宇宙論パラメータとデータ科学的方法

池田 思朗　**数理・推論研究系**

はじめに

2014年度からはじまったCRESTのプロジェクト「広域撮像探査観測のビッグデータ分析による統計計算宇宙物理学」では，すばるHSC(Hyper-Suprime Cam)の戦略枠観測プログラム(SSP: Subaru Strategic Program)を通じて得られたデータによって，宇宙を記述する代表的なモデルである標準宇宙模型の6つのパラメータ(以下，宇宙論パラメータ)の推定精度を高めることを目的としている．

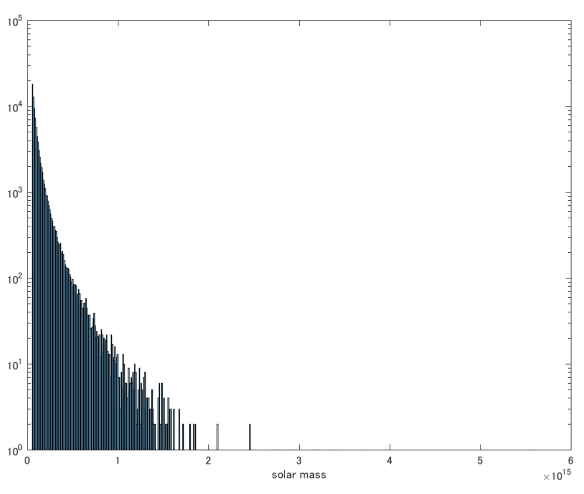
複数のアプローチ

CRESTのプロジェクトでは複数の方法で宇宙パラメータの推定を行おうとしている．ひとつはIa型超新星の赤方偏移を用いる方法である．HSC-SSPでは数多くの超新星を発見すると期待されている．Ia型超新星は絶対等級がほぼ同じだと考えられているため，見かけの明るさが距離を与える．他の観測機器によって赤方偏移を求めれば，地球との相対速度が求まる．これらを合わせれば宇宙の膨張速度を測ることができる．我々はHSCの観測画像から自動的にIa型超新星の候補天体をみつける判別器をNTT CS研と共同で開発し，実際の観測で用いている[1]．Ia型超新星を用いる方法の他には，重力レンズ効果を用いた宇宙論パラメータの推定も行われている．このアプローチについても成果がでつつある．

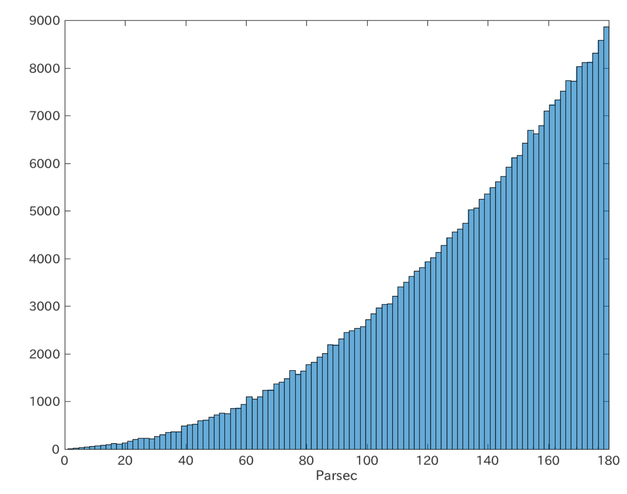
もうひとつ，本プロジェクト内の課題であり今後重要となるものは，シミュテーションとの融合である．宇宙論パラメータを指定すれば，計算機によって宇宙をシミュレートすることが可能となっている．このシミュレーションによって得られた宇宙と実際に観測される宇宙とを比較することによって宇宙論パラメータを推定しようというのだ．ABC (Approximate Bayesian Computation)と同じ発想の研究である．

Fisher 情報量と推定誤差

シミュレーションで得られた宇宙と観測された宇宙とを比較する際，どのような統計量を用いるべきかは確立されていない．



(a) ダークマターの重さの分布．



(b) ある質量以上のダークマターハロー間の距離の分布．

Figure 1: ある宇宙論パラメータを指定して得られた宇宙から統計量を取り出した結果(西道氏からデータの提供をうけた)．

宇宙論パラメータを θ として，用いる統計量を $T(\theta)$ とする．どのような統計量を用いるか，どうやって観測された統計量からパラメータ θ を推定するかは別にして，推定されたパラメータ $\hat{\theta}$ の共分散行列は統計的に振る舞うはずでありCramér-Raoの不等式にしたがうはずである．すなわち，$I(\theta)$ をFisher 情報量行列，N をサンプルサイズとして次のようにかけるはずである．

$$\mathrm{var}(\hat{\theta}) \geq \frac{1}{N} I(\theta)^{-1}.$$

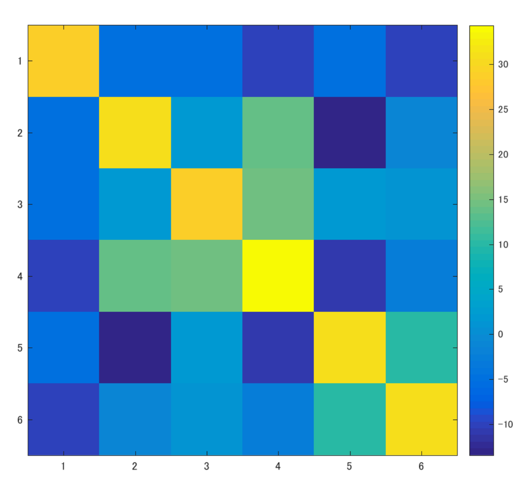
Fisher 情報量行列を求めれば，どの統計量を用いて宇宙論パラメータを推定したかによって推定されたパラメータがどのような分散構造を持つかが予測できる．これは宇宙論のためにどの統計量を用いるべきかを議論する際，重要な指標となる．

