統計科学 文献案内

伊庭幸人

統計数理研究所
〒106 8569 東京都 港区 南麻布 4-6-7
Email iba@ism.ac.jp, URL http://www.ism.ac.jp/~iba/

(1998年10月5日受理)

統計学とその周辺について筆者が書いてきた解説とその他の興味ある文献を、簡単な解説とともに紹介した。総合研究大学院大学共同研究 "非線形現象の数理科学"(1997) の報告集に掲載されたものである。その後、長らく筆者のホームページにあり、私家版の「集成版」にも収録したが、より多くの人に読んでもらうために、「物性研究」に投稿することとした。投稿にあたっては、文1つをを改め (2.3節)、URLの一部を更新した他は、活字を大きくし改行を手直しする程度の変更にとどめた。AICの説明などはやや不完全とも思えるが、変更していない。

目次

統計科学 文献案内

統計数理研究所 伊庭幸人

1　はじめに

本稿では、いままで筆者が書いてきたいくつかの解説を紹介するとともに、最近、面白い(面白そうだ)と思った教科書・モノグラフ・論文について述べた。おもな目的は、ある程度理論科学一般の知識のある読者に統計学とその周辺の発展について知っていただくことである。読者として、物理学の研究者を念頭においている部分もあるが、それ以外の読者にも役立つと思う。

筆者は統計学者として教育を受けたわけではなく、自己流に気にいった部分だけを勉強してきた。また、本稿では意図的に統計学からみたとき境界領域に属する文献を入れた。したがって、"標準的な統計学" からみた場合、大変偏った内容になっていると思う。また、新着の教科書やモノグラフについては、(筆者は学生を指導する立場にないが) もし学生とゼミをやるなら選んでも良いと思うような本については、中身をあまり読んでいないものも入れた。したがって、実際に読んでみるとつまらなかったり、間違っていたりするものもあるかもしれない。これらの点については読者の寛容をお願いしたい。

筆者のホームページは http://www.ism.ac.jp/~iba である。筆者の書いた解説は大部分ここでみられる (但し一部は最終バージョンでない)。ホームページの下にある
http://www.ism.ac.jp/~iba/bunken.html には統計学・統計物理など筆者の関心を持っている分野の文献リストがある。

2　統計学とその周辺

2.1　統計学一般

筆者が統計学の教科書として普段そばに置いているのは、

[竹内 1963]
竹内啓 (1963), 数理統計学, — データ解析の方法 —, 東洋経済新報.

である。大変簡潔で、証明や定義のスタイル (厳密さの程度) も筆者に合っている。30年以上前からある本だという感じは全然しない。個別の話題については、

[竹内ほか 1989]
竹内啓ほか編 (1989), 統計学辞典, 東洋経済新報.

が役に立つことがある。典型的な大項目主義の辞典で、大勢の人が書いている。極端な話、2箇所に別々の著者が同じ主題を解説していて、両方を読んでも話がかみあわない例さえある。しかし、ごたごたしているのがかえって面白い。具体的トピックに関するエッセイ風の文章が好きだという人には、

[竹内 1980]
竹内啓 (1980), 現象と行動のなかの統計数理, 新曜社.

をすすめる。筆者は内容のうち数編を読んだにすぎないが、面白い。

2.2 赤池情報量規準 (AIC)、最小記述長原理 (MDL)

データに統計モデルを最尤法などで当てはめると、パラメータを追加すればするほど当てはまりがよくなるのが普通である。しかし、100個のデータにたいする100変数の多項式による当てはめとか、あらゆる項目で分類した表とかは普通使わない。前者は単に点をつないだのと変わらなくなってしまうし、後者はひとつの区画 (セル) にデータが1個か0個のどちらかになってしまうので、データ処理としての意味がなくなってしまう。これは当たり前のことと思われているが、その根拠を改めてたずねたり、変分原理の形に定式化を試みたりしている人達がいる。“いまあるデータ” に対する当てはまりの良さが最適化すべき対象ではないとしたら、何を最適化すれば良いのだろうか?

ひとつの考え方は、モデルの “いまあるデータに対する当てはまり具合” ではなく “予測能力” を最適化する、ということである。そんなのわかるわけない、と思うのが普通かもしれない。ある意味ではその通りだが、予測能力を論じることが全く無意味なわけではない。3次式を当ては

めたとして、仮にこの3次式に雑音を加えたものが本当のデータ源だと思うことにする。この仮定のもとで、“いまあるデータ”と同じ個数のデータを発生させて、3次式を当てはめる場合と2次式を当てはめる場合を比較しよう。(特定の3次式+雑音)がデータ源だと仮定したので、予測能力をそれに対する当てはまり具合として評価することができる。このとき、仮定したデータ源の3次式が2次式に近くて、“いまあるデータ”の数が少なければ、2次式に基づくモデルが3次式に基づくモデルに勝つことがありうる。利用できるデータの数が有限であり、そのための揺らぎの影響を、図体の大きい3次式の方が余計に受けるというのが本質である。

3次式が本当だと仮定しても、なお、2次式に基づくモデルの方が良い場合があるという点が重要である。その場合に、逆に2次式を仮定したとすれば、よりいっそう2次式の方が有利になることが期待される。したがって、考えているモデル族{2次式+雑音、3次式+雑音}の内部の論理から、2次式の方が良いという判断が可能である。このような観察を定量的に数式化することを試みたのが赤池情報量規準(AIC)である。当てはまりの良さをKullback-Leibler情報量ではかるので、“情報量規準”という名前がついている。

新しい考えを出したときに、教科書から書き直して、自分たちの体系を世の中に広めるというのは、日本では珍しい考え方かもしれない。それをやったのが、

[坂元ほか 1983]
坂元慶行, 石黒真木夫, 北川源四郎 (1983), 情報量統計学, 共立出版.

