Sensitivity analysis: tutorial

- Alen Alexanderian (NC State)
- Pierre Gremaud (NC State)

July 26, 2018

Workshop on Parameter Estimation for Biological Models

Goal

model: $$y = f(x_1, \ldots, x_p)$$

- response y
- inputs $x_1, \ldots, x_p$

we want to:

> *quantify how uncertainties in the model response can be apportioned to uncertainties in model inputs*

the larger the contribution, the more important the input

Challenges

- no agreement on the meaning of important
- one SA method $\Leftrightarrow$ one definition of "importance"
- f may be
  - a black box (computer code/executable)
  - expensive to evaluate ($\Rightarrow$ few values are available)
- p may be large (high dimensional problem)

High-dim is NOT your friend (I)

it is expensive; GSA can be used to reduced parameter space dimension (inspired by P. Constantine)

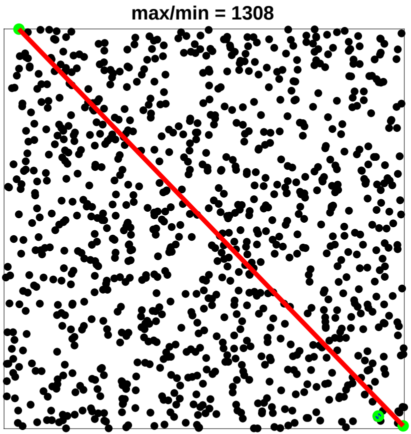
# of parameters p (the dimension)	**# of model runs** (10 points per dim)	**time per run** (1 sec per run)
1	10	10 sec
2	100	1.7 min
3	1,000	17 min
4	10,000	2.8 hrs
5	100,000	28 hrs
6	1,000,000	12 days
10	10^{10}	317 years
20	10^{20}	3 trillion years
		230 $\times$ age of universe

dimension reduction IS your friend

High-dim is NOT your friend (II)

measuring distances gets hard (inspired by Beyer et al. 1999)

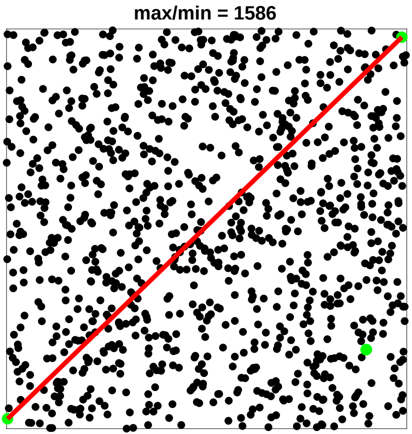
- draw 1000 points from $U([0,1]^p)$
- compute (max distance between 2 points)/(min distance between 2 points)



High-dim is NOT your friend (II)

measuring distances gets hard (inspired by Beyer et al. 1999)

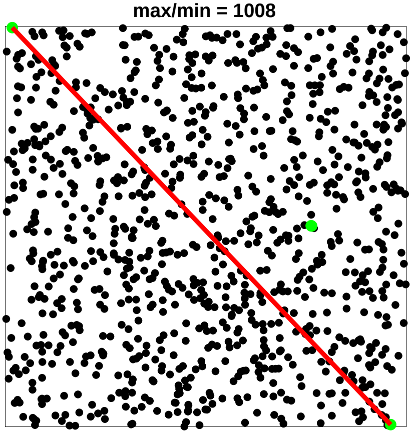
- draw 1000 points from $U([0,1]^p)$
- compute (max distance between 2 points)/(min distance between 2 points)



High-dim is NOT your friend (II)

measuring distances gets hard (inspired by Beyer et al. 1999)

