A Pitfall in Evaluating Systematic Errors

Jim Linnemann, MSU

PHYSTAT 2007

June 29, 2007

Ideal evaluation of Systematics?

Suppose know (Bayesian) pdf of systematic effects:

$\pi(\varphi) \rightarrow \pi(x,y)$ in 2d examples I'll use

e.g. {x,y} = {Jet Energy Scale factor, luminosity}

Let ***f(x,y)*** be what I am assessing systematic error of

single top cross section

Higgs mass

Upper Limit for SUSY in my channel

Nominal values for systematic params are at $\mathbf{x_o, y_o}$.

Redefine as (0,0), i.e. $(x,y) \rightarrow (x-x_o, y-y_o)$

Similarly, let $\mathbf{g(x,y)} = f(x,y) - f(x_o,y_o) = f - f_o$ so $\mathbf{g(0,0)=0}$

Systematic error = (not quite a variance—f_o not E[f])

$$V = \int dx\, dy\; g^2(x,y)\, \pi(x,y)$$

Instead: Do "Standard" Systematic Evaluation

You have a list of systematics; you ran MC at 0 point

Now run MC at + 1 σ for each systematic

Resulting changes are $d_i = f - f_o$

$$S^2 = \sum d_i^2$$

Report Systematic Error:

$$f_o \pm S$$

the "graduate student" solution?

What Justifies This?

1st order Variance Formula:

$$V = \sum_i \sum_j \frac{\partial f}{\partial x_i} \frac{\partial f}{\partial x_j} Cov(x_i, x_j); \quad \text{eval} \frac{\partial f}{\partial x_i} \text{ at } \vec{x} = 0$$

Nice: avoid distribution assumptions on π, just Cov(x)

Claim can ignore cross terms:

$Cov(x_i,x_j) = 0$: systematics (usually) **uncorrelated**

What if your expt. contributes to PDF fits?

First order, so **good for linear** dependence of *f* on x

But we do a bit better:

finite differences to estimate partials (from MC…)

take into account some nonlinearity, right?

One Factor At A Time: OFAT

From my thesis advisor:

Any physicist can find and fix one problem. It should take 2 things ***both*** *wrong at the same time to confuse a physicist.*

Corollary:

Changing more than one thing at a time is asking for trouble.

V(exact) vs. S^2(OFAT): How well do we do?

Take $x_i \rightarrow z_i = x_i / \sigma_i$

consider $z_i = \pm 1$

Take $\pi(x,y) \sim N(0,a) \times N(0,b)$

$f = x + y$	Truly linear	
$V = a^2 + b^2$	$S^2 = V$	OK as expect
$f = x^2 + y^2$	quadratic	
$V = 3a^4 + 2a^2b^2 + 3b^4$	$S^2 = a^4 + b^4$	not so hot
$f = xy$	bilinear	
$V = a^2b^2$ but	$S = 0$	complete failure

What went wrong?

Quadratic terms underestimated

finite diffs not enough

Covariance = 0 does not protect us from xy

xy and derivatives 0 on axes— as if ***f*** indep. of x,y

xy has twisting of *f* surface:

x derivatives depend on y and vice versa

Must consider off-axis points!

If you go to quadratic terms in Taylor series for V, need both xy and x^2, y^2 (consider rotations!)

Barlow: *run at* $\pm 1\sigma$, $d_i = (f^+ - f^-)/2$

makes quadratic → 0 …if you are asleep

You should notice $(f^+ - f_o) \neq -(f^- - f_o)$

don't forget about the 0 point

“Postdoc Solution?”

You have a list of systematics; you ran MC at 0 point

You run MC at $\pm 1\ \sigma$ for each systematic

Resulting changes are $d_i^{\pm}$

Report Systematic Error:

$$S_u^2 = \Sigma \max\{d_i^+, d_i^-\}^2$$

$$S_d^2 = \Sigma \min\{d_i^+, d_i^-\}^2$$

Report: $f_o{}^{+Su}_{-Sd}$

Here we can check for or even account for asymmetry of uncertainties on effects of systematics; should at least notice quadratic, but still BLIND to xy.