(英語版)
Sakamoto,Y., Ishiguro,Y. and Kitagawa,M. (1986), Akaike Information Criterion Statistics, Reidel.

である。赤池による解説は、たとえば、

[赤池 1976]
赤池弘次 (1976), 情報量規準AICとは何か, 数理科学 No.153, 1976年3月号 (特集・情報量規準), pp.5-11.

[赤池 1981]
赤池弘次 (1981), モデルによってデータを測る, 数理科学 No.213, 1981年3月号 (特集・統計モデル), pp.7-10.

にある。これらの号には多くの応用例が含まれている。また、より専

門的・客観的な記述を求める場合には、[竹内ほか 1989] の情報量規準の項目 (この分野の理論の専門家が書いている) が役に立つ。[竹内 1980] のうちの1編も情報量規準を扱っている。

はじめの疑問に対する、AIC とは別の答えとして "情報圧縮" の概念に基づくものがある。事象 i をコード化する (記号列として表現する) 場合には、頻繁な事象ほど短い符号を対応させた方が平均としては短くてすむ。事象 i が起こる確率 p_i が既知の場合、各事象に $-\log p_i$ に比例した長さのコードを対応させた場合が最適であり、この際のコードの長さの期待値は $-\sum_i p_i \log p_i$ となることが知られている (シャノンの情報源符号化定理、少し不正確な表現)。

情報理論の入門講義だと話はここでおしまいであるが、ちょっと考えてみると、p_i をあらかじめ知っている場合というのはそんなにありはしない。文書を圧縮するにしても、文字の頻度とか、ある文字のあとに別の文字がくる頻度とかは文書によって違うだろう。そこで、p_i を文書ごとにデータから推定するという考えが出てくる。このとき、何文字の依存性までを見たらいいか (ほぼ同じことだが、事象 i としてどういうものを選んだいいか) が問題である。初心者は、複雑なモデルを使えば使うほど、与えられた文字列を良く表現するのだから、計算の手間さえ考えなければ、複雑なモデルの方が良いと思うかもしれない。しかし、これは重要な点を見落としている。文書ごとに p_i の推定値が違う、したがって圧縮のやり方も違うのであれば、それを書いた対応表 (辞書) を添付しなければ、もとの文書を復元 (解凍) できない。したがって、この場合、最適化すべきなのは、推定された p_i を使って符号化した結果ではなくて、それに辞書の長さを加えたものなのである。こうして、最適化すべき量があらたに得られた。

上の議論はこのままでは大雑把 (前提が不明確) であるが、こういった思想をより押し進めて、世の中を情報圧縮の観点から理解しよう、というのが "Minimum Description Length Principle (MDL 原理)" という考えである。MDL については、たとえば、

[Rissanen 1983]
Rissanen (1983), A universal prior for integers and estimation by minimum description length, The Annals of Statistics, Vol.11, No.2, pp.416-431, 1983.

をみられたい。一応、決定版らしい本

[Rissanen 1989]
Rissanen,J. (1989), Stochastic Complexity in Statistical Inquiry, World Scientific.

があって、ゼミをやったが、筆者は途中で逃げてしまった。はじめの部分は算術符号の説明などがあって難解だったが、後の方ではそれはあんまり使わないらしい。情報圧縮についての読み物風の解説としては、

[韓 1987]
韓太舜(はん・てすん) (1987), 数理科学 No.290, 1987年8月号(特集・情報圧縮),pp.5-15.

が面白い。

情報圧縮(MDL)と予測(AIC)とではずいぶん違うようであるが、AICの場合も予測の良さはKL情報量ではかるので、実はかなり近い関係にある。それで、予測を問題にした場合も本当はMDLの方が良いのではないかという人もいて、いろいろと議論(なぜか感情的になりやすいテーマなのだ)が行なわれているが、詳細は述べない。

事情通の人のために、2,3 述べておくと、まず、MDLは"真の構造が考えているモデル族の中にあってそれを見つけ出す"といった状況を念頭においているが、AICは"無限に複雑な構造からの有限なデータを暫定的にモデルで表現する"といった状況が念頭にある。また、MDLは辞書の部分を"横軸"(パラメータの次元を持つ軸)でみているので、これを縦軸(無次元の情報量の軸)と比較するには、次元を持った量(考えている世界の大きさに相当する)が必要である。この点を追いかけてゆくと、ベイズ統計における事前分布の扱いという、大変長い歴史を持った問題にゆきつく。実際、AIC対MDLという問題は、Fisher流の検定とベイズ流(Jeffreys流)の検定の違い(しばしばLindley paradoxといわれる)の再現である。

いくつかの論文で赤池は非常な苦労をしてMDLがなぜ不自然と考えられるかを説明しようとしている。赤池の説明には確かにもっともな点があるが、一方で単純な数学的設定のもとではMDLの方が優れているようにみえる。こういう場合にとるべき方法としては、(1) AICの方がMDLより優れているような状況の本質を取り出して数学的に定式化する、(2) 両者の矛盾を止揚するような体系を作り出す、などが考えられる。赤池は(1)も試みているが、後にはむしろ(2)の方向に向かったようである。その結果到達したのが、次節で述べる階層ベイズ的な枠組(ABIC法)である。(1)の方向については、[竹内ほか 1989]の情報量規準の項目が参考になる。

