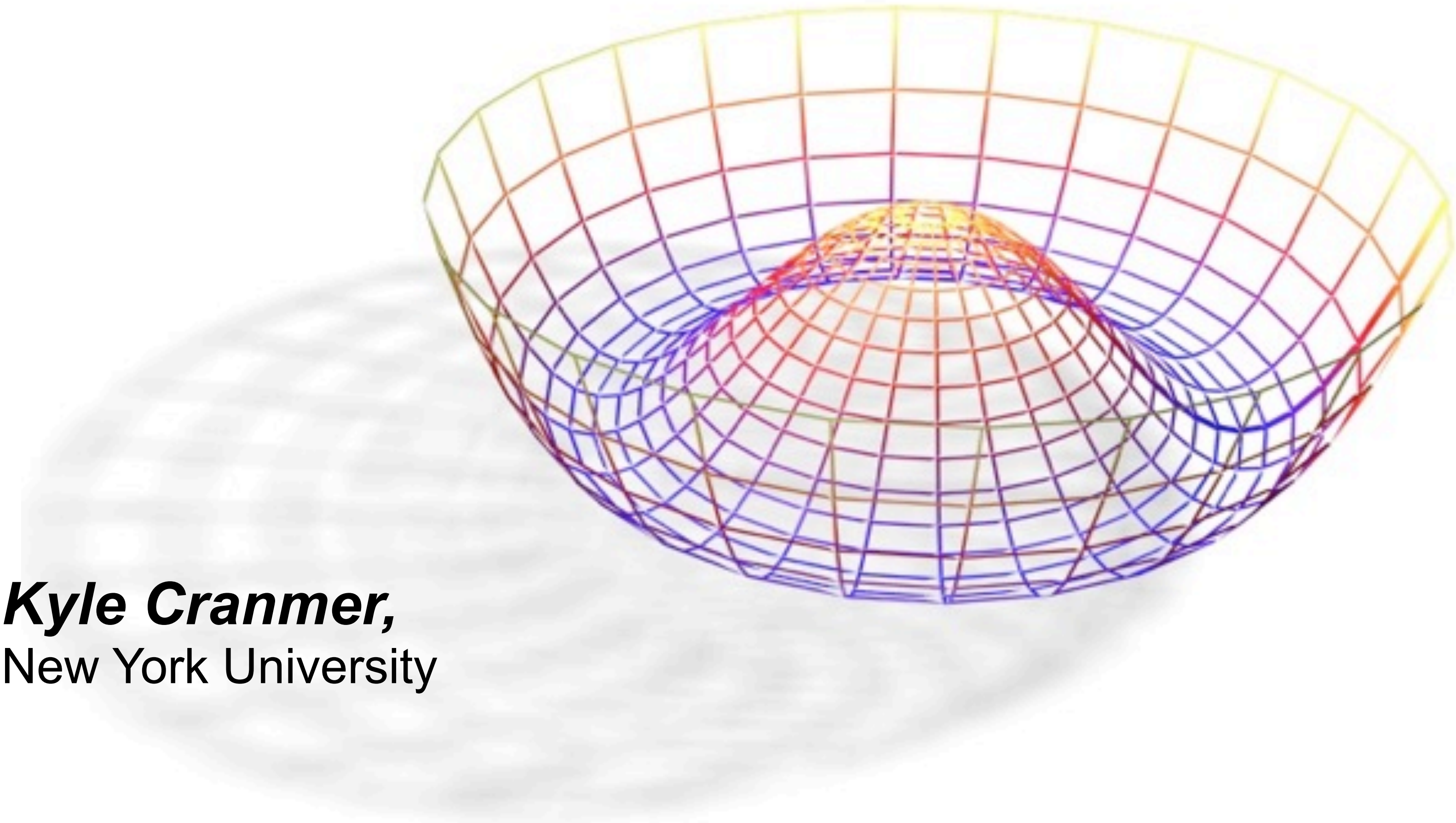
Center for
Cosmology and
Particle Physics

Modeling Systematics



Kyle Cranmer,
New York University

Introduction

We want to assess the systematics associated to known deficiencies in the statistical model being used directly by the statistical framework (ie. the RooFit/RooStats workspaces) and the full simulation and analysis recommendations used to describe the baseline model (aka Geant + some corrections and/or smearing).

Full simulation + smearing + corrections: $f(x|\alpha)$

- very slow, expensive to calculate

Parametrized statistical model $g(x|\alpha)$

- fast, approximates $f(x|\alpha)$, has known deficiencies

where $\alpha=(\mu,\theta)$ includes both the parameters of interest μ and nuisance parameters θ.

