 from adding variable j over all 2^{d-1} subsets of other variables that the model could contain. Their measure is equivalent to the Shapley value. Grömping (2007) surveys variable importance measures for linear regression. These R^2-based measures do not cover the present ANOVA context.

2 Global sensitivity, ANOVA and Sobol' indices

The object of our interest is a function f that depends on d input quantities, $x_j \in \mathcal{X}_j$ for $j = 1, \dots, d$. This f typically represents some properties of a product to be manufactured or a model of a complex system such as climate. Generally f is not available in closed form. We assume that it can be computed at any point $\boldsymbol{x} = (x_1, \dots, x_d) \in \mathcal{X} \equiv \prod_{j=1}^{d} \mathcal{X}_j$ of interest to us. Some of the models of interest are very computationally expensive, and then f may be a fast surrogate function that approximates the original model. See Queipo et al. (2005) for background and an example of computer experimentation.

An important variable x_j is one that makes a big difference to f when we change it. Usual sensitivity analysis considers local changes $x_j \to x_j + \mathrm{d}x_j$ and can be studied via the gradient of f near any point $\boldsymbol{x}_0$ of interest. A global sensitivity analysis considers larger changes to $\boldsymbol{x}$ and averages over points $\boldsymbol{x}_0$ of interest. If we put a probability distribution F_j on each $\mathcal{X}_j$ we can then study the distribution of $f(\boldsymbol{x})$ where $\boldsymbol{x}$ is random with distribution $\prod_{j=1}^{d} F_j$. Commonly F_j is the uniform distribution on an interval which after rescaling could be $[0,1]$, but F_j can be more general. If some parts of $\mathcal{X}_j$ are more important to study than others, then we might choose F_j that puts more probability on the more important parts. If random changes to x_j make a small difference to f, then x_j is not important, and conversely when changes to x_j greatly affect f, then x_j is important. It is usual to take x_j independent. The alternative is much more complicated. See Hooker (2007) or Chastaing et al. (2012) for some approaches in that case.

Taking random $\boldsymbol{x}$ yields a random variable $f(\boldsymbol{x}) \in \mathbb{R}$ that we assume satisfies $\mathbb{E}(f(\boldsymbol{x})^2) < \infty$. We write $\mu = \mathbb{E}(f(\boldsymbol{x}))$ and $\sigma^2 = \mathbb{E}((f(\boldsymbol{x}) - \mu)^2)$. The ANOVA partitions the variance σ^2 of $f(\boldsymbol{x})$ among $2^d - 1$ non-empty subsets of $\{1, \dots, d\}$. The set $\{1, \dots, d\}$ is abbreviated to $[d]$ and for $u \subseteq [d]$ we write $-u$ for the set difference $[d] - u$. We frequently need to work with $u \cup \{j\}$ and it is convenient to write it as simply $u + j$. The cardinality of u is denoted $|u|$. If $u = \{j_1, j_2, \dots, j_{|u|}\} \subseteq [d]$ then we write $\boldsymbol{x}_u$ for the tuple $(u_{j_1}, u_{j_2}, \dots, u_{j_{|u|}})$. Similarly, $\mathrm{d}F_u$ is the distribution of $\boldsymbol{x}_u$ over its domain $\mathcal{X}_u$. We take $u \subset v$ to indicate a proper subset, that is $u \subsetneq u$.

The ANOVA is a decomposition of the form

$$f(\boldsymbol{x}) = \sum_{u \subseteq [d]} f_u(\boldsymbol{x})$$

where f_u depends on $\boldsymbol{x}$ only through x_j for indices $j \in u$. There are many decompositions of this form. The ANOVA version is defined recursively by

$$f_u(\boldsymbol{x}) = \int_{\mathcal{X}_{-u}} \Bigl(f(\boldsymbol{x}) - \sum_{v \subset u} f_v(\boldsymbol{x})\Bigr) \,\mathrm{d}F_{-u}(\boldsymbol{x}_{-u}),$$

starting with $f_\varnothing(\boldsymbol{x})$ which takes the value μ for all $\boldsymbol{x}$. First we subtract effects f_v for any strict subsets of u so as not to attribute structure to set u when that structure that has a simpler explanation. Then we average over $\boldsymbol{x}_{-u}$ yielding a function f_u that depends on $\boldsymbol{x}$ through $\boldsymbol{x}_u$ alone.

