お茶の水女子大学 共創工学部 文化情報工学科　　吉田 裕亮 研究室

# 数学の様々な分野の理論を応用して統計的データ解析の新たな手法の開発を目指します

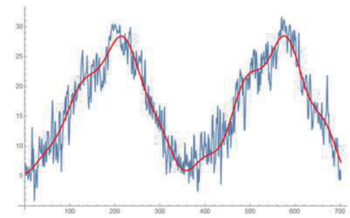
## 研究室の目標

基礎解析学　応用数理　データ解析

### 関数解析学と確率論を合わせてデータ解析に応用

観測や実験で得られたデータには，真に必要な有益な情報と共に，その必要とする情報を隠してしまう不要なノイズが混入しています．純粋数学の多くの基礎理論は，このような観測されたデータから真に有益な情報を抽出するための統計的データ解析に応用することが可能です．当研究室では，数学の関数解析学の分野と確率論を融合させてデータ解析に応用する，理論と応用の両面で研究を行っています．

$$\min_f \left( \sum_{i=1}^{N} \{y_i - f(x_i)\}^2 + \lambda \int \{f''(t)\,dt\}^2 \right)$$



## 研究事例

ランダム行列　時系列解析

### ランダム行列を用いた時系列解析

与えられた長さ $L$ の時系列データ $\{x_i\}_{i=1}^{L}$ に時系列モデルとして MA (移動平均) モデルや AR (自己回帰) モデル，あるいは ARMA モデルで近似する際に，そのモデル適合度 (goodness) の統計的検定を行う新たな手法をランダム行列の理論を応用して考案しました．

時系列データを行毎に折りたたんで，適当なサイズのデータ行列を構成し，それを行方向に相関のあるランダム行列とみて，ランダム行列のゆらぎの理論を用います．

(1) 与えられたデータ $\{x_i\}$ から $N \times M$ の行列 $X = \left(\xi_{i,j}\right)_{N\times M}$ を

$$\xi_{i,j} = x_{(i-1)M+j} \text{ ただし } N \le M \text{ かつ } NM \le L.$$

と構成する．

(2) $X$ から標本共分散行列 $\frac{1}{N}X^tX$ を作成し，行列の 2 次モーメント

$$\mu_2 = \frac{1}{N}\mathrm{Tr}((\frac{1}{N}X^tX)^2)$$

を計算する．

ランダム行列の理論を用いると，

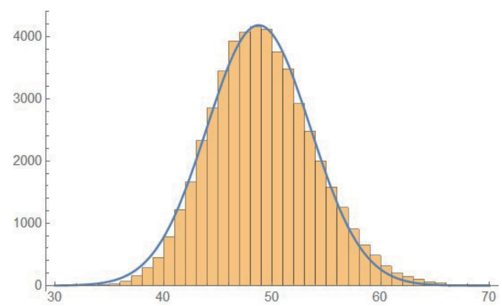
$\mathrm{H}_0$ : 与えられた時系列データがあるモデル $\mathcal{M}$ で生成されている

という帰無仮説の下で，上で考えた $\mu_2$ は，モデル $\mathcal{M}$ のパラメータのみで定まる平均 $\alpha_2$，分散 $v_2/N^2$ の正規分布 $\mathcal{N}(\alpha_2, v_2/N^2)$ に従って観測されることが証明されます．したがって

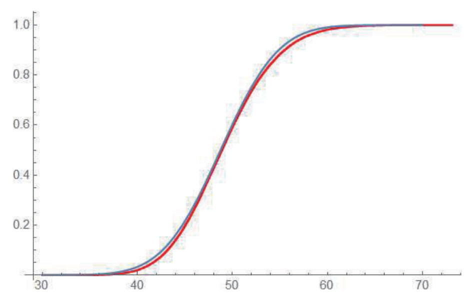
$$z = \frac{\mu_2 - \alpha_2}{\sqrt{v_2/N^2}} \sim \mathcal{N}(0,1).$$

によって，両側 $Z$-検定を行うことが可能です．

理論的には行列サイズが無限大になったとき，すなわち漸近的に正規分布に従うのですが，下図の　N = 40, M = 60での，シミュレーションの例より，この程度のサイズでも十分に正規分布が使えるのが見て取れる



青線が漸近的な正規カーブと30000回のヒストグラム



青線：漸近的な累積分布関数，赤線：30000個での経験累積分布関数

本手法は，比較的少ない計算量で，モデルのあてはまりの検定が可能です．なお本内容は，以下で公表されています．

Statistics & Probability Letters Volume **127**, August 2017, Pages 85-96.　https://doi.org/10.1016/j.spl.2017.03.022