

Lecture 9: Experimental Design Continued

Scribes: Chandrayee Basu and Wenxuan Wang

9.1 Good experiments should have factorial design

Factorial design is accounting for combination or cross product of all the conditions or treatments. For example, in case of CHOMP versus trajOPT, there are two conditions or independent variables: *constraint type* and *order of optimization*.

Constraint type/order of optimization	1 st order	2 nd order
Soft	CHOMP	x
Hard	x	trajOPT

The Table above shows the factorial design of an experiment to understand why CHOMP or trajOPT is the best of the two. If you don't test for all the combinations it is not possible to disambiguate the effect of one variable from the other. So if trajOPT gives better result than CHOMP and we do not test of other combinations of the optimization constraint and optimization order, we cannot form a conclusion if changing the constraint from soft to hard contributed to the improvement or if using 2nd order optimization was better or both were better. It is good if both improved the result but it is not good if one of the independent variables led to worse results. May be we could improve the result even more if we changed the independent variable. Order of exploration also matters. Start with CHOMP, then move to CHOMP-hard constraints. If hard-constraint CHOMP gives better result than CHOMP then move to trajOPT and that is it.

Later we introduced a third variable for collision type: geometry-geometry, point-geometry.

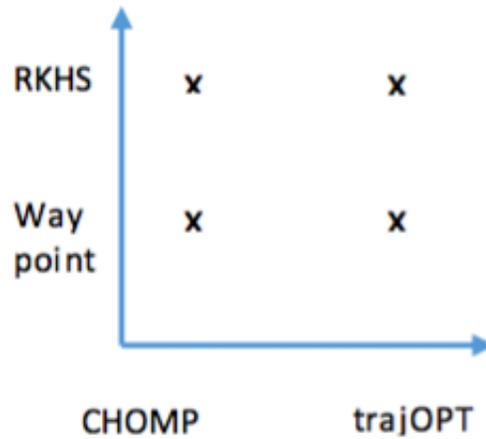
		1 st order	2 nd order
Soft constraint	geometry	x	x
Soft constraint	point	CHOMP	x
Hard constraint	geometry	x	trajOPT
Hard constraint	point	x	x

This above table is called a $2 \times 2 \times 2$ factorial design.

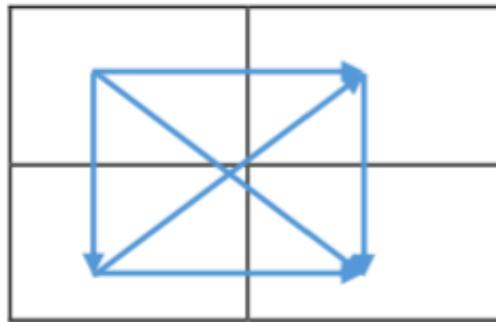
We need to discount all the interaction effects. If one level change gave good result and another gave worse result and the net improvement of outcome by changing both the variables was zero, we need to know there is interaction between the variables.

Next example is about evaluating CHOMP and trajOPT with two different representations of trajectories: Reproducing Kernel Hilbert Space (If interested read more about RKHS) and way points vector. First we had learnt that trajectory is represented by a vector of way points configuration. We could also represent trajectory as a linear combination of splines and play with the parameters of the spline to get optimal trajectory. Reproducing Kernel Hilbert Space (RKHS) is another way to represent trajectory. Let $f(x)$ be a function for the trajectory. It can be represented by the tuple of (α_i, x_i) . κ is a kernel function. x is time.

$$f(x) = \sum_{i=0}^N \alpha_i \kappa(x_i, x) \quad (9.1)$$



This graph shows the 2 x 2 factorial design. We talked about how it was important to test for both RKHS and Way points with CHOMP and trajOPT to be able to say that RKHS gives better results irrespective of the algorithm.



In order to understand the effects of the independent variables separately in the 2 x 2 x 2 factorial design, we can do pair-wise combinations of independent variable as shown in the previous figure.