宇宙模型ではFigure 1にある分布の表現は与えられないので，Fisher 情報行列を解析的に求めることはできない．しかし，情報幾何学から知られているように，Kullback Leibler ダイバージェンスや Hellinger距離はパラメータの変化が小さいならば，次のような二次形式で近似できる．

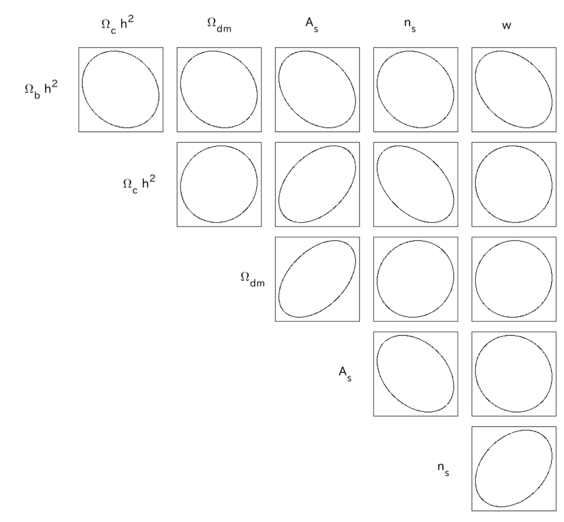
$$D(p(T;\theta);p(T;\theta+\delta)) \simeq \frac{1}{2}\delta' I(\theta)\delta.$$

したがって，θ の近傍で様々な方向に θ を変化させ，それぞれのパラメータの間の Kullback Leibler ダイバージェンスや Hellinger 距離がわかれば，その値から Fisher 情報量行列を推定できる．ダイバージェンスや距離を推定する様々なノンパラメトリックな方法が提案されていることから，そうした方法を用いて $D(p(T;\theta);p(T;\theta+\delta))$ を推定し，Fisher 情報行列を計算すれば，統計量の性質を知ることが可能となる．

結果と考察

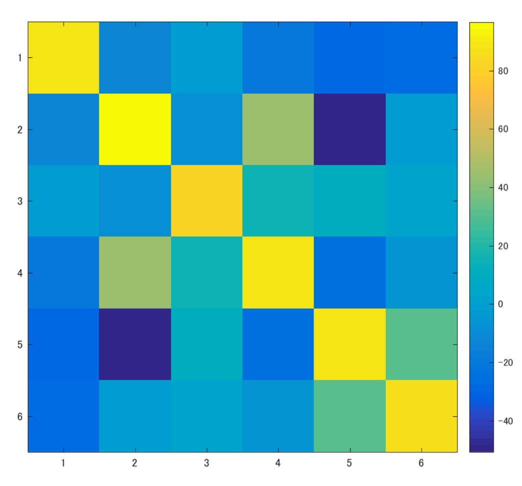


(a) Hellinger 距離を用いて計算した Fisher 情報行列．

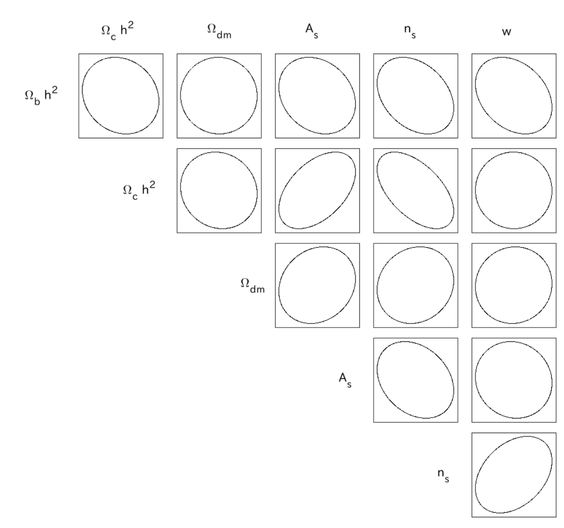


(b) 各変数間の共分散構造．

Figure 2: ダークマターの重さに対する Fisher 情報行列．



(a) Hellinger 距離を用いて計算した Fisher 情報行列．



(b) 各変数間の共分散構造．

Figure 3: ある質量以上のダークマターハロー間の距離の分布に対する Fisher 情報行列．

今回はある宇宙論パラメータの値 θ_o から24方向にパラメータを振り，宇宙をシミュレートした．そのデータから統計量を指定し，k-nn 方を用いてノンパラメトリックに Hellinger 距離を求めた．そうした結果から Fisher 情報量行列を計算したのが上の結果である．この方法を洗練し，宇宙論パラメータを推定するための統計量の提案につなげていきたい．

謝辞

本研究は JST CREST 「広域撮像探査観測のビッグデータ分析による統計計算宇宙物理学」の研究として行った．共同で研究を行った西道啓博, 吉田直紀(東京大学)，上田修功(NTT, RIKEN AIP), 森井幹雄(敬称略)に感謝する．

参考文献

[1] Mikio Morii, Shiro Ikeda, Nozomu Tominaga, Masaomi Tanaka, Tomoki Morokuma, Katsuhiko Ishiguro, Junji Yamato, Naonori Ueda, Naotaka Suzuki, Naoki Yasuda, and Naoki Yoshida. Machine-learning selection of optical transients in Subaru/Hyper Suprime-Cam survey. *Publications of the Astronomical Society of Japan*, 68(6):104(8pages), 2016. arXiv:1609.03249, doi:10.1093/pasj/psw096.