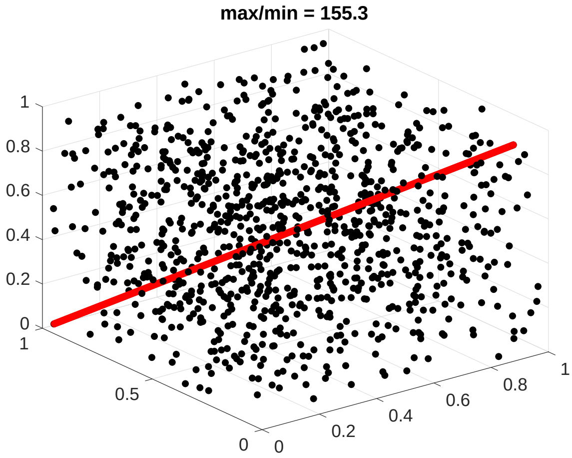
- draw 1000 points from $U([0,1]^p)$
- compute (max distance between 2 points)/(min distance between 2 points)



High-dim is NOT your friend (II)

measuring distances gets hard (inspired by Beyer et al. 1999)

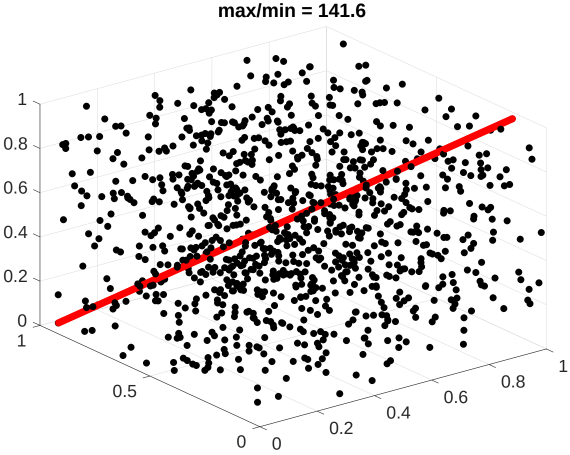
- draw 1000 points from $U([0,1]^p)$
- compute (max distance between 2 points)/(min distance between 2 points)



High-dim is NOT your friend (II)

measuring distances gets hard (inspired by Beyer et al. 1999)

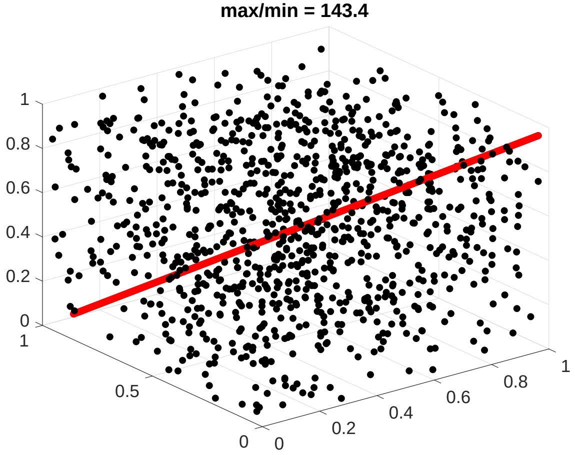
- draw 1000 points from $U([0,1]^p)$
- compute (max distance between 2 points)/(min distance between 2 points)



High-dim is NOT your friend (II)

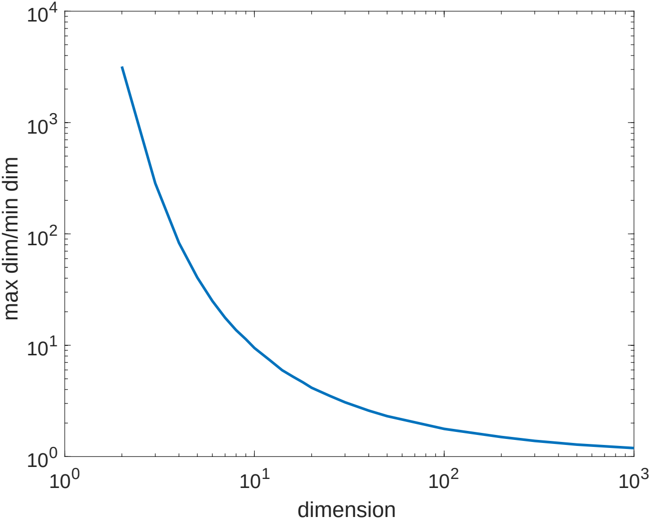
measuring distances gets hard (inspired by Beyer et al. 1999)