2.3 階層ベイズモデル

曲線を表現するのに、多項式などの特定の関数形を当てはめる代りに、“変動の少なさ”、“滑らかさ” などを拘束条件として推定をする方法(罰金つき最尤法)がある。ベイズ統計の枠組ではこれは次のように解釈される。

たとえば、データ $y=\{y_j\}$(j 番目の入力 $t(j)$ に対する出力が y_j, $j=1,2...K$) が与えられているとして、これから、関数 $y=f(t)$ を構成することを考える。関数の表現として、多数の点を結んだ折れ線を考え、各頂点をパラメータ $x=\{x_t\}$ $(t=1,2...N)$ とする(簡単のために x の添字(自然数) と t を同一視する)。“関数が滑らか” という予備知識を x の確率分布 (“事前分布”) として表現すると、

$$\pi(x)=\frac{1}{Z_\pi}\exp(-\lambda\textstyle\sum_t(x_{t+1}-2x_t+x_{t-1})^2) \tag{1}$$

のようになる(λ は滑らかさの要求の度合を表わす定数、Z_π は分布の規格化定数)。

x を与えたときのデータ y の確率分布 $P(y|x)$ が

$$P(y|x)=(\frac{1}{\sqrt{2\pi\sigma^2}})^K\exp(-\frac{1}{2\sigma^2}\textstyle\sum_j(x_{t(j)}-y_j)^2) \tag{2}$$

で表現されるとする(分散 σ^2 のガウス雑音を仮定)。すると、データ y を得たあとの x の分布 (“事後分布”) は

$$P_{pos}(x)=\frac{P(y|x)\pi(x)}{\sum_x P(y|x)\pi(x)} \tag{3}$$

すなわち、

$$P_{pos}(x)=\frac{\exp(-E_{pos}(x))}{\sum_x\exp(-E_{pos}(x))} \tag{4}$$

$$E_{pos}(x)=\frac{1}{2\sigma^2}\sum_j(x_{t(j)}-y_j)^2+\lambda\sum_t(x_{t+1}-2x_t+x_{t-1})^2 \tag{5}$$

となる。しばしば、(3) 式を “ベイズの定理” と呼ぶが、内容は条件つき確率の定義のようなものであり、高校の数学でも理解できる式である。事後確率 (4) を最大にする x を選ぶことは、“事後エネルギー”(5) を最小にする x を選ぶことに等しい。データにバネがついていて、x の各成分がひっぱられているが、x の成分どうしも相互作用している (滑らかでない

曲線に対しては“罰金”がある)ので、両者の兼ね合いで形がきまる、というイメージである。

ここでパラメータ x も確率分布していると考えているが、これは普通の統計学の考えとはちょっと違っている。このような枠組(立場)を“ベイズ統計”という。いまの問題では、(5)の最小化そのものは必ずしもベイズ統計の枠組でなくとも論じられるが、ベイズの枠組では、“超パラメータ”λ, σ を推定する手がかりが得られるという利点がある(階層ベイズ法)。階層ベイズ法、あるいは階層的なベイズモデルは、隠れユニット・潜在変数・欠測データなどを統合的に扱う上で便利なので、現在各方面で話題になっている。赤池をはじめとする統計数理研究所のグループはこの枠組の開発と応用に大きな寄与をした。

筆者が統計数理研究所に入所した時点では、階層ベイズ法を利用したデータ解析法が一応確立されて、多くの応用例が発表されたあとであった。筆者の興味はこの枠組の統計物理との類似とそれを利用した計算アルゴリズムの開発にあったが、枠組そのものやその物理との類似などについてうまく解説した文献があまりなかったので、いきおい自分で解説を書くことになった。階層ベイズ法についての短い解説(EM法、認知科学などとの関連や批判的な議論を含む)は、

[伊庭 1996a]
伊庭幸人(1996), 学習と階層, — ベイズ統計の立場から —,
「物性研究」65-5 (1996年2月号), pp.657-677.

を、統計物理との関連に重点を置いた解説は

[伊庭 1993]
伊庭幸人(1993), ベイズ統計と統計物理, — 有限温度での情報処理 —, 「物性研究」60-6 (1993年9月号), pp.677-699.
伊庭幸人(1996), ベイズ統計と統計物理(物性研究1993年9月号)への訂正と追加, 「物性研究」65-5 (1996年2月号), pp.678-685.

を見られたい。また、少し古いが、長い解説

[伊庭 1990]
伊庭幸人(1990), 統計物理と統計的情報処理(1.00版), — 大規模・非ガウスモデルをめぐる話題 —,
(図を除いて `http://www.ism.ac.jp/~iba/` で入手可能).

が筆者のホームページにある。

階層ベイズ法に関する赤池の議論は、

[Akaike 1980]
Akaike,H.(1980) Likelihood and Bayes procedure, in Bayesian Statistics, Eds. Bernardo,J.M, DeGroot,M.H., Lindley,D.V., and Smith,A.F.M., University press, Valencia.

にある。長くはげしい討論つき。統計数理研究所の研究者による本としては、たとえば、

[Kitagawa and Gersch 1996]
Kitagawa,G and Gersch,W.(1996), Smoothness Priors Analysis of Time Series, Lecture Notes in Statistics, No.116, Springer.

