

実験計画法 Design of Experiments (DoE)

明治大学 理工学部 応用化学科
データ化学工学研究室 金子 弘昌

実験計画法とは？

- ✓ 効率的に実験もしくはシミュレーションをして、目的を達成するための方法
- ✓ 実験パラメータのすべての組み合わせの中から、いくつかの組み合わせを、情報量がなるべく大きくなるように選択する
- ✓ 選択された実験パラメータの組み合わせで実験したあとに、仮想的な実験を行うためのモデルを構築する
- ✓ モデルに基づいて、次の実験 (シミュレーション) を計画する
- ✓ 実験計画と実験 (シミュレーション) とを繰り返すことを、適応的実験計画法という

実験計画法のイメージ 1/3

- ✓ 化学反応 $A + B \rightarrow C + D$ の、 C の収率を上げたい！
- ✓ 反応温度 25°C で実験すると収率はばらつくため、複数回実験する必要がある
 - 3 回が多い
- ✓ 反応温度 $20^{\circ}\text{C} \cdot 30^{\circ}\text{C} \cdot 35^{\circ}\text{C} \cdot 40^{\circ}\text{C} \cdot 45^{\circ}\text{C} \cdot 50^{\circ}\text{C}$ でもやってみたい！
収率を最大にする反応温度があるのでは！？
 - 3×7 回
- ✓ 反応時間・触媒の量もいろいろ変えて、収率が上がるか確認したい！
 - $3 \times 7^3 = 1029$ 回！？

実験計画法のイメージ 2/3

✓いったん落ち着いて考えよう

✓収率を最大化する前に、モデル f

収率 = f (反応温度、反応時間、触媒の量)

を求めることを目標にしたらどうか！？

✓このモデルがあれば、反応温度、反応時間、触媒の量を変えたときに収率がどうなるか、実験せずに確認(シミュレーション)できる！

✓シミュレーションの結果、収率が最も大きい条件で実験しよう！

実験計画法のイメージ 3/3

✓たとえば線形モデルであれば、

$$\text{収率} = a_1 \times \text{反応温度} + a_2 \times \text{反応時間} + a_3 \times \text{触媒の量} + \text{定数項}$$

と仮定

✓ a_1 、 a_2 、 a_3 を求める (モデリング) ための実験を行う

- 最低 3 回！
- 実際はもっと実験回数が多かったり、線形モデルではなくモデルに非線形項 (交差項・二乗項) を入れるなど、非線形モデルにしたりする

✓ 適応的実験計画法

- モデリングと実験とを繰り返す

実験パラメータの候補をどのように選択するか？

5

- ✓ 実験パラメータである説明変数 X と収率などの目的変数 y との間で、良好な回帰モデルが構築できるように選択する
- ✓ 回帰モデルとは？・・・最小二乗法による重回帰分析を仮定する
- ✓ 回帰モデルは、 $(X^T X)^{-1} X^T y$ で与えられる
最小二乗法による重回帰分析の詳細についてはこちら
<https://atachemeng.com/ordinaryleastquares/>
- ✓ y のデータがない (まだ実験していない) 状況で、どのように “良好な” モデルであると判断するか？
- ✓ $(X^T X)^{-1}$ に着目

D 最適基準が大きくなるように選択する

✓ $(\mathbf{X}^T\mathbf{X})^{-1}$ に着目

✓ この逆行列が計算できないと、そもそも回帰モデルが得られない

✓ 逆行列が計算できない状況に近いと、回帰モデルは“良好でない”
と考えた

- 逆行列が計算できない状況・・・ $\mathbf{X}^T\mathbf{X}$ の行列式 ($\det(\mathbf{X}^T\mathbf{X})$) が小さい状況

✓ $\mathbf{X}^T\mathbf{X}$ の行列式を D 最適基準とよび、 $\mathbf{X}^T\mathbf{X}$ の行列式が大きくなるように実験パラメータの候補を選択する！

- これなら X のデータのみから選択できる♪

どうやって実際に選択するか？

- ✓ 実験計画を立てるための MATLAB や Python のプログラムを作りました！

https://github.com/hkaneko1985/design_of_experiments

プログラムの使い方

- ✓ variable1, variable2, ... : 実験パラメータの種類 (反応温度、反応時間、触媒 など)
 - variable3, variable4, ... と増やすことが可能
 - コードの最初の例にあるように、variable1, variable2, ... のそれぞれに、候補の値を入れる (0.2, 0.6, 0.8, 1, 1.2 など)

- ✓ number_of_experiments : 選択する実験パラメータの候補の数

- ✓ 実行内容
 - variable1, variable2, ... のすべての組み合わせを出して、all_experiments.csv に保存
 - その中からD最適基準が大きくなるように number_of_experiments だけ選択して、selected_experiments.csv に保存

実験もしくはシミュレーションの後には？

- ✓ 選択された実験パラメータの候補での実験もしくはシミュレーションが終わったとする
- ✓ その後は、ベイズ最適化で次の実験パラメータの候補を選択する
 - 適応的実験計画法
 - ベイズ最適化の詳細についてはこちら
<https://datachemeng.com/bayesianoptimization/>

[補足] D 最適基準以外の最適基準の例

- ✓ A 最適基準 : $\mathbf{X}^T\mathbf{X}$ の逆行列の対角成分の和を最小化
- ✓ E 最適基準 : $\mathbf{X}^T\mathbf{X}$ の固有値の最小値を最大化
- ✓ I 最適基準 : $\mathbf{X}^T(\mathbf{X}^T\mathbf{X}/m)^{-1}\mathbf{X}$ の対角成分の平均値を最小化
 - m : サンプル数
- ✓ minimax 基準 : $\mathbf{X}^T\mathbf{X}$ の対角成分の最大値を最小化