The ANOVA decomposition satisfies $\int_{\mathcal{X}_j} f_u(\boldsymbol{x}) \,\mathrm{d}F_j(\boldsymbol{x}_j) = 0$ whenever $j \in u$. From this it follows that $\mathbb{E}(f_u(\boldsymbol{x}) f_v(\boldsymbol{x})) = 0$ whenever $u \neq v$. Defining $\sigma_u^2 = \mathrm{var}(f_u(\boldsymbol{x}))$, orthogonality of effects leads to the ANOVA identity

$$\sigma^2 = \mathrm{var}(f(\boldsymbol{x})) = \sum_{u \subseteq [d]} \sigma_u^2.$$

More background on the ANOVA appears in Owen (2013).

Sobol's variable importance indices for subset u are

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2, \quad \text{and} \quad \overline{\tau}_u^2 = \sum_{v : v \cap u \neq \varnothing} \sigma_v^2.$$

They satisfy $\underline{\tau}_u^2 \le \overline{\tau}_u^2$ and $\overline{\tau}_u^2 = \sigma^2 - \underline{\tau}_{-u}^2$. Normalized versions $\underline{\tau}_u^2/\sigma^2$ and $\overline{\tau}_u^2/\sigma^2$ are frequently used to quantify relative importance, but the normalization is not needed here. By identifying variable j with the singleton $\{j\}$ we can interpret $\underline{\tau}^2_{\{j\}}$ and $\overline{\tau}^2_{\{j\}}$ to be two different importance measures for x_j.

Because $\underline{\tau}^2_u = \mathrm{var}(\mathbb{E}(f(\boldsymbol{x}) \mid \boldsymbol{x}_u))$, it is the variance explained by $\boldsymbol{x}_u$ and is therefore the natural choice for the importance of the set u. When $\underline{\tau}^2_u$ is large, it means that the combined effect of all x_j for $j \in u$ makes an important contribution to the variance of $f(\boldsymbol{x})$. When $\overline{\tau}^2_u$ is small, it means that the joint effects of x_j for $j \in u$ make little difference even allowing for all interactions between them and x_k for $k \neq u$. Sometimes that means these variables are so unimportant that they can be 'frozen' at a fixed value (Sobol', 1996) thus reducing the dimension of the function f.

3 Shapley value

Economists use an attribution method known as the Shapley value (Shapley, 1953). The presentation here follows that in Winter (2002). Let $\mathrm{val}(u) \in \mathbb{R}$ be the value attained in a game, by the subset $u \subseteq \{1, \dots, d\} \equiv [d]$. It is always assumed that $\mathrm{val}(\varnothing) = 0$. The Shapley value for individual item $j = 1, \dots, d$ is $\phi_j = \phi_j(\mathrm{val})$, defined below. Shapley value has several appealing properties.

1) (Efficiency) $\sum_{j=1}^d \phi_j = \mathrm{val}([d])$.
2) (Symmetry) If $\mathrm{val}(u+i) = \mathrm{val}(u+j)$ for all $u \subseteq [d] - \{i, j\}$, then $\phi_i = \phi_j$.
3) (Dummy) If $\mathrm{val}(u+i) = \mathrm{val}(u)$ for all $u \subseteq [d]$, then $\phi_i = 0$.
4) (Additivity) If val and val$'$ have Shapley values ϕ and ϕ' respectively then the game with value $\mathrm{val}(u) + \mathrm{val}'(u)$ has Shapley value $\phi_j + \phi'_j$ for $j \in [d]$.

The Shapley value is the only attribution method with these four properties. The Shapley value of an individual variable j is defined by the following formula

$$\phi_j = \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} \bigl(\mathrm{val}(u+j) - \mathrm{val}(u)\bigr).$$

Notice that multiplying $\mathrm{val}(\cdot)$ by a scalar multiplies all of the ϕ_j by that same scalar. We will refer to that property as 'linearity' below.

