# データ量子モデルとその応用

東京大学大学院　工学系研究科　　池上 孝則

## 1　はじめに

最尤推定法では、データ$X$から推定した確率密度関数$f(x|\theta)$が存在するとき、$x$がデータで$\theta$を確率変数と見る関数$g(\theta|x)$を尤度関数 $L(\theta|x)$と位置づけており，データが複数存在する場合にはそれらを積演算で統合した関数が尤度関数となる。しかし $g(\theta|x)$は母数$\theta$を関数の要素に含むことから独立とは言えず，故に乗法定理は適用できないはずである。また，従来の尤度関数は直感に照らして齟齬があるし，何より積演算で実現する直列型データ処理はシステム構成的に不安定である。

そこで本報告では，和演算で構成する並列型データ処理に基づく尤度関数の実現を目的として，統計学に量子力学的知見を組み込んだ「データ量子モデル」を提案する。

## 2　データ量子モデルの概要

データ量子モデルでは，データ目線の確率密度関数$g(\theta|x)$を「母集団分布を構成する波動関数」と位置付けて$\varphi(x)$と表し，それを「データ量子」と呼ぶ。またデータ処理上の要請から，そのプラットフォームとしてヒルベルト空間で構成される「超世界」を設定し，複素数を用いて以下のように表す。

$$\varphi(x_j) = \frac{\mathrm{i}}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(\mu-x_j)^2}{2\sigma^2}} = \frac{1}{\sqrt{2\pi\sigma^2}}\mathrm{e}^{-\frac{(\mu-x_j)^2}{2\sigma^2}+\frac{\pi}{2}\mathrm{i}}$$

超世界でのデータ統合はデータ量子の重ね合わせと解釈されるので，全てのデータの統合後の状態を示すデータ分布関数$\Psi$はデータ量子の和演算で定義され，$\Psi(X) = \sum_{j=1}^{n}\varphi(x_j)$ となる。また，$\Psi(X)$の実世界での姿は，ボルンの確率解釈からそのノルムをとることにより，確率密度を表すデータ密度関数$\boldsymbol{\zeta}(X)$として以下の式で観測することができる。

$$\boldsymbol{\zeta}(X) = |\Psi|^2 = \Psi^*\Psi \propto \sum_{j=1}^{n}\sum_{k=1}^{n}\mathrm{e}^{-\frac{(\mu-x_j)^2+(\mu-x_k)^2}{2\sigma^2}}$$

## 3　データ量子モデルの有効性の検証

データ量子モデルの有効性の検証は，まず正規分布や一様分布などの理論分布，三角分布や指数分布などの幾何分布，及びこれらの分布に外れ値を付加した分布を母集団分布に設定してその要件を満たす乱数を発生し、それに基づいて正規分布関数 $\phi$ 及びデータ密度関数 $\zeta$ を算出した上で規格化し，母集団分布と推定分布関数の重複率で評価した。その結果，母集団分布が正規分布の場合には重複度に有意な差は存在しないが，それが正規分布から外れるにつれ，また非対称性が大きくなるにつれてデータ密度関数 $\zeta$ の優位が顕著となり，外れ値を含む母集団分布に対してはその推定におけるパフォーマンスの差は決定的となる（図1，2）。

データ密度関数 $\zeta$ は母集団分布に対する情報欠損が少ない母集団分布の推定関数であるため，記述統計及び数理統計のいずれの分野でも様々な応用が期待できる。

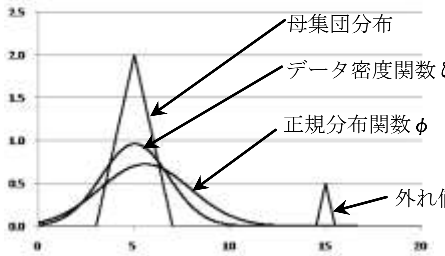

図１　母集団分布の推定

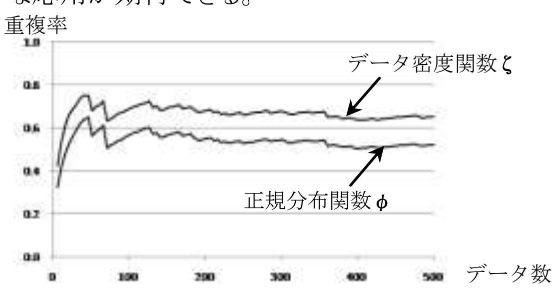

図２　母集団分布と推定分布の重複率の変化