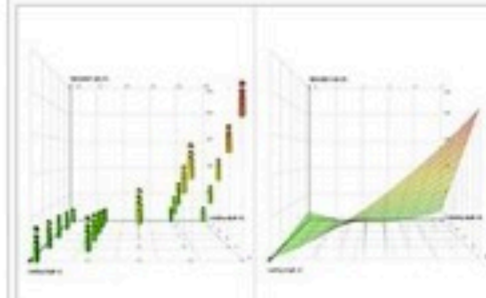
Design of experiments

Design of experiments

From Wikipedia, the free encyclopedia

It has been suggested that *Experimental research design* be merged into this article or section. (Discuss) *Proposed since May 2012.*

In general usage, **design of experiments (DOE)** or **experimental design** is the design of any information-gathering exercises where variation is present, whether under the full control of the experimenter or not. However, in statistics, these terms are usually used for controlled experiments. Other types of study, and their design, are discussed in the articles on opinion polls and statistical surveys (which are types of observational study), natural experiments and quasi-experiments (for example, quasi-experimental design). See Experiment for the distinction between these types of experiments or studies.



Design of experiments with full factorial design (left), response surface with second-degree polynom

In the design of experiments, the experimenter is often interested in the effect of some process or intervention (the "treatment") on some objects (the "experimental units"), which may be people, parts of people, groups of people, plants, animals, materials, etc. Design of experiments is thus a discipline that has very broad application across all the natural and social sciences.

2x2 factorial experiment

	A	B
(1)	−	−
a	+	−
b	−	+
ab	+	+

Currently we are only probing our “response surface” (the acceptance as a function of nuisance parameter) changing one parameter at a time.

- full factorial experiments may not be feasible -->

Nesting non-nested models

Consider the mixture *h* of the non-nested models *f* and *g*

$$h(x|\alpha, \epsilon) = \epsilon f(x|\alpha) + (1-\epsilon) g(x|\alpha)$$

$$h(x|\alpha, \epsilon) = f(x|\alpha)^{\epsilon} g(x|\alpha)^{(1-\epsilon)} \ .$$

Statement of the Problem: How does one provide approximately calibrated confidence intervals in the parameter of interest μ for all values of θ when $\varepsilon=1$ is true, but the likelihood function can only be evaluated at $\varepsilon=0$?

We want

$$f(-2\ln\lambda_g(\mu)|\mu,\epsilon=1,\theta)$$

Instead consider

$$f(-2\ln\lambda_h(\mu,\epsilon=0)|\mu,\epsilon=1,\theta)=\chi^2_D(\Lambda)$$

What is non-centrality parameter?

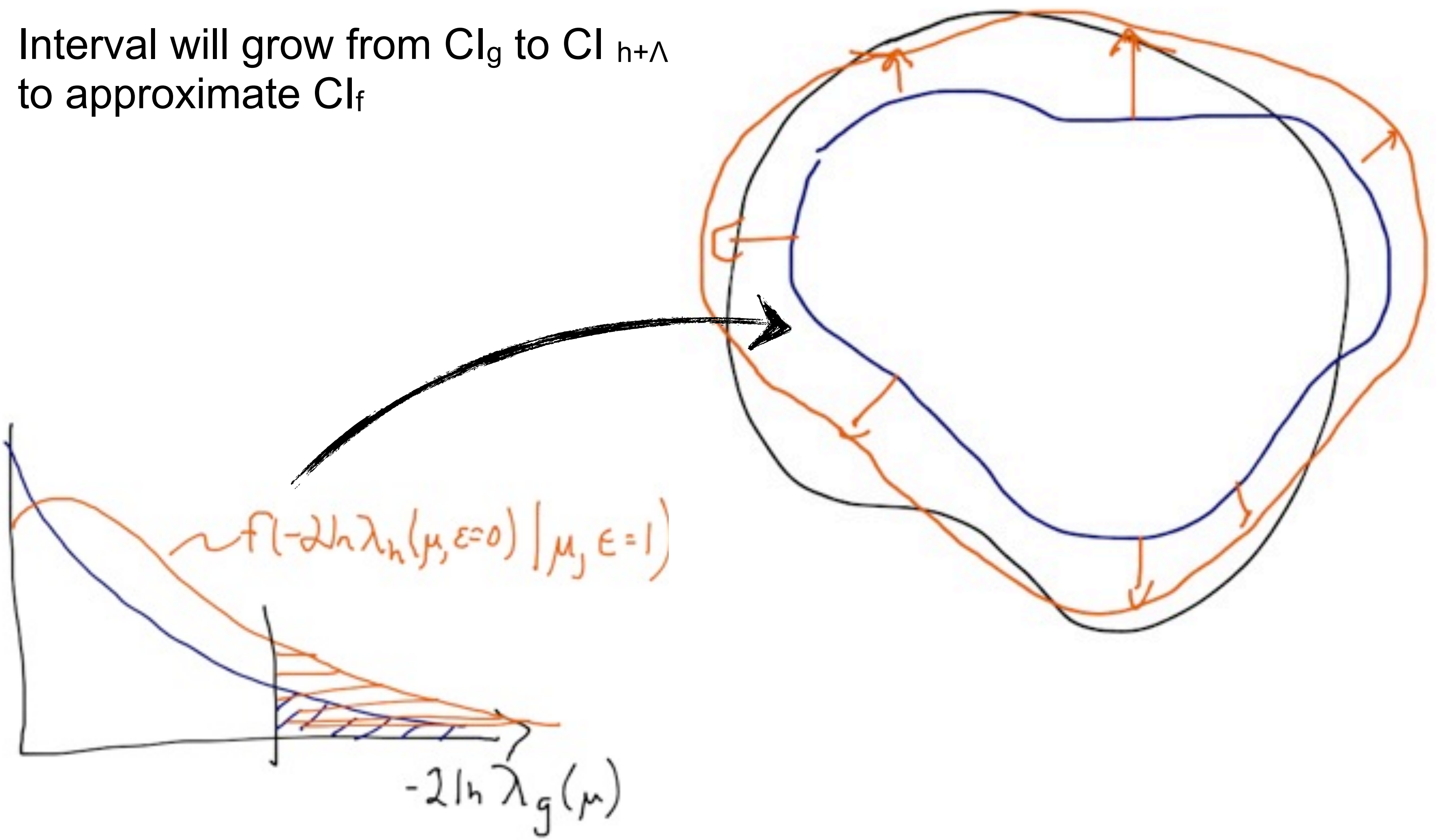
- can determine from Fisher information matrix