- draw 1000 points from $U([0,1]^p)$
- compute (max distance between 2 points)/(min distance between 2 points)



High-dim is NOT your friend (II)

- $\forall p$, results from 100 sets of 1000 points are averaged



High-dim is NOT your friend (III)

the geometry gets weird

- unit (hyper-)spheres and (hyper-)cubes centered at 0
- distance 0-vertex: $p = 2$

$$\sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = \frac{1}{\sqrt{2}} \approx 0.707 < 1 \qquad \text{inside the sphere}$$

- distance 0-vertex: $p = 4$

$$\sqrt{\left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2 + \left(\frac{1}{2}\right)^2} = 1 \qquad \text{on the sphere}$$

- distance 0-vertex: in general = $\frac{\sqrt{p}}{2}$

  $\Rightarrow$ (way) outside the sphere for $p > 4$; $p = 400 \Rightarrow$ dist = 10

What kind of model do you have?

- diagnostics (understanding) vs. prognostics (predictions)
- data-driven vs. law-driven
- law-driven models are good for understanding but are generally overparametrized

Overparametrization is bad

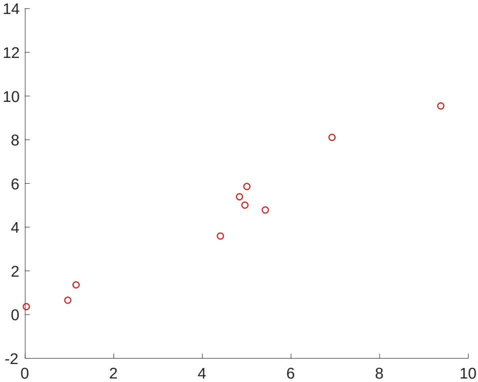
- makes you work in a needlessly high-dimensional space
- loses predictive power

> *In desperation I asked Fermi whether he was not impressed by the agreement between our calculated numbers and his measured numbers. He replied, "How many arbitrary parameters did you use for your calculations?" I thought for a moment about our cut-off procedures and said, "Four." He said, "I remember my friend Johnny von Neumann used to say, with four parameters I can fit an elephant, and with five I can make him wiggle his trunk."*

Freeman Dyson

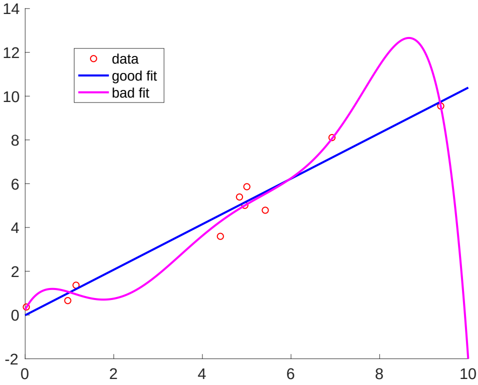
Overparametrization is bad

- makes you work in a needlessly high-dimensional space
- loses predictive power
- easily leads to overfitting



Overparametrization is bad

- makes you work in a needlessly high-dimensional space
- loses predictive power
- easily leads to overfitting



Rationale for SA (inspired by Saltelli)

- model corroboration: is the inference robust?
- research prioritization: which factor most deserves further analysis/measurement?
- model simplification: can factors/compartments be fixed or simplified?
- model reliability: identify factors which interact and may lead to extreme values

