第4回　理学部門談話会

日時：2011年7月20日（水）
　　13：30-15：00
場所：理学部第１会議室（理学部２号館６F）

話題及び提供者

「統計学の生み出したスーパースター，EMアルゴリズム」
（野間口　謙太郎）

「くりこみについて」
（仲野　英司）

「動物の心が知りたい」
（種田　耕二）

教職員，大学院生，学生，一般の方々のご来場をお待ちしています．

（問い合わせ：suzuki@kochi-u.ac.jp）

統計学の生み出したスーパースター，EMアルゴリズム

数学コース　野間口謙太郎

[**最尤法**]　数理統計学では，**母数**$\theta \in \Theta$をもった**確率モデル**$f(x;\theta)$（xの確率密度関数）に従って**データ**xが出現するという前提をおきます．その下で，母数についての**推測（推定・仮説検定）**を行います．その際，統計家が絶大の信頼を寄せている手法が**尤度法**というものです．

尤度法の原理は簡単です．θの推定を考えるなら，データxに対して，

$$\hat{\theta} = \mathrm{argmax}_{\theta\in\Theta} f(x;\theta)$$

を見つけることです（単純な最大化問題）．これを**最尤推定量**（maximal likelihood estimator, **MLE**）と呼びます．ちなみに，$f(x;\theta)$は母数θの関数と見なされたとき，**尤度関数**と呼ばれます．

[**完全データと不完全データ**] 通常のモデルの下での通常のデータであれば，上の最大化問題は簡単に解けるのが通常です．このときのデータを完全データと呼びましょう．しかし，通常と異なり，ある部分が欠けたデータyについて上の最大化問題を解かなければならない場合もあります．このyを不完全データと呼びましょう．

[**不完全データに関する最尤法**] 完全データxが部分的に失われているので，失われた部分を$z = x_{miss}$とおくと

$$x = (y, z) = (x_{obs}, x_{miss})$$

と書けます．そこで，欠損が起きるとyになるようなxの集まりを

$$\chi(y) = \{x : x_{obs} = y\}$$

と表記すると，yの従う密度関数は次のように表現できます．

$$g(y;\theta) = \int_{x\in\chi(y)} f(x;\theta)dx$$

たいていの場合，この最大化問題はもつれた家庭環境を修復する程度に面倒です．

[**尤度関数の分解**] 条件付き密度関数の定義を思い出すと，

$$f(x|y;\theta) = \frac{f(x;\theta)}{g(y;\theta)}, \quad x \in \chi(y)$$

両辺の対数を取ると，

$$\log g(y;\theta) = \log f(x;\theta) - \log f(x|y;\theta)$$

このとき，右辺の第1項は完全データの（対数）尤度関数なので，その最大化は非常に容易です．完全データの最尤推定量の計算を邪魔しているのが，右辺の第2項

です．きっと何処かにいる愛人のように接触を拒みます．昔と違い x の何かが欠けてしまったので（x は観測されない），昔のように親しく直接は付き合えません．そこで，両辺を条件付き分布 $f(x|y;\eta)$ で期待値を取ります．すると，見えなかった欠測部分 z に関係する項がその期待値で置き換えられ，y で表現できます（y は観測されている）．これで互いに調停の場に立てるようになり交渉できるようになります．

$$\begin{aligned}\log g(y;\theta) &= \int_{x\in\chi(y)} \log f(x;\theta)\ f(x|y;\eta)dx - \int_{x\in\chi(y)} \log f(x|y;\theta)\ f_\eta(x|y;\eta)dx \\ &= Q(\theta|\eta) - H(\theta|\eta)\end{aligned}$$

このとき，第 1 項 $Q(\theta|\eta)$ には完全データの尤度関数のもっていた優しさ（易しさ）が残っていてほしいと，強い期待を抱かせます．ですから，忌まわしい存在 $H(\theta|\eta)$ を完全に無視して，$Q(\theta|\eta)$ だけで話し合いを続けたいな，と考えるわけです．

[**EM アルゴリズム**] 初期値 $\theta_0 \in \Theta$ から出発して，$n = 0, 1, 2, ...$ に対して次の 2 つのステップを収束するまで繰り返す．θ_n を **EM 列**と呼びます．

$$\begin{array}{lcl}\text{E(xpectation)-step} & : & Q(\theta|\theta_n) \text{ を計算せよ} \\ \text{M(aximization)-step} & : & Q(\theta|\theta_n) \text{ を最大にする } \theta = \theta_{n+1} \text{ を見つけよ}\end{array}$$

[**EM 列はどこへ収束するか？**] EM アルゴリズムで改良できない母数の集まり（**EM 最適解**）が次のように定義できます．

$$S = \left\{\theta \in \Theta : \theta = \text{argmax}_{\eta\in\Theta} Q(\eta|\theta)\right\}$$

S の中の要素は，尤度関数の極値や変曲点，鞍点などで構成されます．非常に緩やかな条件の下で，EM 列は収束し，その収束先は S の中の要素になります．つまり，収束先は，尤度関数の極値や変曲点，鞍点です．

[**なぜ EM アルゴリズムは支持されたのか？**] 複雑な確率モデルであっても，ある仮想的な変数 z を設定して，観測データ y（いわゆる欠測のないデータであっても）に z を追加すると，(y, z) が簡単なモデルからのデータであると見なせる場合が多いからです．z は絶対に存在しないし観測できないデータなのに，それを想定することにおもしろみを感じ，実用性を見いだしたからでしょう．

[**談話会では**] 主に EM 列の収束に話を絞ってお話しします．

Ｈ２３年７月理学部談話会

題名「くりこみについて」

講演者　仲野英司

要旨

我々自身も含め、身の周りのものほとんどは、小さく見ていけば素粒子から構成されています。電子も素粒子の一つです。「素粒子で出来ている」となぜ言えるのかというと、それらが見えるからです。物理でいう「見える」とは、目で見るというだけでなく、素粒子間の相互作用というものを通して反応があるということです（ここで反応とは、例えば、木に石をぶつけて跳ね返るという状況と同じ意味です)。これらの反応の過程を詳細に記述するための道具が場の量子論になります。場の量子論は、反応の結果、どのように物理量が観測されるかを教えてくれます。

1920 年ごろに量子力学が誕生し、電子や原子などのミクロな世界を記述できるようになりました。より精密な記述のために、その後すぐ量子力学と相対性理論を合わせて作った場の量子論が考えられました。しかし、当初、場の量子論では意味のある計算結果を引き出すことが出来ませんでした。“くりこみ”（繰り込み）は、1950 年ごろにこの場の量子論の病気を治すために考え出された処方箋でした。これによって大分改善されましたが、完全に治ったかどうかは今でも不明です。

現代では、くりこみの見方が少し違ってきています。むしろ、積極的に物理学や数学で利用されるようになってきました。例えば、水などの相転移点の解析や、大きなスケールの運動の記述など、さまざまです。

本講演では、なるべく、くりこみについて直感的に理解できるように解説したいと思います。最後に、くりこみに関連した自分の研究を少し紹介します。