For variable importance we define the value of the set u to be the variance explained by x_j for $j \in u$, that is $\mathrm{val}(u) = \underline{\tau}^2_u$. With this definition for value, we have

$$\begin{aligned}\phi_j &= \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} (\underline{\tau}^2_{u+j} - \underline{\tau}^2_u) \\ &= \frac{1}{d} \sum_{u \subseteq -\{j\}} \binom{d-1}{|u|}^{-1} \sum_{v \subseteq u} \sigma^2_{v+j}. \qquad (1)\end{aligned}$$

Theorem 1. *Let the value of a subset of variables be* $\mathrm{val}(u) = \underline{\tau}^2_u$*, where* $\underline{\tau}^2_u$ *is derived from an ANOVA decomposition with variance components* σ^2_u*. Then* *the Shapley value of variable j is*

$$\phi_j = \sum_{u\subseteq[d], j\in u} \frac{\sigma_u^2}{|u|}.$$

Proof. Using linearity of the Shapley value, we can write

$$\mathrm{val}(u) = \sum_{v\subseteq[d], v\neq\varnothing} \mathrm{val}^{(v)}(u)$$

where $\mathrm{val}^{(v)}(u) = \sigma_v^2 1_{u=v}$. The Shapley value for $\mathrm{val}^{(v)}$ has $\phi_j^{(v)} = 0$ for $j \notin v$ by the dummy property. It has $\phi_j^{(v)} = \sigma_v^2/|v|$ for $j \in v$ because of symmetry and the fact that $\phi_j^{(v)}$ sum to $\mathrm{val}^{(v)}([d])$ (efficiency). The conclusion then follows from additivity of the Shapley value. □

The defining properties of the Shapley value let us avoid a lengthy manipulation of binomial coefficients that would follow from simplifying equation (1).

The Sobol' indices for singletons are

$$\underline{\tau}_{\{j\}}^2 = \sigma_{\{j\}}^2 \quad \text{and} \quad \overline{\tau}_{\{j\}}^2 = \sum_{y:j\in v} \sigma_u^2.$$

Neither of these match the Shapley value because they do not sum to $\underline{\tau}_{[d]}^2 = \sigma^2$, and so fail property one, efficiency. As we show next, the problem cannot be fixed by simply rescaling them to have the desired sum.

The index $\underline{\tau}_u^2$ does not take account of the contribution of variable j to variance components σ_u^2 with $j \in u$ and $|u| > 1$. The index $\overline{\tau}_u^2$ includes multiple counting of variance components: the contribution of σ_u^2 for $|u| > 1$ is counted in $\overline{\tau}_{\{j\}}^2$ for every $j \in u$. Neither of these issues can be fixed by re-scaling, that is, by multiplying each ϕ_j by a constant. No rescaling will correct the zero-weight given by $\underline{\tau}_{\{j\}}^2$ to σ_u^2 when $j \in u$ and $|u| > 1$. For $\overline{\tau}_{\{j\}}^2$, a different rescaling would be required for each different cardinality $|u|$.

The Sobol' indices bracket the Shapley value. We easily find that

$$\underline{\tau}_{\{j\}}^2 \le \phi_j \le \overline{\tau}_{\{j\}}^2. \tag{2}$$

The bracketing property (2) holds because every $\sigma_u^2 \ge 0$. These variance components can be expressed in terms of $\underline{\tau}_u^2$ via Mobius inversion, $\sigma_u^2 = \sum_{v\subseteq u}(-1)^{|u-v|}\underline{\tau}_v^2$, an inclusion-exclusion relationship.

In economic contexts, the analogue of σ_u^2 is

$$\beta_u \equiv \sum_{v\subseteq u}(-1)^{|u-v|}\mathrm{val}(v).$$

This quantity can be negative. A bad player might subtract value from a team, or we might simply find that $\mathrm{val}(\{1,2\}) < \mathrm{val}(\{1\}) + \mathrm{val}(\{2\})$, in which case $\beta_{\{1,2\}} < 0$. This may happen even if $\mathrm{val}(\{1,2\}) > \max(\mathrm{val}(\{1\}), \mathrm{val}(\{2\}))$. A game with all $\beta_u \ge 0$ is "totally monotone game". The bracketing inequality (2) extends to any totally monotone game. For example, belief functions (Shafer, 1976) are totally monotone.

4 The Shapley properties of Sobol' indices

When $\mathrm{val}(u) = \underline{\tau}_u^2$, the Sobol' indices for singletons each satisfy three of the four Shapley conditions from Section 1, as we show here. As we saw above they do not satisfy the efficiency condition of summing to total value $\underline{\tau}_{[d]}^2 = \sigma^2$.