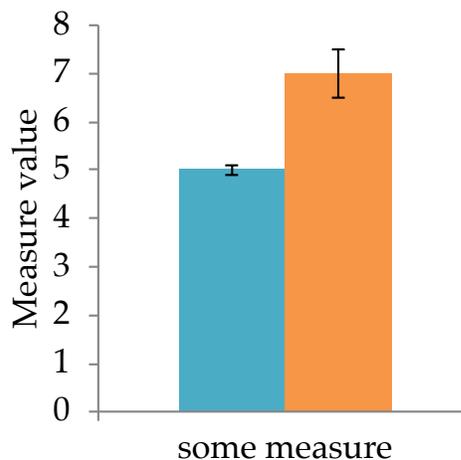
- It is possible to run pair-wise tests for up to two independent variables each with two levels. However, if that number of levels is higher, checking for cross product of all the levels becomes expensive and may be undesirable.
- If the effect on the outcomes are monotonic, they it is worthwhile only the consider the boundaries.
- It is also worthwhile checking if it is possible to reduce the number of levels or drop some of the conditions altogether
- We could also randomize for conditions

9.2 Statistical Analysis

9.2.1 Error Bars

Standard errors must be displayed on a bar graph. Standard error defines the standard deviation of the distribution of the means. It is important to include error bars in bar graphs because high error bar means that the sample mean deviates significantly from the population mean. Let us consider the case of 1 independent variable (algorithm) with 2 levels and a within subject study, i.e. all the motion planning problems are subjected to both CHOMP and CHOMP-2nd order.

Motion Planning	CHOMP	CHOMP-2 nd order
1	8	3
2	6	2
3	7	5



Suppose orange shows some outcome of CHOMP and blue is CHOMP-2nd order. This figure for Motion Planning task 1 shows that CHOMP is better than CHOMP-2nd. But CHOMP has larger standard error than CHOMP-2nd. The results are less reliable. In order to understand if 1st order CHOMP is definitely better than 2nd order CHOMP you should do a t-test.

9.2.2 t-Test

t-Test measures if means of two groups are statistically different. where t-value for 1 independent variable, 2 levels and within-subjects study is defined as below.

$$t = \frac{\bar{x}}{s/\sqrt{N}} \quad (9.2)$$

\bar{x} = sample mean, s = sample standard deviation, N = number of samples For one independent variable, 2 levels and between subject study the formula is as follows:

$$t = \frac{\bar{x}_1 - \bar{x}_2}{\frac{s_1}{\sqrt{N_1}} + \frac{s_2}{\sqrt{N_2}}} \quad (9.3)$$

9.2.3 p-value

p-value was defined in the class as probability of obtaining a value that is equal to or more extreme than what was actually observed when the Null Hypothesis is true. Two independent variables, two levels, between-subjects design gets complicated. We can do pair-wise difference between the means across all the independent variables, which gives us 6 comparisons. There is pitfall of using standard p-value of 0.05 in that case.

Standard p-value = 0.05

Let's say 100 such pairwise comparisons of means instead of 6. $P(\geq 1 \text{ error}) = 1 - P(\text{all correct}) = 1 - (0.95)^{100} = 99.4\%$

There are more sophisticated tools which are used to lower the threshold error or restrict the p-value below 0.05. These are called Bonferroni corrections, in which each individual hypothesis is tested at $0.05/\text{number of hypothesis}$.

9.2.4 Construct Regression

Sometimes statisticians do Construct Regression to check if there is already an interaction between the independent variables. Let a and b be two independent variables, each with two levels? [0,1].

A regression model is constructed between a and b as:

$$\beta_1 a + \beta_2 b + \beta_3 ab + \beta_4$$

We look at what sign each of the terms above. If β_3 is high, that means, there is interaction between variables a and b.

Note: Use Minitab, R or JMP to do statistical analysis.