[坂元 1985]
坂元慶行 (1985), カテゴリカルデータのモデル分析, 共立出版.

(英語版)
Sakamoto,Y.(1991), Categorical Data Analysis by AIC, Kluwer.

がある。前者では時系列の階層モデルが、後者の後半では密度推定や空間統計の問題が扱われている。

階層ベイズ法に関連した主張は古くからあるが、大規模なモデルについて実際に計算ができるようになったのは比較的最近であり、いろいろなグループによって同様の考えが展開されている。たとえば、いわゆる最大エントロピー法のグループ (Gull や Skilling など) は少しづつ主張の重点を移した結果 (?)、現在では階層ベイズ法とほぼ同じ枠組を主張しているようである。この流れでは、MacKay の仕事が Neural Network やそれに関連した統計物理の分野で普及している。MacKay の仕事の新しい点は主として多層パーセプトロンへの応用であるが、階層ベイズ法の解説自体が評価されている面もある (日本の統計学者の寄与に全く触れていない点は良くないが)。Ph.D 論文 ([MacKay 1991]) は数編にわけて雑誌 "Neural Computation" に掲載された (1992) が、MacKay のホームページ `http://131.111.48.24/mackay/`で見ることもできる。より新しい解説としては、

[MacKay 1996]
MacKay, D.J.C. (1996), Bayesian methods for backpropagation networks, in Models of Neural Network III, — Association, Generalization and Representation —, Eds. E.Domany, J.L.van Hemmen, K.Schulten, Springer.

がある。別の重要な分野としては、音声認識やタンパク質の配列の解析で利用されている "隠れマルコフモデル (HMM)" による情報処理がある。HMM は階層モデルの一種であり、[Kitagawa and Gersch 1996] でいう (一般の) 状態空間モデルと近い関係にある。

ベイズ統計が流行だというので、"ベイズ" と書いてある本や論文をみても、上記のような話題が論じられているとは限らない。Baxter には " イジング模型の 399 番目の解" という論文があるが、ベイズ統計学者にもそのくらいの種類の流派があるようである。逆に、平滑化や罰金つき最尤法に対して、階層ベイズ法以外のアプローチをすることもできる。平滑化に対する別の立場の例としては、

[Wahba 1990]
Wahba,G.(1990), Spline Models for Observational Data, SIAM, Philadelphia, Pennsylvania.

をあげておく。

階層ベイズモデルに対して統計物理の手法を応用して理論的な接近を試みた論文としては、たとえば、

[Bruce and Saad 1994]
Bruce,A.D. and Saad,D. (1994), Statistical mechanics of hypothesis evaluation, Journal of Physics A (Mathematical and General), 27, pp.3355-3363.

がある。

2.4　人工知能・認知科学との関連

以上のような解説を読むと、いわゆる "人工知能" と統計学的な考えの関係について議論したくなってくる。この問題について広い見地から論じたのが、

[伊庭 1996b]
伊庭幸人 (1996), 基礎的問題から見た情報統合, 人工知能学会誌, Vol.11, No.2, (1996年3月号), pp.193-200.

である。ただし紙数の制限からきわめて圧縮した記述になっている。

10年前に筆者が物理学科の大学院にいて、統計学に興味を持ちはじめたときは、Neural Network の流行の初期で、形式論理ベースの人工知能がまだ盛んであった。当時は、[坂元 1985] の前半 (AIC を使って分割表を解析する部分) を読んで、"AIC(変分原理) と Simulated Annealing(最適化手法) を組み合わせれば人工知能ができる" などと思っていた。また、後述の Geman の論文をみて文献を請求したり、Hinton たちのボルツマンマシンの仕事をみて "うーむ" と思った記憶がある。それに比べると、[伊庭 1996b] ではずいぶん悲観論になっているかもしれない。

[伊庭 1996b] では人工知能一般について述べたが、階層ベイズモデルや罰金付最尤法の初期視覚・運動制御への応用については

[Poggio et al. 1985]
Poggio,T., Torre,V. and Koch,C.(1985), Computational vision and regularization theory, Nature, vol.317, pp.314-319.

[川人 1996]
川人光男 (1996), 脳の計算理論, 産業図書.

を参照されたい (画像についてのより工学的・統計学的要素の多い研究については、4.3節を参照)。

これらの研究は David Marr の、

[マー 87]
Marr,D. (1987), ビジョン — 視覚の計算理論と脳内表現 —, 乾 敏郎, 安藤 広志 訳, 産業図書.

(原著)
Marr,D. (1982), VISION, A Computational Investigation into the Human Representation and Processing of Visual Information, Freeman.

に強く影響されている。もっとも Marr 自身は統計モデルを用いた定式化を重要とみなしていないし、変分原理的な側面も意外と希薄である。Poggio たちの定式化は本質を明らかにする上で有効だったというのが筆

者の立場であるが、Marr の良い点は別のところにあり、Poggio たちの仕事はそれを堕落させたという主張の人もある。また、Marr の主張そのものが今日の (ロボット) 研究において悪影響を残しているという人、Marr が大好きなので批判されるとわがことのように悲しいという人、Marr に影響を受けたのは素人であってプロ (の画像処理屋) への影響はほとんどないという人、などなど、いろいろな意見があって議論はつきないが、この位にする。ちなみに Marr 自身は重病を押して上の本の原稿を完成したのち、35 歳の若さでなくなったので、彼自身の意見はもう聞くことができない。