To verify the symmetry property, suppose that $\underline{\tau}_{u+i}^2 = \underline{\tau}_{u+j}^2$ holds for all $u \subseteq [d] - \{i, j\}$ where $i \neq j$. It then follows that

$$\sum_{v \subseteq u} \sigma_{v+i}^2 = \sum_{v \subseteq u} \sigma_{v+j}^2$$

holds for all $u \subseteq [d] - \{i, j\}$. For $u = \varnothing$, we find that $\sigma_{\{i\}}^2 = \sigma_{\{j\}}^2$ so that $\underline{\tau}_{\{i\}}^2 = \underline{\tau}_{\{j\}}^2$ and Sobol's lower importance measure $\underline{\tau}^2$, for i and j are then equal. Proceeding by induction on $|u|$ we find that $\sigma_{u+i}^2 = \sigma_{u+j}^2$ for all $u \subseteq [d] - \{i, j\}$. Now

$$\overline{\tau}_{\{i\}}^2 - \overline{\tau}_{\{j\}}^2 = \sum_{w: i \in w} \sigma_w^2 - \sum_{w: j \in w} \sigma_w^2.$$

A set w containing neither i nor j does not appear in the difference above and a set containing them both cancels. If w contains i but not j then we get a term $\sigma_{w'+i}^2$ in the left sum where $w' = w - \{i\}$. But then the term $\sigma_{w'+j}^2$ appears in the right sum and cancels it because $\sigma_{w'+i}^2 = \sigma_{w'+j}^2$. An analogous cancellation takes place for a set w containing j but not i. It follows that Sobol's upper importance measure satisfies $\overline{\tau}_{\{i\}}^2 = \overline{\tau}_{\{j\}}^2$ under this condition. That is, both measures satisfy symmetry.

It is easy to see that both Sobol' measures are linear in the value. Finally suppose that $\overline{\tau}_{u+i}^2 = \overline{\tau}_u^2$ for some i and all $u \subseteq [d]$. It then follows that $\sigma_v^2 = 0$ whenever $i \in v$ and so $\overline{\tau}_{\{i\}}^2 = 0$ in this case. Because $\overline{\tau}_{\{i\}}^2 \geq \underline{\tau}_{\{i\}}^2 \geq 0$, this argument establishes the dummy property for both measures.

The Sobol' indices satisfy three of the four Shapley properties. We can show that Sobol' indices satisfy the Shapley properties but for a different value function. Let

$$\underline{\mathrm{val}}(u) = \sum_{j \in u} \sigma_{\{j\}}^2, \quad \text{and} \quad \overline{\mathrm{val}}(u) = \sum_{v \subseteq u} |v| \sigma_v^2.$$

The first value function only counts singletons, or main effects, in the language of ANOVA. The second value function weights variance components by their cardinality. Liu and Owen (2006) show that

$$\sum_{j=1}^{d} \overline{\tau}_{\{j\}}^2 = \sum_{u \subseteq [d]} |u| \sigma_u^2.$$

It is easy to show that $\sum_{j=1}^d \underline{\tau}_{\{j\}}^2 = \underline{\mathrm{val}}([d])$ and $\sum_{j=1}^d \overline{\tau}_{\{j\}}^2 = \overline{\mathrm{val}}([d])$. To conclude that the Sobol' indices are Shapley values for these altered value functions requires us to verify the other three properties. Linearity is immediate. The remaining two properties follow from essentially the same arguments used above.

5 Discussion

The Sobol' indices for singletons either ignore interactions, or count them multiplicatively. A large $\underline{\tau}^2_{\{j\}}$ tells us that variable j is important and a small $\overline{\tau}^2_{\{j\}}$ tells us that it is not. But outside of such cases, the Shapley value is a compelling choice as a measure of the importance of an input variable, because it shares the variance component σ^2_u equally among all $j \in u$. The computational disadvantage of Shapley values is that they are written in term of all $2^d - 1$ variance components. Where kriging-based emulators are used (e.g., Sacks et al. (1989) and Queipo et al. (2005) among others) the cost of computing Shapley value becomes much more reasonable, at least for small d.

Liu and Owen (2006) present estimators for cardinality moment quantities like $\sum_u |u|^k \sigma^2_u$ where k is an integer between 1 and d inclusive. For the Shapley value we need something like $\sum_{u:j\in u} |u|^{-1}\sigma^2_u$, that is a -1'st moment on sets containing j. It is not clear that there is a shortcut to estimate this quantity without estimating all 2^{d-1} ANOVA variances for subsets u containing j.