Elementary (linear) example (I)

$$y = f(x_1, x_2) = ax_1 + bx_2, \quad a, b > 0$$

1. calculus:

$$S_i = \frac{\partial y}{\partial x_i}, i = 1, 2 \quad \Rightarrow \quad S_1 = a, S_2 = b$$

problems with this approach:

- ignore range of values for x_1 and x_2
- for f nonlinear, this is a local approach (the derivatives have to be evaluated somewhere)
- need to be able to compute derivatives (problematic for black box functions)

Elementary (linear) example (II)

2. Sobol':

- considers x_i's as random variables; for instance

$$x_i \sim N(0, \sigma_i^2)$$

- apportion to them their relative contribution to the variance of the response

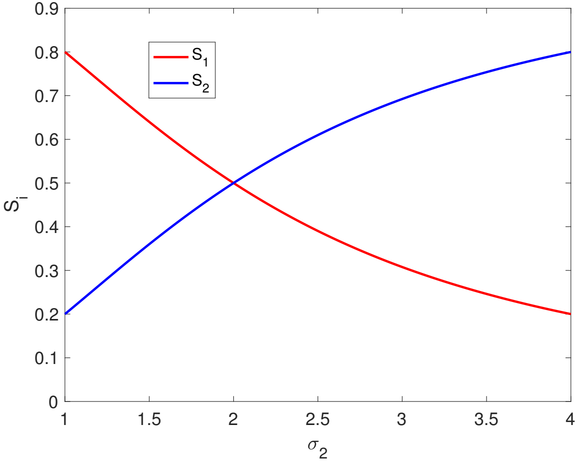
- with above distributions: $y \sim N(0, \sigma_Y^2)$

$$\sigma_Y^2 = a^2\sigma_1^2 + b^2\sigma_2^2 \Rightarrow 1 = \underbrace{\frac{a^2\sigma_1^2}{a^2\sigma_1^2 + b^2\sigma_2^2}}_{S_1} + \underbrace{\frac{b^2\sigma_2^2}{a^2\sigma_1^2 + b^2\sigma_2^2}}_{S_2}$$

- note the importance of the σ_i's!

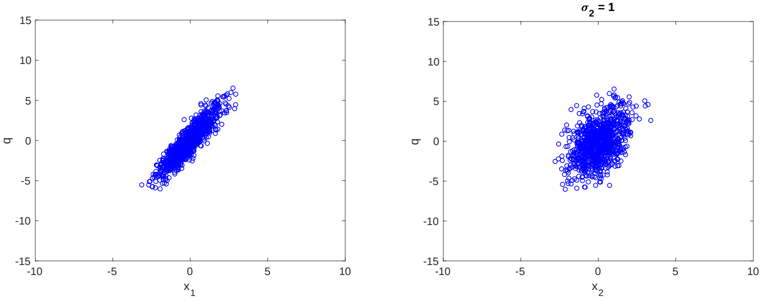
Elementary (linear) example (III)

Let's take $a = 2$, $b = 1$, $\sigma_1 = 1 \Rightarrow S_1 = \frac{4}{4+\sigma_2^2}$ and $S_2 = \frac{\sigma_2^2}{4+\sigma_2^2}$.



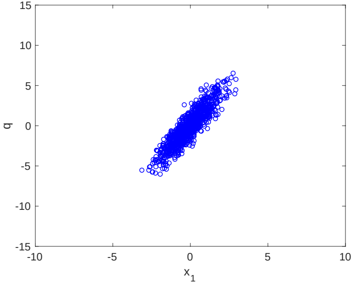
Elementary (linear) example (IV)

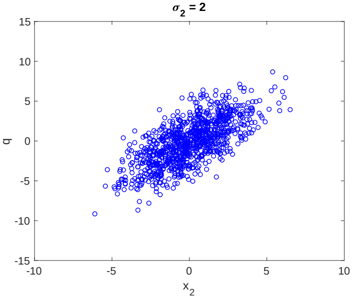
Let's take $a = 2$, $b = 1$, $\sigma_1 = 1 \Rightarrow S_1 = \frac{4}{4+\sigma_2^2}$ and $S_2 = \frac{\sigma_2^2}{4+\sigma_2^2}$.



