

Risa/Asir と統計システム R の連携

高山信毅 (神戸大学/JST CREST)

2013.03.17

Holonomic Gradient Method (ホロノミック勾配法, [N₃OST₂; 2011]): 統計的推定への応用を念頭において, ホロノミック関数 $f(\theta, x)$ および領域 R が与えられた時, **正規化定数**

$$z(\theta) = \int_R f(\theta, x) dx$$

の近似値を計算する方法.

- ① $z(\theta)$ の満す微分方程式を求める.
- ② $z(\theta)$ およびその偏微分の近似値をうまい θ で求める. 初期値.
- ③ 微分方程式を数値解析して $z(\theta)$ の値をきめていく.

-
- ① **D-加群の積分アルゴリズム** `nk_restriction.rr`, `yang.rr` [...] or **問題に応じた工夫**
 - ② **数値積分** [...] or **問題に応じた工夫**

Von-Mises 分布. $f(\theta, x) = \exp(\theta_1 \cos x + \theta_2 \sin x)$.

$$z(\theta) = \int_0^{2\pi} \exp(\theta_1 \cos x + \theta_2 \sin x) dx$$

$F = (z, \partial_1 z)^T$, $\partial_i = \partial/\partial\theta_i$. Pfaffian system は

$$\frac{\partial F}{\partial\theta_1} = \begin{pmatrix} 0 & 1 \\ \frac{\theta_1^2}{\theta_1^2 + \theta_2^2} & \frac{\theta_2^2 - \theta_1^2}{\theta_1(\theta_1^2 + \theta_2^2)} \end{pmatrix} F$$

$$\frac{\partial F}{\partial\theta_2} = \begin{pmatrix} 0 & \theta_2/\theta_1 \\ \frac{\theta_1\theta_2}{\theta_1^2 + \theta_2^2} & \frac{-2\theta_2}{\theta_1^2 + \theta_2^2} \end{pmatrix} F$$

Fisher の最尤法. X_i が (観測) データ

$$\prod_i \frac{f(\theta, X_i)}{z(\theta)}$$

を最大にする θ を見つける. HGD, D=descent.

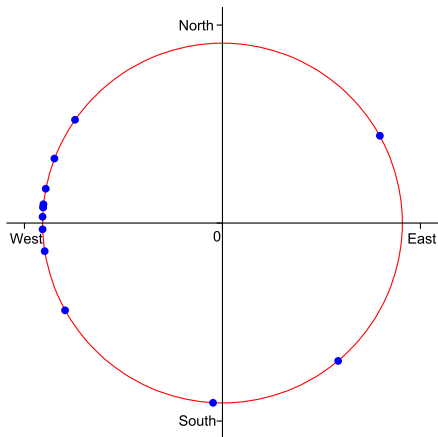


Figure: 札幌上空約 1 万メートル地点における午前 9 時の風向きの 13 日間のデータ (2011 年 1 月 1 日から 14 日 (11 日は欠損): 気象庁の気象統計情報から)

$(\cos t, \sin t)$ 達を上記の 13 件のデータとする. HGD により

```
A=test1(); isFit(A);
```

$$\theta = (\theta_1, \theta_2) = (-0.1286, -1.5963).$$

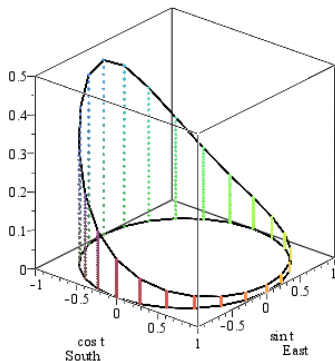


Figure: 推定したパラメータによる風向分布

HGM, HGD の研究 1(本日の発表に関係あるもの).

Fisher 分布. 清, 柴田, 竹村, 小原, 高山: arXiv:1110.0721, to appear in J. Multivariate Analysis.

$X, \Theta = (\theta_{ij})_{ij}$ は \mathbf{R} 上の 3 次正方行列とし, Θ^\top で Θ の転置行列を表す. $\partial_{ij} = \frac{\partial}{\partial \theta_{ij}}$ と置き, μ を $SO(3)$ 上の不変確率測度とする. サンプル平均行列 X に関する, $SO(3)$ 上の Fisher 分布の尤度関数 $f(X; \Theta)$ は次のように表される.

$$f(X; \Theta) = \frac{\exp(\text{Tr}(\Theta^\top X))}{z(\Theta)}, \quad \text{ここで} \quad z(\Theta) = \int_{SO(3)} \exp(\text{Tr}(\Theta^\top X)) d\mu$$

正規化定数 $z(\Theta)$ は Fisher 積分とも呼ばれる.

`hgm.so3nc()`, デモ `h-mle/SO/Prog/hgm-comets.r`

HGM, HGD の研究 2(本日の発表に関係あるもの).

