

0.1 Introduction

Group sampling is an experimental design introduced to handle large parameter sets in a sensitivity analysis [?]. This design allows an analyst to identify influential parameters and determine their influence. It even allows obtaining sensitivity analysis information from so-called supersaturated designs. These are designs, where the number of measurements is smaller than the number of parameters.

The assumption is made, that the influence of each parameter is negligible. This assumption is rejected if there is data providing strong evidence of this influence. With this approach, the problem can be viewed as one of statistical testing [?]. This makes it possible to test for influential parameters and later reuse this information to analyze the nature of these influences. At the core of group sampling is the idea, that information about parameters can be extracted if multiple parameters are put in a group and tested together.

How the amount of testing can be reduced by clever designs, can be explained by recent efforts to reduce the number of tests needed to test as many people as possible for the SARS-CoV-2-Virus. [?] describes several approaches for pooled testing.

Here, a slightly altered version will be explained to more easily show, how the number of tests can be reduced by pooled testing. We assume we want to test 9 individuals for SARS-CoV-2 and want to reduce the number of tests needed. We also assume that the chance of more than one person being infected is negligibly small. By grouping our 9 individuals into groups and only using one test for the whole group, we can reduce the needed tests without significantly impacting the chance of detection of an infected individual.

This can be done by grouping these 9 individuals in a specific way. First, we create a matrix with the same amount of elements as the individuals to test and the dimensions $m \times n$. In this case, a 3x3-matrix. Now we assign individuals to a corresponding place in the matrix. This is done by assigning the element x_{ij} to the individual $I_{(mi+j)}$. The groups for testing are then created by taking each column and each row of the matrix as a group, resulting in six groups total. For ease of reference, the groups created from the rows are g_i where i is the index of the row and the groups created from the columns are h_j where j is the index of the column. This form of grouping gives us a way to re-identify an infected person if there is one. Let's assume I_5 is infected with SARS-CoV-2 in this example. Subsequently, the groups containing this individual would test as positive. In our example, if I_5 is infected, the groupings g_2 and h_5 would test as positive. To re-identify the individual, the overlap between those two groupings needs to be identified. This can be done by looking at the indices of the groupings $f(g_i, h_j) = m * i + j$.

With this, the amount of needed tests to test 9 individuals is reduced from 9 to 6 tests. This design, in theory, can be scaled up but may suffer in reliability when increased in size. In the case of the SARS-CoV-2 example, the tests used on the group might not be able to identify if a group is infected if the amount of individuals in this group is higher than a certain threshold.

The same idea of extracting as much information as possible out of groupings is used in group sampling. Here, the parameters are grouped randomly in groups of the same size. For each group, a test is done, and the influence values are stored. By repeating this grouping multiple times with different groups, information about the influence of a single parameter can be extracted.

0.2 Group sampling for sensitivity analysis

?] describes how group sampling can be applied to perform sensitivity analysis on a given model. Suppose we have a model with 1000 parameters X_i with $i \in [1 - 1000]$ where only a few parameters are influential. We group the parameters into groups of equal size. With 1000 parameters a sensible amount of groups could be $M=10$, giving us 10 groups containing 100 parameters. Group G_1 would be $G_1 = \{X_1, X_2, X_3, \dots, X_{100}\}$, $G_2 = \{X_{101}, X_{102}, X_{103}, \dots, X_{200}\}$ and Group $G_{10} = \{X_{901}, X_{902}, X_{903}, \dots, X_{1000}\}$. This kind of grouping would be repeated N times, with randomly assigned parameters for each group, resulting in $N * M$ groups $g_{n,m}$, where for each grouping $g_{n,M}$ all sets of parameters are distinct. Each group is now treated as a parameter of its own, assigning all parameters in a group the same value, giving us an abstraction of the model with only 10 parameters. Now, on these 10 groups, we can perform a set of simulations to determine the influence of the group as if it were a single parameter.

In Table 1 an example with 10 parameters, two groups $M=2$ and 5 rounds of random grouping can be found. All parameters with bold font are in the first group of the grouping M . For each group an influence value is determined and each parameter in the group is assigned this influence value for this grouping. If we look at the second grouping, the parameters X_1, X_4, X_6, X_7, X_9 are in this first group G_1 and have the influence value of 9.3. The parameters $X_2, X_3, X_5, X_8, X_{10}$ are in the second group G_2 and have the influence value of 5.1. In this example, the parameter X_4 is an influential parameter. We can already see, with the second grouping alone, the influential parameter is most likely contained in group G_1 for the second grouping. If we take the average value of each parameter for each grouping, we can estimate the influence of each parameter. In this example, we can see the parameter X_4 has the highest influence and is most likely our influential parameter.

Table 1: Parameter groupings and their influences

M	G_1	G_2	X_1	X_2	X_3	X_4	X_5	X_6	X_7	X_8	X_9	X_{10}
1	8.1	1.8	8.1	1.8	1.8	8.1	1.8	8.1	1.8	1.8	8.1	8.1
2	9.3	5.1	9.3	5.1	5.1	9.3	5.1	9.3	9.3	5.1	9.3	5.1
3	10	6.5	6.5	6.5	6.5	10	6.5	10	10	6.5	10	10
4	9.8	5.4	5.4	9.8	5.4	9.8	9.8	9.8	9.8	5.4	5.4	5.4
5	4.7	0.9	4.7	0.9	4.7	4.7	4.7	0.9	0.9	4.7	0.9	0.9
			6.8	4.8	4.7	8.4	5.6	7.6	6.4	4.7	6.7	5.9

From this example, we can also see how non-influential parameters can look influential if they share the group with the influential parameter too often. If we look at X_6 , the parameter looks influential, even though if the actual influential parameter is not in the same group, the group influence is small. This makes it harder to determine the actual influential parameter. ?] give us the probability of two parameters, X_i and X_j sharing a group t times with:

$$P_{ij}(t, N, S) = \binom{N}{t} \left(\frac{1}{S}\right)^t \left(\frac{S-1}{S}\right)^{N-t} \quad (1)$$

For our example with $N = 5$ groupings and a group size of $S = \frac{|X|}{M} = 5$, this gives us a probability of X_4 and X_6 sharing the same group $t=3$ times with $P_{4,6}(3, 5, 5) = 0.051$ and a total probability of any parameter sharing the group with the influential parameter X_4 t times with $1 - (1 - P_{ij}(3, 5, 5))^9 = 0.377$.