Estimating Mean Dimensionality

Ruixue Liu

Art B. Owen

Stanford University

Motivation

- functions per se are objects of study
  - function mining
  - global sensitivity analysis
- quadrature can be easy for low effective dimension
- and some other tasks

Anova

1. For f depending on d input variables
2. $f \in L^2[0,1]^d$ or product of probability spaces
3. Apportion variance to subsets of variables

Sensitivity indices

1. Measure importance of subsets of variables
2. By collecting up variances

Notation

Points	$x = (x^1, \ldots, x^d) \in [0,1]^d$
Subsets	$u \subseteq 1:d \equiv \{1, \ldots, d\}$
Cardinality	$\lvert u \rvert$
Complement	$-u \equiv 1:d - u$
Projections	$x^u \in [0,1]^u$ has x^j for $j \in u$
	e.g. $x^{\{1,2,5\}} = (x^1, x^2, x^5)$ and $x^{\emptyset} = ()$
Moments	$I \equiv \int f(x)dx \qquad \sigma^2 = \int (f(x) - I)^2 dx$

Gluing points

$$x^u : z^{-u} \equiv y \quad \text{where} \quad y^j = \begin{cases} x^j, & j \in u \\ z^j, & j \in -u \end{cases}$$

Dependence

$$\begin{aligned} f(x^u : z^{-u}) \quad &= \quad f(x) \quad \forall x, z \in [0,1]^d \\ &\Longleftrightarrow \quad f(x) \text{ depends only on } x^u \\ &\Longleftrightarrow \quad f(x) \text{ does not depend on } x^{-u} \end{aligned}$$

e.g. $\int_{[0,1]^u} f(x)dx^u$ integrates out x^u, depends only on x^{-u}

Subset decompositions

$$f(x) = \sum_{u \subseteq 1:d} f_u(x) \qquad f_u \quad \text{depends only on} \quad x^u$$

ANOVA Decomposition

$$f(x) = \sum_{u \subseteq \{1,\ldots,d\}} f_u(x)$$

Effect of subset u

$$\begin{aligned} f_u(x) &= \int \Big(f(x) - \sum_{v \subsetneq u} f_v(x) \Big) dx^{-u} \\ &= \int f(x) dx^{-u} - \sum_{v \subsetneq u} f_v(x) \end{aligned}$$

Examples

$$f_\emptyset(x) = \int f(x) dx = I \quad \text{by convention}$$

$$f_{\{j\}}(x) = \int f(x) - I dx^{-\{j\}} = \int f(x) dx^{-\{j\}} - I$$

Variances

$$\sigma^2 = \sum_{u \subseteq \{1,\dots,d\}} \sigma_u^2 \qquad \text{where,} \qquad \sigma_u^2 \equiv \begin{cases} \int f_u(x)^2 dx & u \neq \emptyset \\ 0 & u = \emptyset \end{cases}$$

Variable importance measures Sobol'

$$\underline{\tau}_u^2 = \sum_{v \subseteq u} \sigma_v^2$$

$$\overline{\tau}_u^2 = \sum_{v \cap u \neq \emptyset} \sigma_v^2$$

Identities

$$0 \le \underline{\tau}_u^2 \le \overline{\tau}_u^2 \le \sigma^2$$

$$\underline{\tau}_u^2 + \overline{\tau}_{-u}^2 = \sigma^2$$

$\frac{\underline{\tau}_u^2}{\sigma^2}$ and $\frac{\overline{\tau}_u^2}{\sigma^2}$ are global sensitivity indices

Sobol’ identities

$$\begin{aligned} I^2 + \underline{\tau}_u^2 &= \int f(x^u : x^{-u}) f(x^u : z^{-u}) dx dz^{-u} \\ \underline{\tau}_u^2 &= \frac{1}{2} \int (f(x^u : x^{-u}) - f(z^u : x^{-u}))^2 dx dz^u \end{aligned}$$