3　動的モンテカルロ法とその周辺

3.1　動的モンテカルロ法入門

物理、とくに物性物理では、モンテカルロ法というとメトロポリス法や heat bath 法などマルコフ連鎖を利用したものがほとんどであるから、単にモンテカルロ法で通じてしまう。単純なランダムサンプリングとか global な重み関数を利用したサンプリングと区別するときは表題のように "動的モンテカルロ法" と呼ぶようである。ここ 10 年ほどで統計学や工学でもマルコフ連鎖を利用した方法が使われるようになったが、これらの分野では他のタイプのモンテカルロ法が従来から使われてきたので、区別のための総称がどうしても必要である。統計学者は "マルコフ連鎖モンテカルロ法"(MCMC または $(MC)^2$ と略す) という名称を使っている。統計学では "Gibbs Sampler" というのもよく聞くが、これはマルコフ連鎖を利用する方法の総称ではなく、特定のサンプリング法 (物理でいう heat bath 法とほぼ同じ?) の意味である。Neural Network 関係の人や工学者は "ボルツマンマシン" という名前が好きだが、これは本当は困る。"ボルツマンマシン" は、モデルの名前、学習法の名前、動的モンテカルロ法の同義語、simulated annealing の同義語、のすべてに使われるため誤解の原因になるからである。

古典スピン系を材料にして、統計物理の基礎を動的モンテカルロ法と絡めて解説したのが、

[伊庭 1996c]
伊庭幸人 (1996), 統計学者・数理工学者のための統計物理入門 — 暫定版 (2.1版), — 格子スピン模型とマルコフ連鎖モンテカルロ法を中心にして —, ISM Reseach Memorandum No.592, January, 1996. (`http://www.ism.ac.jp/~iba/`で図を除いて入手可能)

である。これは、ある事情で [伊庭 1990] と同時に書いた原稿を、本来の用途が没になったあと、書き直したものである。書き直してレビューとして某誌に送ったが、載せてもらえなかった。そのあと、その一部を利用して、ほとんど新規に書いたのが、

[伊庭 1996d]
伊庭幸人 (1996), マルコフ連鎖モンテカルロ法とその統計学への応用, 統計数理 44, No.1,(特集・計算統計学の発展), pp.49-84.

である。こちらはアルゴリズム中心の記述になっている。

統計物理・物性物理の方でモンテカルロ法といえば、Binder が編集したシリーズが有名である。

[Binder 1986]
Binder,K.(ed.)(1986), Monte Carlo Methods in Statistical Physics (2nd ed.), Topics in Current Physics, Vol.7, Springer.

[Binder 1987]
Binder,K.(ed.)(1987), Applications of the Monte Carlo Method in Statistical Physics (2nd ed.), Topics in Current Physics, Vol.36, Springer.

[Binder 1995]
Binder,K.(ed.)(1995), The Monte Carlo Method in Condensed Matter Physics (2nd ed.), Topics in Applied Physics, Vol.71, Springer.

おのおの別の内容で、それぞれについて改版していく方針らしい。現在はどれも第2版だと思う。筆者が修士課程の一年のとき、ゼミで、1番目の本の初版をやった。冒頭の概説が当たったのだが、専門的でまいった。

概説くらいはもっと取り付きやすくして欲しいと思った。Binder たちもこれではまずいと思ったのか、2 冊目の冒頭の概説、

> [Binder and Stauffer 1984]
> Binder,K. and Stauffer,D. (1984), A simple introduction to Monte Carlo simulation and some specialized topics, in Applications of the Monte Carlo Method in Statistical Physics, Topics in Current Physics, Vol.36, Ed. K. Binder, Springer. (2nd ed.(1987) にも載っていると思うが未確認.)

はもっと親切になっている。また、Binder と Heermann による教科書も Springer から出た。

> [Binder and Heermann 1992]
> Binder,K. and Heermann,D.W.(1992), Monte Carlo Simulation in Statistical Physics (2nd ed.), Springer.

これら (特に後者) は、相転移の解析法に多くの頁がさかれていて、統計学や工学への応用に興味がある人にはあまり向いていない。後者では、プログラミングの初歩も説明されているので、計算機の経験の少ない人には良いかもしれない。

最近は、統計学者の書いた統計学者のための教科書が出ているが、これは 4.1 節で取り上げる。

3.2　分布からのサンプリング .vs. 最適化

統計物理や素粒子物理以外で動的モンテカルロ法を用いるときは、simulated annealing 法としての使用、つまり、最適化アルゴリズムとしての利用法に限られると思っている人が多い。たしかに、巡回セールスマン問題などでは、有限温度での状態に直接の意味を見い出すことは困難である。こうした問題では、確率分布を考えることは local optima を避けるための手段に過ぎない。しかし、統計学においては、もともと確率分布が主題になっているわけであるから、分布からのサンプリング自体が重要な関心事であり、サンプリングの手段としての動的モンテカルロ法が有効である。情報理論やいわゆる “Neural Network” の分野の一部でも同じである。サンプリングの手段として動的モンテカルロ法を用いることを、simulated annealing (温度 $T \to 0$) に対して、“温度 $T=1$ での利用”

と呼ぶこともできるが、物理以外では、もとの問題に温度という概念があるわけではないから、その意味では無理のある表現である。

もちろん、simulated annealing も統計学の問題に使える。4.3 節で紹介する Geman and Geman の論文での動的モンテカルロ法の応用はほとんど annealing によるものである (それ以外の応用についても触れられているのだが、とくに 1984 年の論文では、よほど注意深く読まないと気がつかないだろう)。動的モンテカルロ法が simulated annealing として適用される問題と確率分布からのサンプリング法 (あるいはそれを利用しての期待値・周辺分布の計算法) として適用される問題の違いについては、統計学者のほうでも混乱があって、最近ようやく整理されてきたというのが実情であろう。