橋口, 沼田, 高山, 竹村 : arxiv:1201.0427, to appear in J. Multivariate Analysis.

$$\frac{\Gamma_m(a)\Gamma_m(c-a)}{\Gamma_m(c)} {}_1F_1(a; c; Y)$$
$$= \int_{0 < X < I_m} \exp(\text{Tr } XY) |X|^{a-(m+1)/2} |I_m - X|^{c-a-(m+1)/2} dX,$$

$0 < X < I_m$ は X および $I_m - X$ が正定値対称行列であることを意味する. $dX = \prod_{i \leq j} dx_{ij}$ は X の上三角成分についてのルベーク測度. それから

$$\Gamma_m(a) = \pi^{\frac{1}{4}m(m-1)} \prod_{i=1}^m \Gamma\left(a - \frac{i-1}{2}\right).$$

X_1, \dots, X_n を, \mathbf{R}^m における平均 0, 共分散行列 Σ の多変量正規分布に従う, 独立な確率変数とする ($n \geq m$). X_i を縦ベクトルとみなし, $m \times n$ 行列 X を (X_1, \dots, X_n) と定義し, $m \times m$ 行列 Y を XX^T で定義する. Y は自由度 n の Wishart 分布に従う確率変数となる. 対称行列 Y の最大固有値を ℓ_1 とする. このとき

$$\Pr[\ell_1 < x] = C \exp\left(-\frac{x}{2} \text{Tr} \Sigma^{-1}\right) x^{\frac{1}{2}nm} {}_1F_1\left(\frac{m+1}{2}; \frac{n+m+1}{2}; \frac{x}{2} \Sigma^{-1}\right),$$

$$C = \frac{\Gamma_m\left(\frac{m+1}{2}\right)}{2^{\frac{1}{2}nm} (\det \Sigma)^{\frac{1}{2}n} \Gamma_m\left(\frac{n+m+1}{2}\right)}.$$

これは χ^2 分布 ($m=1$) の一般化. `hgm.cwishart(m,n,x)` 後出
beta は $\Sigma^{-1}/2$.

Wikipedia より

R 言語 (アールげんご) は、オープンソースでフリーソフトウェアの統計解析向けプログラミング言語、及びその開発実行環境である。R 言語は、ニュージーランドのオークランド大学の Ross Ihaka と Robert Gentleman により作られた。現在では、R Development Core Team (S 言語開発者である John M. Chambers も参画。R Project Contributors) によって、メンテナンスと拡張がなされている。

... 略 ...

2009 年 7 月、SAS Institute は R Interface Coming to SAS/IML Studio によって SAS から R 言語へのインタフェースを提供することを発表した。公式ページには、**新たな統計手法はたいていの場合まっさきに R 言語上で実装されるという現状**をふまえ SAS ユーザーの要望に応じてインタフェースの提供を行なう、との旨が述べられている。

Asir と R の連携 1 (プロトタイプ)

tk_jack.rr から C のコード, R のパッケージ hgm へ.

```
import("tk_jack.rr");  
tk_jack.zonal([2,1],3);
```

- ① 微分方程式. OpenXM/src/hgm/mh/src/code-n.c
- ② 初期値. Koev-Edelman (2006) のアルゴリズム. jack-n.c

http://www.math.kobe-u.ac.jp/HOME/taka/2012/tmp/hgm_1.2.tar.gz

```
R CMD install hgm_1.2.tar.gz
```

```
library(hgm)
a<-hgm.cwishart(m=3,n=5,beta=c(1,2,3),x=10)
b<-hgm.cwishart(m=3,n=5,beta=c(1,2,3),x=10,
                method="a-rk4",err=c(0.001,0.0))
```

```
b<-hgm.cwishart(m=4,n=10,beta=c(1,2,3,4),x0=1,x=10,
                approxdeg=20,mode=c(1,1,(16+1)*100));
c<-matrix(b,ncol=16+1,byrow=1);
plot(c)
```

Ropenxm パッケージ.

```
library(Ropenxm)
mat<-matrix(c(1,0,0,0, 1,0,1,0, 1,0,0,1,
              1,1,0,0, 1,1,1,0, 1,1,0,1),
            nrow=4,ncol=6)
mb<-oxm.markov(mat)
mb$markov
```

Todo

OoHG (OpenXM RFC 104, 2004) と httpd-rpc.sh.
OpenXM 100 形式の通信を HTTP GET (または POST) で wrap する方法を定義する. OpenXM/doc/OpenXM-spec/ox-rfc-104.oxw

```
cd OpenXM/src/kan96xx/Doc
openxm ./httpd-rpc.sh
a<-system("hostname",intern=TRUE)
getURL(
  url=paste("http://",a,":8090/?rpc=1-20000",
            sep=""))
```

より詳しくは?

- ① D 加群の積分アルゴリズムと推定理論, 数学セミナー
2012.02, 41-46.
- ② <http://www.math.kobe-u.ac.jp/Movies/oxvh> の元の
2012-05-10-hgd(HGD 概説), 2012-09-05-r-package(R
package なんちゃって作成法),
2012-09-10-wishart(cwishart アルゴリズム, 学会講演),
2013-03-14-r-hgm(R hgm package 開発メモ), “.mov” も.
- ③ <http://www.youtube.com/watch?v=SgATvugfwNM>
KNOPPIX/Math 2011 での R の利用 (入門)