Inverse linear relationship

$$\begin{aligned} \underline{\tau}_u^2 &= \sum_{v \subseteq u} \sigma_v^2 \\ \sigma_u^2 &= \sum_{v \subseteq u} (-1)^{|u-v|} \underline{\tau}_v^2 \end{aligned}$$

Recipe

1. (Q)MC estimate of $\underline{\tau}_u^2$
2. Solve for σ_u^2 stable for small $|u|$, cancellation for large $|u|$

Superset importance

$$\theta_u^2 = \sum_{w \supseteq u} \sigma_w^2$$

Examples

For $|u| = 1$, $\theta_{\{j\}}^2 = \overline{\tau}_{\{j\}}^2$ For $|u| = d$, $\theta_{\mathcal{D}}^2 = \sigma_{\mathcal{D}}^2$ (highest interaction)

Solve for σ_u^2

$$\sigma_u^2 = \sum_{w \supseteq u} (-1)^{|w-u|} \theta_w^2$$

Stable for large $|u|$, cancellation for small $|u|$

$$\theta_u^2 = \frac{1}{2^{|u|}} \int \Bigl(\sum_{v \subseteq u} (-1)^{|u-v|} f(x^v : z^{-v}) \Bigr)^2 dx^u dz$$

Effective Dimension

Effective dimension is s in proportion p (eg 0.99) if

$$\sum_{|u|\le s} \sigma_u^2 \ \ge\ p\sigma^2 \qquad \text{(superposition sense)}$$

$$\sum_{u\subseteq\{1,\dots,s\}} \sigma_u^2 \ \ge\ p\sigma^2 \qquad \text{(truncation sense)}$$

We’ll employ superposition

Effective dimension is a percentile eg 99’th

Corresponding moments are useful too

Dimension Distribution

Select random subset $U \subseteq 1:d \qquad \Pr(U=u) = \dfrac{\sigma_u^2}{\sigma^2}$

$$\text{Effective dimension} \le s \iff \Pr(|U| \le s) \ge p$$

$$\iff \sum_{j=1}^{s} \nu(j) \ge p$$

where

$$\nu(j) = \frac{1}{\sigma^2} \sum_{|u|=j} \sigma_u^2, \quad j = 1, \dots, d.$$

Dimension distribution moments

$$\nu(j) = \frac{1}{\sigma^2} \sum_{|u|=j} \sigma_u^2, \quad j = 1, \dots, d.$$

Mean dimension

$$E(|U|) = \frac{1}{\sigma^2} \sum_{j=1}^{d} j\nu(j) = \frac{1}{\sigma^2} \sum_{u \subseteq 1:d} |u| \sigma_u^2$$

Mean square dimension

$$E(|U|^2) = \frac{1}{\sigma^2} \sum_{j=1}^{d} j^2\nu(j) = \frac{1}{\sigma^2} \sum_{u \subseteq 1:d} |u|^2 \sigma_u^2$$

Motivation

Dimension moments easier to estimate

Can provide bounds on quantiles

E.g. $E(|U|) \le 1.05 \implies \nu(1) \ge 0.95$

Moments from variable importance

$$E(|U|) \quad = \quad \frac{1}{\sigma^2}\sum_{j=1}^{d}\overline{\tau}^{\,2}_{\{j\}}$$

Proof $\displaystyle\sum_{j=1}^{d}\overline{\tau}^{\,2}_{\{j\}} = \sum_{j=1}^{d}\sum_{u\cap\{j\}\neq\emptyset}\sigma_u^2 = \sum_{u}\sigma_u^2\sum_{j=1}^{d}1_{u\cap\{j\}\neq\emptyset} = \sum_{u}|u|\sigma_u^2$