“期待値・周辺分布の計算” 対 “最適化” という構図は動的モンテカルロ法以外にも広く存在する。たとえば、(tanh のような関数で階段関数を置き換える) Hopfield-Tank タイプの最適化手法と平均場近似の関係は統計物理学者の大多数にとっては (おそらく Hopfield 自身にとっても) 自明であろう。しかし、物理学者以外では相当多くの人が、(周辺) 確率を近似する方法としてそれが使えるということを知らなかった。逆に物理学者の多くは高次元の確率分布を扱っている人達が他にいるとは思わずに、“彼ら” のやりたいのは “最適化” だと思い込んでいたようである。それで、かなりあとになって

> [Peterson and Hartman 1989]
> Peterson,C. and Hartman,E. (1989), Explorations of the mean field thoery learning algorithm, Neural Networks, 2, pp.475-494.

が出て、話題になったりした。その後、画像に対する平均場近似の応用が論じられたりしたが、いまでも、“Hopfield-Tank の方法” は知っていても平均場近似としての意味は知らないという人は多いと思う。

ついでながら、“ニューロでやる”、“ニューロで解く” などという表現は教養のある人はなるべく使わないでほしい。本物の “ニューロ” つまり脳を研究している人に失礼だという点はさておくとして、意味が特定できないのである。多層パーセプトロンを当てはめて予測を行なうのも、Hopfield-Tank 法で巡回セールスマン問題を解くのも、連想記憶にパターンを記憶させるのも、動的モンテカルロ法を温度 1 で利用して分布の推定をするのも、みんな “ニューロで解く” になってしまう (そして、次に出るのは大抵 “ニューロは嫌いだ” である。とほほ。)。

ここで述べたことを表にまとめると、

最適化	期待値の計算
Simulated Annealing 法	動的モンテカルロ法 (= マルコフ連鎖モンテカルロ法)
Hopfield-Tank 法	平均場近似

といった感じになるだろう。左側は最適化手法を必要とする分野一般で、右は統計物理学、統計学など、多変数の確率分布を扱う分野で有効である。この表はもっと拡張できて、

最適化	期待値・積分の計算
Simulated Annealing 法	動的モンテカルロ法 (= マルコフ連鎖モンテカルロ法)
Hopfield-Tank 法	平均場近似
動的プログラミング	転送行列法・転送積分法 非ガウスフィルタ Baum-Welch 法 (の尤度を計算する部分)
遺伝的アルゴリズム	Green 関数モンテカルロ・拡散モンテカルロ 転送行列モンテカルロ モンテカルロフィルタ
マルチグリッド法 (による最適化)	実空間繰り込み群

のようなものも書ける。この表の解説を書くと大変長くなるので、いろいろ異論や質問もあると思うが、ここはこれだけにする。

4　動的モンテカルロ法と統計学

4.1　動的モンテカルロ法と統計学一般

動的モンテカルロ法の統計学への応用についてのまとまった情報は、

[Gilks et al. 1996]
Gilks,W., Spiegelhalter,D. and Richardson,S.(eds.)(1996), Markov Chain Monte Carlo in practice, Chapman and Hall.

にある。その他、[伊庭 1996d] の文献リスト、特にそこにあげたいくつかの総合報告を参照されたい ([伊庭 1996d] は筆者のホームページでみられ

る。ただし、現在入っているのは最終原稿より前の版)。また、動的モンテカルロ法に関する統計関係のプレプリントサーバーが、
`http://www.stats.bris.ac.uk/MCMC/`にある。

統計物理や素粒子物理では、通常の動的モンテカルロ法以外にいわゆる分子動力学法やそれとモンテカルロ法を混合した方法(ハイブリッド法)が使われている。これを多層パーセプトロンなどに適用した研究が、

[Neal 1996]
Neal,R.M. (1996), Bayesian Learning for Neural Networks, Lecture Notes in Statistics 118, Springer.

である。階層ベイズ法や動的モンテカルロ法一般についても解説されており、読みやすそうである。Neal は Hinton の学生で MacKay の影響も受けているらしい(ホームページ
`http://www.cs.toronto.edu/~radford/`)。

4.2 遺伝の問題への応用

ワークショップでの講演の際に例としてあげた、遺伝の問題 (pedigree analysis, linkage analysis) への応用については、たとえば、

[Thompson 1991]
Thompson,E.A.(1991), Probabilities on complex pedigrees; the Gibbs sampler approach, in Computing Science and Statistics: Proc. of the 23rd symposium on the interface(Ed. E.M.Keramides), pp.371-378, Interface Foundation, Fairfax Station, Va.

を参照されたい。遺伝子伝搬に関する問題や DNA 配列に関する問題は、離散変数を含んだ複雑な統計モデルが考えられるため、動的モンテカルロ法の応用分野として注目されている。

4.3 画像処理・空間統計への応用

筆者がはじめて読んだ “ベイズ統計” と題する文献は、

[Geman and Geman 1986]
Geman,D. and Geman,S.(1986), Bayesian image analysis, in

Disordered Systems and Biological Organization, NATO ASI series F20, Eds. E.Bienenstock et al., Springer.