A great strength of Sobol' indices is that they are quite easy to measure directly. Specifically, suppose that $\boldsymbol{z}$ is independently sampled from the same distribution as $\boldsymbol{x}$ and construct $\boldsymbol{y}$ by combining components $\boldsymbol{x}_u$ and $\boldsymbol{z}_{-u}$. Then the remarkable identity $\mathbb{E}(f(\boldsymbol{x})(f(\boldsymbol{y}) - f(\boldsymbol{z}))) = \underline{\tau}^2_u$ (Mauntz, 2002; Sobol' et al., 2007) can be used to directly estimate $\underline{\tau}^2_u$ using a single $2d$-dimensional quadrature or a Monte Carlo sample. Similarly, Sobol' (1990) shows that $(1/2)\mathbb{E}((f(\boldsymbol{x}) - f(\boldsymbol{y}))^2) = \overline{\tau}^2_{-u}$ so any desired upper index can also be estimated by a quadrature of dimension at most $2d$.

The easy computation of Sobol' indices suggests that they may sometimes make a suitable substitute for Shapley value. We have shown that the Shapley value for variable importance is bracketed between the two kinds of Sobol' index. This property extends from variance explained to any totally monotone game.

Acknowledgments

This work was supported by grant DMS-0906056 from the U.S. National Science Foundation. I thank three anonymous reviewers and the editor, Jim Berger, for their helpful comments.

References

Chastaing, G., Gamboa, F., and Prieur, C. (2012). Generalized hoeffding-sobol decomposition for dependent variables-application to sensitivity analysis. *Electronic Journal of Statistics*, 6:2420–2448.

Grömping, U. (2007). Estimators of relative importance in linear regression based on variance decomposition. *The American Statistician*, 61(2).

Hooker, G. (2007). Generalized functional anova diagnostics for high-dimensional functions of dependent variables. *Journal of Computational and Graphical Statistics*, 16(3):709–732.

Kruskal, W. (1987). Relative importance by averaging over orderings. *The American Statistician*, 41(1):6–10.

Lindeman, R. H., Merenda, P. F., and Gold, R. Z. (1980). *Introduction to bivariate and multivariate analysis.* Scott, Foresman and Company, Homewood, IL.

Lipovetsky, S. and Conklin, M. (2001). Analysis of regression in game theory approach. *Applied Stochastic Models in Business and Industry*, 17(4):319–330.

Liu, R. and Owen, A. B. (2006). Estimating mean dimensionality of analysis of variance decompositions. *Journal of the American Statistical Association*, 101(474):712–721.

Mauntz, W. (2002). Global sensitivity analysis of general nonlinear systems. Master's thesis, Imperial College. Supervisors: C. Pantelides and S. Kucherenko.

Owen, A. B. (2013). Variance components and generalized Sobol' indices. *Journal of Uncertainty Quantification*, 1(1):19–41.

Queipo, N. V., Haftka, R. T., Shyy, W., Goel, T., Vaidyanathan, R., and Tucker, P. K. (2005). Surrogate-based analysis and optimization. *Progress in Aerospace Sciences*, 41(1):1–28.

Sacks, J., Welch, W. J., Mitchell, T. J., and Wynn, H. P. (1989). Design and analysis of computer experiments (c/r: P423-435). *Statistical Science*, 4:409–423.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M., and Tarantola, S. (2008). *Global Sensitivity Analysis. The Primer.* John Wiley & Sons, Ltd, New York.

Shafer, G. (1976). *A Mathematical Theory of Evidence.* Princeton University Press, Princeton, NJ.

Shapley, L. S. (1953). A value for n-person games. In Kuhn, H. W. and Tucker, A. W., editors, *Contribution to the Theory of Games II (Annals of Mathematics Studies 28)*, pages 307–317. Princeton University Press, Princeton, NJ.

Sobol', I. M. (1990). On sensitivity estimation for nonlinear mathematical models. *Matematicheskoe Modelirovanie*, 2(1):112–118. (In Russian).

Sobol', I. M. (1993). Sensitivity estimates for nonlinear mathematical models. *Mathematical Modeling and Computational Experiment*, 1:407–414.

Sobol', I. M. (1996). On "freezing" unessential variables. *Moscow University Maths Bulletin*, 6:92–94.

Sobol', I. M., Tarantola, S., Gatelli, D., Kucherenko, S. S., and Mauntz, W. (2007). Estimating the approximation error when fixing unessential factors in global sensitivity analysis. *Reliability Engineering & System Safety*, 92(7):957–960.

Winter, E. (2002). The Shapley value. *Handbook of game theory with economic applications*, 3:2025–2054.