Elementary (linear) example (IV)

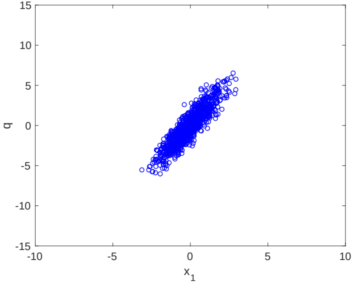
Let's take $a = 2$, $b = 1$, $\sigma_1 = 1 \Rightarrow S_1 = \frac{4}{4+\sigma_2^2}$ and $S_2 = \frac{\sigma_2^2}{4+\sigma_2^2}$.

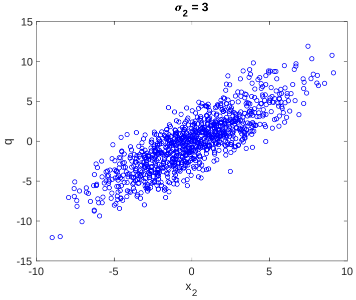




Elementary (linear) example (IV)

Let's take $a = 2$, $b = 1$, $\sigma_1 = 1 \Rightarrow S_1 = \frac{4}{4+\sigma_2^2}$ and $S_2 = \frac{\sigma_2^2}{4+\sigma_2^2}$.





Sobol' indices: same idea for general case

- how about using $\text{var}(y|x_i = x_i^\star)$ to build an importance measure of x_i?
- not a good idea!
  - answer would depend on $x_i^\star$ ($\Rightarrow$ local)
  - it can be that $\text{var}(y|x_i = x_i^\star) > \text{var}(y)$!
- both issues disappear upon averaging

$$\mathbb{E}_{x_i}[\text{var}_{x_{\sim i}}(y|x_i)]$$

indeed by the law of total variance

$$\star \qquad \text{var}_{x_i}(\mathbb{E}_{x_{\sim i}}[y|x_i]) + \mathbb{E}_{x_i}[\text{var}_{x_{\sim i}}(y|x_i)] = \text{var}(y)$$

- first order index: $S_i = \dfrac{\text{var}(\mathbb{E}[y|x_i])}{\text{var}(y)}$

First order Sobol' indices

- first order index: $S_i = \frac{\text{var}(\mathbb{E}[y|x_i])}{\text{var}(y)}$, $i = 1, \ldots, p$
- $0 \leq S_i \leq 1$, $i = 1, \ldots, p$
- S_i "close" to 1 $\Rightarrow$ x_i important
- S_i "small" $\not\Rightarrow$ x_i is not important

Total Sobol' indices

Using again (★) but with $x_{\sim i}$ instead of x_i

$$\mathrm{var}(\mathbb{E}[y|x_{\sim i}]) + \mathbb{E}[\mathrm{var}(y|x_{\sim i})] = \mathrm{var}(y)$$

and thus

$$\underbrace{\mathrm{var}(y) - \mathrm{var}(\mathbb{E}[y|x_{\sim i}])}_{\text{remaining variance if } x_{\sim i} \text{ were known}} = \mathbb{E}[\mathrm{var}(y|x_{\sim i})]$$

- total index: $S_{T_i} = \frac{\mathbb{E}[\mathrm{var}(y|x_{\sim i})]}{\mathrm{var}(y)} = 1 - \frac{\mathrm{var}(\mathbb{E}[y|x_{\sim i}])}{\mathrm{var}(y)}$

$S_{T_i} = 0 \Leftrightarrow x_i$ non-important

$\Leftarrow$:

x_i non-import. $\Rightarrow \mathrm{var}(y|x_{\sim i}) = 0 \Rightarrow \mathbb{E}[\mathrm{var}(y|x_{\sim i})] = 0 \Rightarrow S_{T_i} = 0$