だったかもしれない。単なるannealingの応用とは何か違うとぼんやり感じたことを記憶している。有名なのは、

[Geman and Geman 1984]
Geman,S. and Geman,D.(1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, pp.721-741.

の方であるが、こちらは定理AとかBとかあってあまり馴染めない。1984年以降のGemanたちの仕事の一部は[Geman and Geman 1986]に出ているが、より詳しくは[Iba 1996d]の文献、あるいは筆者のホームページの文献リストを参照されたい(といっても、最近の仕事はfollowしていないが)。

画像関係を中心に動的モンテカルロ法を解説した本に、

[Winkler 1995]
Winkler,G.(1995), Image analysis, Random fields and Dynamic Monte Carlo Methods, — A Mathematical Introduction —, Springer.

がある。副題からすると、著者自身はマルコフ連鎖の数学的な面に興味があるようであるが、応用的なこともひととおり触れてある。(非ガウス・離散変数の)ベイズモデルを用いた画像処理については、簡潔な

[Besag 1989]
Besag,J. (1989), Towards Bayesian image analysis, Journal of Applied Statistics, 16, No.3, pp.395-407.

も良いかもしれない。同じ雑誌のひとつ前の号

Journal of Applied Statistics, 16, No.2, (1989).

は統計的画像処理の特集である。たとえば、

[Grenander and Keenan 1989]
Grenander,U. and Keenan,D.M.(1989), Towards automated image understanding, Journal of Applied Statistics, 16, No.2, pp.207-221.

は、画像理解についての独自の見解を各種の例を通じて述べており、興味深い。

統計数理研究所の尾形と種村は、統計学で動的モンテカルロ法が広く話題になる以前から、これを利用した仕事を多数発表している。空間の点配置の解析に関する応用が多い。

[Ogata and Tanemura 1981]
Ogata,Y. and Tanemura,M.(1981), Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure, Annals of the Institute of Statistical Mathematics, 33 B, pp.315-338.

[長谷川・種村 1986]
長谷川政美, 種村正美 (1986), なわばりの生態学, — 生態のモデルと空間パターンの統計 —, 東海大学出版会.

[Ogata and Tanemura 1989]
Ogata,Y. and Tanemura,M.(1989), Likelihood estimation of soft-core interaction potentials for Gibbsian point patterns, Annals of the Institute of Statistical Mathematics, 41, pp.583-600.

これらは先駆的な研究として高く評価されるべきものである。尾形は階層ベイズモデルに対する動的モンテカルロ法の応用も行なっている。

文献一覧

[赤池 1976]
赤池弘次 (1976), 情報量規準 AIC とは何か, 数理科学 No.153, 1976 年 3 月号 (特集・情報量規準), pp.5-11.

[Akaike 1980]
Akaike,H.(1980), Likelihood and Bayes procedure, in Bayesian Statistics, Eds. Bernardo,J.M, DeGroot,M.H., Lindley,D.V., and Smith,A.F.M., University press, Valencia.

[赤池 1981]
赤池弘次 (1981), モデルによってデータを測る, 数理科学 No.213, 1981 年 3 月号 (特集・統計モデル), pp.7-10.

[Besag 1989]
Besag,J. (1989), Towards Bayesian image analysis, Journal of Applied Statistics, 16, No.3, pp.395-407.

[Binder 1986, 1987, 1995]
Binder,K.(ed.)(1986), Monte Carlo Methods in Statistical Physics (2nd ed.), Topics in Current Physics vol.7, Springer.
Binder,K.(ed.)(1987), Applications of the Monte Carlo Method in Statistical Physics (2nd ed.), Topics in Current Physics vol.36, Springer.
Binder,K.(ed.)(1995), The Monte Carlo Method in Condensed Matter Physics (2nd ed.), Topics in Applied Physics vol.71, Springer.

[Binder and Heermann 1992]
Binder,K. and Heermann,D.W.(1992), Monte Carlo Simulation in Statistical Physics (2nd ed.), Springer.

[Binder and Stauffer 1984]
Binder,K. and Stauffer,D. (1984) A simple introduction to Monte Carlo simulation and some specialized topics, in Applications of the Monte Carlo Method in Statistical Physics, Topics in current physics vol.36, Ed. K. Binder, Springer. (2nd ed.(1987) にも載っていると思うが未確認.)

[Bruce and Saad 1994]
Bruce,A.D. and Saad,D. (1994), Statistical mechanics of hypothesis evaluation, Journal of Physics A (Mathematical and General), 27, pp.3355-3363.

[Geman and Geman 1984]
Geman,S. and Geman,D.(1984), Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, IEEE Transactions on Pattern Analysis and Machine Intelligence, 6, pp.721-741.

[Geman and Geman 1986]
Geman,D. and Geman,S.(1986), Bayesian image analysis, in Disordered Systems and Biological Organization, NATO ASI series F20, Eds. E.Bienenstock et al., Springer.

[Gilks et al. 1996]
Gilks,W., Spiegelhalter,D. and Richardson,S.(eds.)(1996), Markov Chain Monte

Carlo in practice, Chapman and Hall.

[Grenander and Keenan 1989]

Grenander,U. and Keenan,D.M.(1989), Towards automated image understanding, Journal of Applied Statistics, 16, No.2, pp.207-221.

[韓 1987]

韓太舜 (はん・てすん) (1987), 数理科学 No.290, 1987年8月号 (特集・情報圧縮),pp.5-15.