DOE
Design of Experiments
not your US funding agency

OFAT is not a statistician's term of endearment. They wish your thesis advisor had talked to them first:

Always change more than one at a time

Assume each run long enough to measure effects of interesting size

Search for effects in order of likely importance

all linear (main effects)

then bilinear (2^{nd} order interactions)

then 3fold etc

Typically a few effects dominate

One expects "interactions" to be small if *each* main effect of interaction is small (i.e. bare xy term rare)

Interaction: twisting in response plane, i.e. *slope* wrt a variable depends on value of another variable

Typical Goals of DOE

1) Optimization/search

Best pattern of points for searching for

best yield for curing tracker epoxy

least variance of mass vs. cuts

Look for pattern to find a hilltop

which direction, if any, uphill from here?

i.e. good point set for numerical derivatives

2) Robustification (Taguchi)

Look for max or min (stationary)

worry about simultaneously maximizing multiple objectives

Look for ridge (separate important from unimportant params)

strangely named metrics to optimize

Response surface methodology: characterize shape of ***f***

pattern of points for data to fit to 2nd degree curves

geometry to characterize classes of curves:

hilltop, ridge, rising ridge…

“composite designs” add points to basic design to better characterize area (e.g. near maxima)

Glossary

Factor x_i variable; systematic parameter
or from Analysis of Variance: linear combinations
Level values used: 2 level example $\pm 1\sigma$; 3 levels {+ 0 -}
Additive f linear in x_i's
Main Effects linear terms
Active factors main effects which are significant
Interaction multilinear terms $x_i x_j$ or trilinear or higher
Curvature Quadratic term
Respose Surface f(x,y,…)
Twisting of Response Surface $\partial_x f(x,y) \neq \partial_x f(x,0)$
Confounding Fractional Design can't Distinguish all interactions
can detect whether one of class active
ideally confound higher order with lower order
Factorial Design plan for sampling x_i space
Full: L^k all combinations of L levels of k factors
Fractional: L^{k-m} not all combinations
k has "subtracted" off m things confounded

OFAT vs. Design

OFAT advantages

Simpler to set up (fewer changes from nominal)

OK if main effects dominate

Easier to analyze w/o specialized software

One bad run loses less information

Can identify curvature if use 0

Design advantages

Can estimate interactions (or show negligible)

More important savings, the more variables

Less error (all runs contribute to each effect)

Can identify curvature if use 0

All DOE's change more than one factor at a time

2^2 full factorial design 2 levels +1, -1;

Zx	Zy
+1	+1
+1	-1
-1	+1
-1	-1

"Screening designs" in higher dimensions:

Not full 2^k combinations for 2 levels

See all main effects, and Groups of interactions

confound several low order, or low with high order

Calculating Main Effects and Interactions

Look at sign of factors in {x,y} runs:

Sgn {x,y}	++	+-	-+	--
Sgn (xy)	+	-	-	+
run	1	2	3	4

[(1 - 3) + (2 - 4)]/4 = main effect in x

compare the 2 terms for consistency: look for twisting

each term parallel to axes

rather than on axes like [(+0) – (-0)]/2

[(1 - 2) + (3 - 4)]/4 = main effect in y

[(1 - 2) + (4 - 3)]/4 = interaction xy

Or: fit Ax+By+Cxy to points

Sample calculations w/ DOE

without 0 point

$f=x+y$ no interactions

$V = a^2 + b^2$ $= S^2$ OK DOE=V

$f = xy$

$V = a^2b^2$ $S = 0$ BAD DOE=V

$f = x^2 + y^2$

$V = 3a^4+2a^2b^2+3b^4$ $S=a^4+b^4$ Ouch DOE= 0 Worse

DOE from sums of squares of main effects

Both need to explicitly look at 0 point to *notice* curvature
and can be extended to estimate effects better

OFAT CAN'T see xy even with 0 point added, but DOE can

Summary

- Even if your systematics *are* independent, your measurement probably correlates them for you
- If you worry about curvature (up-down asymmetry) you need to worry about xy too
- OFAT is blind to multi-linear (xy-like) effects
- You MUST leave OFAT to see xy-like terms
- OFAT evaluation of systematics misses some of nonlinear effects
- Don't forget the point at nominal parameter values
- Statisticians have heard before from scientists who insist OFAT is the best/only way
- DOE might even help you—worth a think

A speculation:

Could saddle point/Laplace ideas help if assume something about π?

References

See Nancy Reid’s talk at this conference

B. Gunter, Computers In Physics 7 May (1993) (not online alas—complain to AIP)

Can look at NIST handbook or Wiki for definitions and some discussions

Box Hunter & Hunter “Statistics for Experimenters” (Wiley)

less terse, but longer than Cox & Reid