多変数超幾何系と統計, 高山信毅.

推定とは 1.　S : $\mathbf{R}^d$ の適当なクラスの集合 (たとえば semi-algebraic set).

$T_i(t)$: $t \in S$, smooth.

$\theta \in \mathbf{R}^n$ はパラメータ. u を S 上の指数型非正規化確率分布, $Z(\theta)$ を u の正規化定数と呼ぶ.

$$u(\theta, t) = \exp\left(\sum_{i=1}^{n} \theta_i T_i(t)\right), \quad Z(\theta) = \int_S u(\theta, t) dt \qquad (1)$$

$u(\theta, t)/Z(\theta)$ は θ でパラメータづけされた S 上の確率分布を定義する.

例: $A = (a_1, \ldots, a_n)$, $a_i \in \mathbf{N}_0^d$ (列ベクトル).

$T_i(t) = t^{a_i} := \prod_{j=1}^{d} (t_j)^{(a_i)_j}$ とおく *.

例 (一変数の正規分布): $d = 1$, $A = (1, 2)$, $S = \mathbf{R}^2$.

$u(\theta, t) = \exp(\theta_1 t + \theta_2 t^2)$, $\theta_2 < 0$. $Z(\theta) = \dfrac{\sqrt{\pi}}{\sqrt{-\theta_2}\exp(\theta_1^2/(4\theta_2))}$.

---

*t_j で t の j-成分を表す. 同様の記法を混乱の恐れがない限り用いる. a_i は行列 A の i 列.

推定とは 2.　与えられたデータ $t(j) \in S$, $j = 1, \ldots, N$ に対して, 確率の積

$$\prod_{j=1}^{N} u(\theta, t(j))/Z(\theta)$$

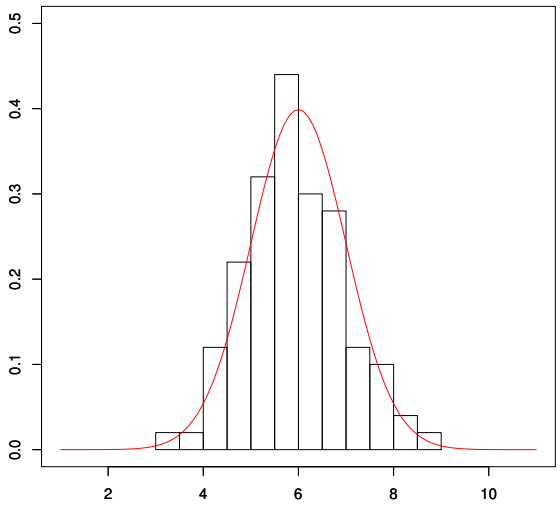
を最大化するパラメータ θ を見つけることを, Fisher の最尤推定という (Fisher's maximal likelihood estimation, MLE).

$$\begin{aligned} & \mathrm{maxarg}_\theta \prod_{j=1}^{N} u(\theta, t(j))/Z(\theta) \\ = \quad & \mathrm{maxarg}_\theta \sum_j \log u(\theta, t(j)) - N \log Z(\theta) \\ = \quad & \mathrm{maxarg}_\theta \sum_i \theta_i \frac{1}{N} \sum_j T_i(t(j)) - \log Z(\theta) \end{aligned} \tag{2}$$

$-\log Z(\theta)$ は上に凸. MLE 方程式:

$$\frac{\partial}{\partial \theta_i} \log Z = \frac{1}{N} \sum_j T_i(t(j)) = \bar{T}_i \tag{3}$$

推定とは 3.



例 (一変数正規分布): $A = (1, 2)$, $\theta_2 = -\frac{1}{2}$ とすると $\frac{1}{N}\sum_j t(j) = \frac{\partial}{\partial\theta_1}\log Z = -\frac{\theta_1}{2\theta_2} = \theta_1$.
$\theta_2 = -1/2$ (既知固定). histogram のような 100 個のデータで $\bar{T}_1 = 6.0$. MLE は $\theta_1 = \bar{T}_1 = \frac{1}{100}\sum_{j=1}^{100} t_1(j) = 6.0$ (平均). $\partial_i \bullet \log Z$ ($\partial_i = \frac{\partial}{\partial\theta_i}$) が簡単な式で書かれてるので MLE 方程式を解くのは簡単.

問題 MLE 方程式を (近似的に) 解け.

MLE 方程式, 解の存在と一意性 1.

Theorem

たとえば Brown の本 [†], Michalek, Sturmfels, Uhler, Zwiernik[‡], Th 2.2, 参照

$C = \{\theta \in \mathbf{R}^n \mid \log Z(\theta) < \infty\}$, $K = \operatorname{conv} T(S)$ とおく [§]. ある条件のもと, $\nabla \log Z$ は C と K の内部の全単射.

$$u(\theta, t) = \exp\left(\sum_{i=1}^{n} \theta_i T_i(t)\right), \quad Z(\theta) = \int_S u(\theta, t) dt$$

[†] L.D.Brown, Fundamentals of Statistical Exponential Families with Applicatioins in Statistical Decision Theory, 1986, Th 3.6

[‡] Exponential Varieties, 2016

[§] $T(t) = (T_1(t), \ldots, T_d(t))$

MLE 方程式, 解の存在と一意性 2.