[長谷川・種村 1986]

長谷川政美, 種村正美 (1986), なわばりの生態学, — 生態のモデルと空間パターンの統計 —, 東海大学出版会.

[伊庭 1990]

伊庭幸人 (1990), 統計物理と統計的情報処理 (1.00版), — 大規模・非ガウスモデルをめぐる話題 —, (図を除いて `http://www.ism.ac.jp/~iba/` で入手可能).

[伊庭 1993]

伊庭幸人 (1993), ベイズ統計と統計物理, — 有限温度での情報処理 —, 「物性研究」60-6 (1993年9月号), pp.677-699.

伊庭幸人 (1996), ベイズ統計と統計物理 (物性研究1993年9月号) への訂正と追加, 「物性研究」65-5 (1996年2月号), pp.678-685.

[伊庭 1996a]

伊庭幸人 (1996), 学習と階層, — ベイズ統計の立場から —, 「物性研究」65-5 (1996年2月号), pp.657-677.

[伊庭 1996b]

伊庭幸人 (1996), 基礎的問題から見た情報統合, 人工知能学会誌, Vol.11, No.2, (1996年3月号), pp.193-200.

[伊庭 1996c]

伊庭幸人 (1996), 統計学者・数理工学者のための統計物理入門 — 暫定版 (2.1版), — 格子スピン模型とマルコフ連鎖モンテカルロ法を中心にして —, ISM Reseach Memorandum No.592, January, 1996. (`http://www.ism.ac.jp/~iba/` で図を除いて入手可能)

[伊庭 1996d]

伊庭幸人 (1996), マルコフ連鎖モンテカルロ法とその統計学への応用, 統計数理 44, No.1,(特集・計算統計学の発展), pp.49-84.

[川人 1996]

川人光男 (1996), 脳の計算理論, 産業図書.

[Kitagawa and Gersch 1996]

Kitagawa,G and Gersch,W.(1996), Smoothness Priors Analysis of Time Series, Lecture Notes in Statistics, No.116, Springer.

[MacKay 1991]

学位論文。 数編にわけて雑誌 "Neural Computation" に掲載された (1992) が、MacKay のホームページ `http://131.111.48.24/mackay/` で見ることもできる。

[MacKay 1996]

MacKay, D.J.C. (1996), Bayesian methods for backpropagation networks, in Models of Neural Network III, — Association, Generalization and Represen-

tation —, Eds. E.Domany, J.L.van Hemmen, K.Schulten, Springer.
[マー 87]
Marr,D. (1987), ビジョン — 視覚の計算理論と脳内表現 —, 乾 敏郎, 安藤 広志 訳, 産業図書.
(原著) Marr,D. (1982), VISION, A Computational Investigation into the Human Representation and Processing of Visual Information, Freeman.
[Neal 1996]
Neal,R.M. (1996), Bayesian Learning for Neural Networks, Lecture Notes in Statistics 118, Springer.
[Ogata and Tanemura 1981]
Ogata,Y. and Tanemura,M.(1981), Estimation of interaction potentials of spatial point patterns through the maximum likelihood procedure, Annals of the Institute of Statistical Mathematics, 33 B, pp.315-338.
[Ogata and Tanemura 1989]
Ogata,Y. and Tanemura,M.(1989), Likelihood estimation of soft-core interaction potentials for Gibbsian point patterns, Annals of the Institute of Statistical Mathematics, 41, pp.583-600.
[Peterson and Hartman 1989]
Peterson,C. and Hartman,E. (1989), Explorations of the mean field thoery learning algorithm, Neural Networks, 2, pp.475-494.
[Poggio et al. 1985]
Poggio,T., Torre,V. and Koch,C.(1985), Computational vision and regularization theory, Nature, vol.317, pp.314-319.
[Rissanen 1983]
Rissanen (1983), A universal prior for integers and estimation by minimum description length, The Annals of Statistics, Vol.11, No.2, pp.416-431, 1983.
[Rissanen 1989]
Rissanen,J. (1989), Stochastic Complexity in Statistical Inquiry, World Scientific.
[坂元 1985]
坂元慶行 (1985), カテゴリカルデータのモデル分析, 共立出版.
(英語版) Sakamoto,Y.(1991), Categorical Data Analysis by AIC, Kluwer.
[坂元ほか 1983]
坂元慶行, 石黒真木夫, 北川源四郎 (1983), 情報量統計学, 共立出版.
(英語版) Sakamoto,Y., Ishiguro,Y. and Kitagawa,M. (1986) Akaike Information Criterion Statistics, Reidel.
[竹内 1963]
竹内啓 (1963), 数理統計学, — データ解析の方法 —, 東洋経済新報.
[竹内 1980]
竹内啓 (1980), 現象と行動のなかの統計数理, 新曜社.
[竹内ほか 1989]
竹内啓ほか編 (1989), 統計学辞典, 東洋経済新報.

[Thompson 1991]
Thompson,E.A.(1991), Probabilities on complex pedigrees; the Gibbs sampler approach,New in Computing Science and Statistics: Proc. of the 23rd symposium on the interface(Ed. E.M.Keramides), pp.371-378, Interface Foundation, Fairfax Station, Va.
[Wahba 1990]
Wahba,G.(1990), Spline Models for Observational Data, SIAM, Philadelphia, Pennsylvania.
[Winkler 1995]
Winkler,G.(1995), Image analysis, Random fields and Dynamic Monte Carlo Methods, — A Mathematical Introduction —, Springer.