Similarly

$$E\bigl(|U|^2\bigr) \quad = \quad (2d-1)E(|U|) - \frac{2}{\sigma^2}\sum_{j=2}^{d}\sum_{k=1}^{j-1}\overline{\tau}^{\,2}_{\{j,k\}}$$

$$E(|U|) \quad = \quad \frac{1}{\sigma^2}\sum_{j=1}^{d}\theta^2_{\{j\}}$$

$$E\bigl(|U|^2\bigr) \quad = \quad \frac{2}{\sigma^2}\sum_{j=2}^{d}\sum_{k=1}^{j-1}\theta^2_{\{j,k\}} + E(|U|).$$

Need d or $d(d+1)/2$ integrals, not 2^d

Example: Asian call with barrier

Parameter values

- stock price: $dS_t = rS_t dt + \sigma_S S_t dW_t$
- payoff function: $(S_T - K)^+ 1_{\{S_t > B, \quad t=1,\dots,T\}}$
- $r = 0.06, \sigma_S = 0.25, S_0 = 40, K = 40, B = 36, T = 10$

Results for $\sigma^2_{1:d}$

Here $d = 10$ and

$$\sigma^2_{1:d} \quad = \quad \theta^2_{1:10} \quad \approx \quad 0.0003\sigma^2$$

Integrand is discontinuous

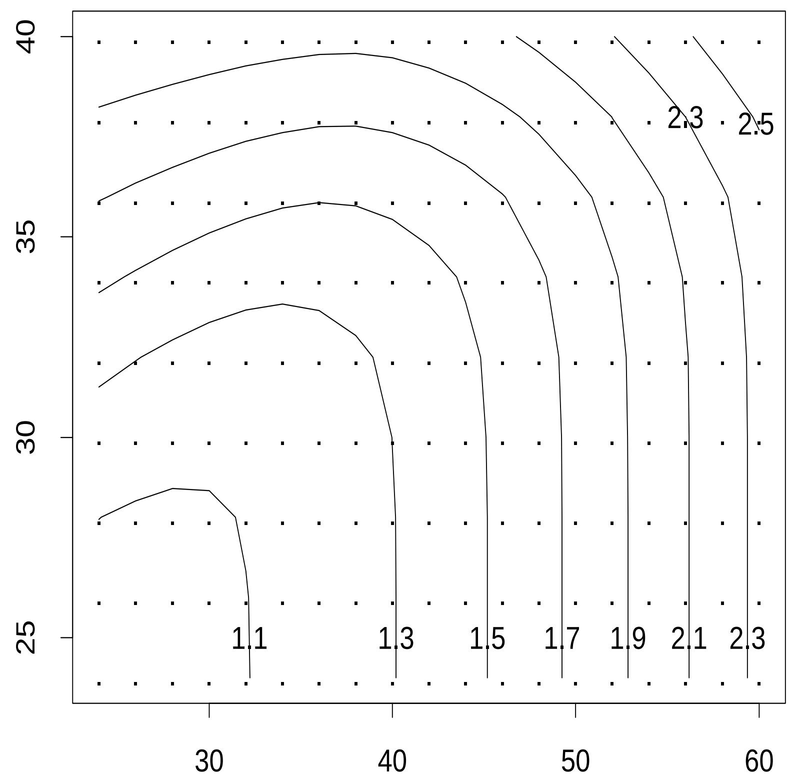
f_u continuous for $|u| < d = 10$ by integration

Discontinuous ANOVA component negligible

. . . as measured by variance

Empirically variance is related to QMC error   Schlier and Caflisch & Morokoff

Mean Dimension of Barrier Option



$E(|U|)$ for grid of strike prices K (horizontal) and barriers B (vertical)

Extreme value functions

$$f(x) = \min(x^1, \ldots, x^d)$$

By calculus (no sampling)

$$\begin{aligned}
\underline{\tau}_u^2 &= \frac{|u|}{(d+1)^2(2d-|u|+2)} \\
E(|U|) &= \frac{2(d+1)}{d+3} \to 2 \\
\lim_{d\to\infty} \nu(k) &= 2^{-k}
\end{aligned}$$

Acknowledgements

Chair: H. Niederreiter & D. Talay

Organizing commitee: M. Bossy & M. Deaconu & E. Tanré