$\Rightarrow$:

$S_{T_i} = 0 \Rightarrow \mathbb{E}[\mathrm{var}(y|x_{\sim i})] = 0 \underset{\mathrm{var} \geq 0}{\Rightarrow} \mathrm{var}(y|x_{\sim i}) = 0 \Rightarrow$
x_i not important

ANOVA (Reader's Digest version)

- assume x_i, iid, $x_i \sim U(0,1)$
- split $x = (x_i, x_{\sim i})$ and decompose f as

$$f(x) = f_0 + f_1(x_i) + f_2(x_{\sim i}) + f_{12}(x_i, x_{\sim i})$$

  where
  - $f_0 = \int f(x)\, dx$,
  - $f_1(x_i) = \int (f - f_0)\, dx_{\sim i}, \quad f_2(x_{\sim i}) = \int (f - f_0)\, dx_i$
  - $f_{12} =$ remainder
- above functions have zero average $\Rightarrow \perp \Rightarrow$

$$\begin{aligned} \text{var}(y) &= \int (f(x) - f_0)^2\, dx = \int f(x)^2\, dx - f_0^2 \\ &= \underbrace{\int f_1^2\, dx}_{\text{var}(f_1)} + \underbrace{\int f_2^2\, dx}_{\text{var}(f_2)} + \underbrace{\int f_{12}^2\, dx}_{\text{var}(f_{12})} \end{aligned}$$

another way to look at things

Sobol' indices can equivalently be defined as

$$S_i = \frac{\mathrm{var}(f_1)}{\mathrm{var}(y)}, \qquad S_{T_i} = \frac{\mathrm{var}_{T_i}}{\mathrm{var}(y)}$$

where

$$\mathrm{var}_{T_i} = \mathrm{var}(f_1) + \mathrm{var}(f_{12}) = \text{ total variance corresponding to } x_i$$

exercise:

$$\mathrm{var}_{T_i} = \frac{1}{2}\iint (f(x) - f(x'))^2\, dx\, dx_i'$$

where $x' = (x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_p)$.

another way to look at things

Sobol' indices can equivalently be defined as

$$S_i = \frac{\mathrm{var}(f_1)}{\mathrm{var}(y)}, \qquad S_{T_i} = \frac{\mathrm{var}_{T_i}}{\mathrm{var}(y)}$$

where

$$\mathrm{var}_{T_i} = \mathrm{var}(f_1) + \mathrm{var}(f_{12}) = \text{ total variance corresponding to } x_i$$

exercise:

$$\mathrm{var}_{T_i} = \frac{1}{2} \iint (\overbrace{f(x) - f(x')}^{\frac{\partial f}{\partial x_i}(\hat{x})(x_i - x_i')})^2 \, dx \, dx_i'$$

where $x' = (x_1, \ldots, x_{i-1}, x_i', x_{i+1}, \ldots, x_p)$.

back to the "calculus approach"

previous exercise $\Rightarrow$ one can show (Sobol', Kucherenko, 2009)

$$S_{T_i} \leq \frac{1}{\pi^2 \operatorname{var}(y)} \underbrace{\int \left(\frac{\partial f}{\partial x_i} \right)^2 dx}_{\nu_i}$$

- ν_i is another importance measure
- ν_i is derivative based rather than variance based
- ν_i may be simpler to compute than S_{T_i} if $\partial_{x_i} f$ is available
- if not, more approximations have to be considered
- parameter distributions are needed for both Sobol' and DGSMs (derivative based global sensitivity measures)

how to actually compute all this?

- lots of integrals of the type $\int g(x)\, dx_1 \ldots dx_p$
- problems with 100's of parameters are the rule in practice, not the exception
- functions with 100's of variables that can be integrated through calculus are (really) the exception, not the rule
- $\Rightarrow$ calculus is hopeless
- standard quadratures that work in dimension 2 or 3 are WAY too expensive in a high dimensional setting
- we need something else!