$$I_{ij}(\mu,\epsilon,\theta)=E\left[\partial_i \log L\, \partial_j \log L|\mu,\epsilon,\theta\right]$$

- but that requires being able to evaluate full simulation f(μ,θ)
- use relationship to Kullback-Leibler divergence:

A picture

Interval will grow from CI_g to $CI_{h+\Lambda}$ to approximate CI_f



Estimaing non-centrality parameter

Center for Cosmology and Particle Physics

Use relationship between KL-distance and integral of fisher information metric to approximate the non-centrality parameter

$$D_F = \int_0^1 dt \sqrt{I_{ij}(\alpha(t))\dot{\alpha}_i(t)\dot{\alpha}_j(t)} \approx \int_0^1 d\epsilon \sqrt{I_{\epsilon\epsilon}} = \sqrt{I_{\epsilon\epsilon}}\ ,$$

$$\sqrt{2KL(p||q)} \to D_F \qquad \text{as } p \to q$$

$$KL(p||q) = \int dx\, p(x) \ln \frac{p(x)}{q(x)}$$

$$\Lambda \approx 2 \int dx\, f(x|\mu,\hat{\theta}) \ln \frac{f(x|\mu,\hat{\theta})}{g(x|\mu,\hat{\theta})}$$

Requires running simulation at the best fit parameters from $\overline{g(x|\mu,\hat{\theta})}$

- one more full simulation run

Moving past one-parameter-at-a-time

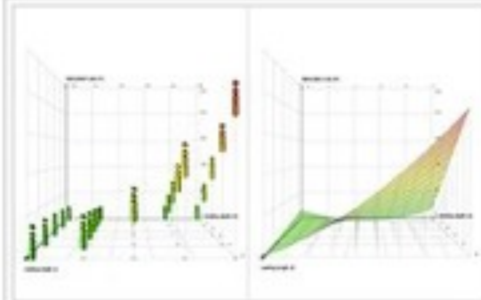
Factorial experiment

From Wikipedia, the free encyclopedia

In statistics, a full **factorial experiment** is an experiment whose design consists of two or more factors, each with discrete possible values or "levels", and whose experimental units take on all possible combinations of these levels across all such factors. A full **factorial design** may also be called a **fully crossed design**. Such an experiment allows studying the effect of each factor on the response variable, as well as the effects of interactions between factors on the response variable.

For the vast majority of factorial experiments, each factor has only two levels. For example, with two factors each taking two levels, a factorial experiment would have four treatment combinations in total, and is usually called a *2×2 factorial design*.

If the number of combinations in a full factorial design is too high to be logistically feasible, a fractional factorial design may be done, in which some of the possible combinations (usually at least half) are omitted.



Designed experiments with full factorial design (left), response surface with second-degree polynomial (right)

Treatment combinations for a 2^{5-2} design

Treatment combination	I	A	B	C	D = AB	E = AC
de	+	−	−	−	+	+
a	+	+	−	−	−	−
be	+	−	+	−	−	+
abd	+	+	+	−	+	−
cd	+	−	−	+	+	−
ace	+	+	−	+	−	+
bc	+	−	+	+	−	−
abcde	+	+	+	+	+	+

When there are many factors, many experimental runs will be necessary, even without replication. For example, experimenting with 10 factors at two levels each produces 2^{10}=1024 combinations. At some point this becomes infeasible due to high cost or insufficient resources. In this case, fractional factorial designs may be used.