Monte Carlo integration

- $X \sim U(0,1)$
- key observation

$$I = \int_0^1 g(x)\,dx \text{ can be regarded as } \mathbb{E}[g(X)] = \int_0^1 g(x)\,dx$$

- estimator:

$$I_N = \frac{1}{N}\sum_{i=1}^{N} g(X_i)$$

where the X_i's, $i = 1, \ldots, N$ are N iid $U(0,1)$ RVs

- realizations of I_N are sample means of $g(X)$

$$\frac{1}{N}\sum_{i=1}^{N} g(x_i)$$

Monte Carlo: analysis

- estimates from I_N are exact on average:

$$\mathbb{E}[I_N] = \int_0^1 \frac{1}{N}\sum_{i=1}^{N} g(x)\,dx = \int_0^1 g(x)\,dx = I$$

- in what sense/how fast do realizations of I_N converge to I?
- $\mathrm{var}(I_N) = \mathrm{var}(\frac{1}{N}\sum_{i=1}^N g(X_i)) = \frac{1}{N^2}\mathrm{var}(\sum_{i=1}^N g(X_i))$
  $= \frac{1}{N^2}\sum_{i=1}^N \mathrm{var}(g(X_i)) = \frac{\mathcal{V}}{N}$
  where $\mathcal{V} = \mathrm{var}(g(X)) = \int_0^1 g^2(x)\,dx - I^2$
- Chebyshev inequality $\Rightarrow P\left(|I_N - I| > \frac{\delta}{\sqrt{N}}\right) \leq \frac{\mathcal{V}}{\delta^2}, \forall \delta > 0$
  in other words:

$$I_N \xrightarrow{P} I \text{ at rate } \mathcal{O}(N^{-1/2})$$

Monte Carlo: analysis (II)

- central limit theorem gives "error estimate"

$$I_N - I \approx \sqrt{\frac{\mathcal{V}}{N}}\mathcal{N}$$

  where $\mathcal{N}$ is a $N(0,1)$ RV
- slow rate $\mathcal{O}(N^{-1/2})$ but does not depend on dimension p
- various techniques can be used to speed up convergence
  - variance reduction
  - quasi Monte Carlo, etc...

additional challenges

- the parameters may be correlated
- the quantity of interest may be a vector rather than a scalar (see afternoon session) or a function ...
- f itself may be stochastic
- usually, parameter distributions are unknown (robustness?)
- sampling may be very expensive and/or the dimension very high $\Rightarrow$ metamodels/surrogates

surrogates

- a surrogate $\hat{f}$ approximates: $f \approx f$ in some sense
- $\hat{f}$ is typically built from a "few" samples of f
- $\hat{f}$ is (much) cheaper to evaluate than f
- if $\mu(f)$ is some importance measure of f, we hope

$$\mu(\hat{f}) \approx \mu(f)$$

the above may be true for poor approximations of f (lots of work to be done here!)

example of surrogates

- linear regression: fit the following model to data

$$f \approx \beta_0 + \beta_1 x_1 + \cdots + \beta_p x_p$$

  $\hat{f}$ can be the right hand side or only variables with "significant" β_i's can be retained (screening)
- regression trees/forests, MARS
- Gaussian processes
- Polynomial Chaos Expansion
- and many more...

tutorial summary

- variance based methods work well but may be expensive
- derivative based methods (or elementary effects) tend to be cheaper
- (some) surrogate models can be used for dimension reduction
- sometimes, simple models (linear regression) work shockingly well!
- sometimes, they don't...
- our quantity of interest y is usually not directly what comes out of a "disciplinary solver" but depends on it
- ignoring correlations may be disastrous
- taking correlations into account is hard
- there is much to do: join us!