離散版の Z とその MLE 問題の例. $A: d \times n$ 行列. 成分は 0 以上の整数. $\operatorname{rank} A = d$ で $(1, 1, \ldots, 1)$ なる形の行を含む. $\beta \in \mathbf{N}_0^d$ を固定. $p_i = \exp(\theta_i)$: パラメータ. $\frac{p^c}{c!}$ は有限集合 $S = \{c \in \mathbf{N}_0^n \mid Ac = \beta\}$ の上の非正規化確率分布.

$$Z(\beta; p) = \sum_{Ac=\beta, c \in \mathbf{N}_0^n} \frac{p^c}{c!} = \sum_{Ac=\beta, c \in \mathbf{N}_0^n} \frac{\exp(\sum_{i=1}^n \theta_i c_i)}{c!} \tag{4}$$

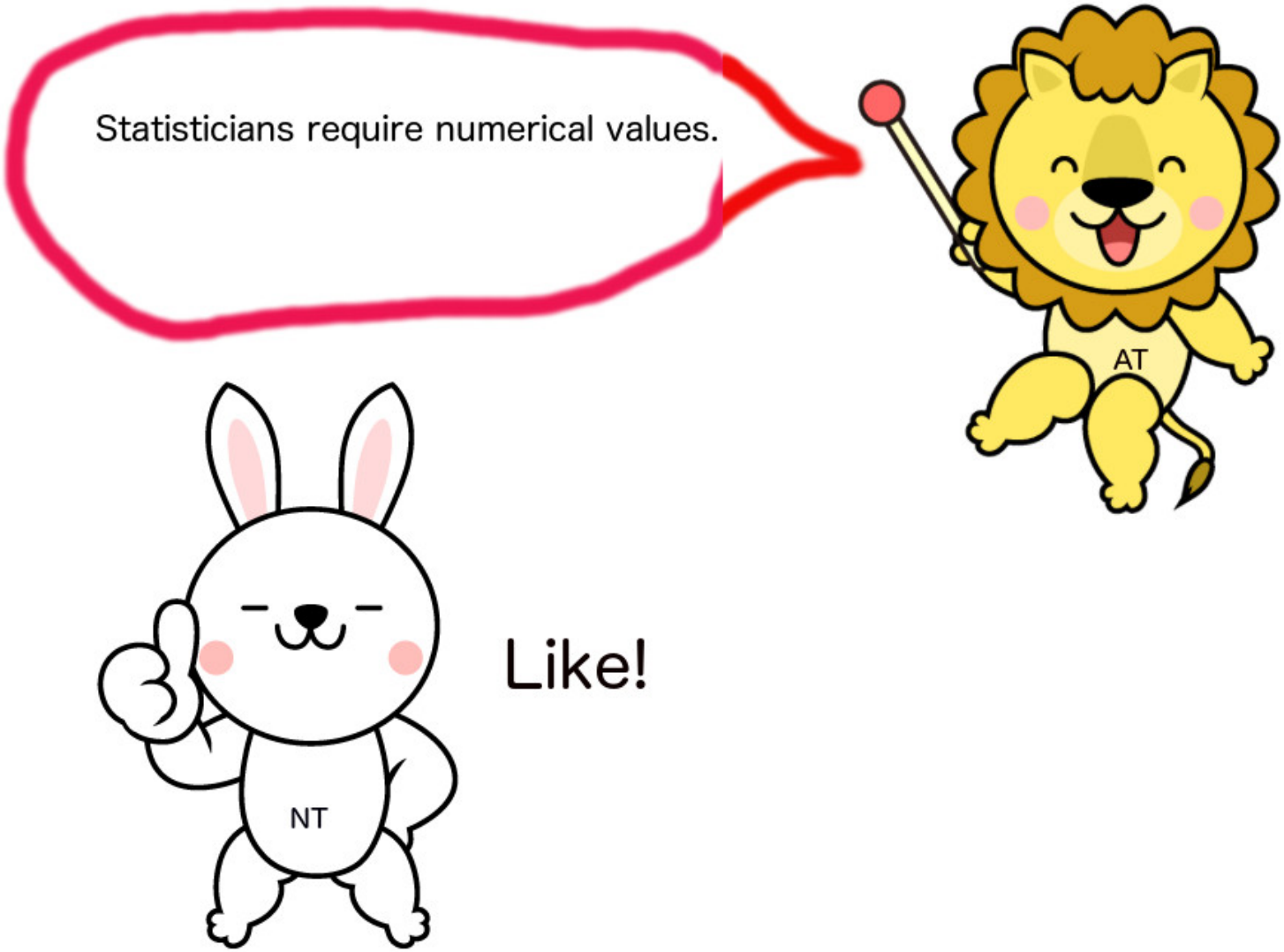
はその正規化定数.

Theorem (T, Kuriki, Takemura (2018)[¶])

Newton 多面体 $\operatorname{New}(Z)$ *(*Z は p の多項式とみなす*)* の次元が $n - d$ なら $\nabla \log Z$ は $\mathbf{R}^n / \operatorname{Im} A^T$ と $\operatorname{relint}(\operatorname{New}(Z))$ の全単射.

[¶] A-Hpergeometric Distributions and Newton Polytopes

MLE 方程式を数値的に解く



MLE 方程式を解きたい. (2) より最大値最小値問題を解く ⇒ MLE. gradient descent には Z, $\frac{\partial Z}{\partial \theta_i}$ の数値, できたら 2 階微分の数値が必要.

1. Z をモンテカルロ法などで数値積分する. ただしこの方法では高精度の計算は期待できない. 積分変数の次元が低い場合はさまざまな数値積分アルゴリズムを適用できる.
2. Z の級数展開を求める. 一変数ではこの方法およびそのバリエーションで高精度かつ高速な計算が可能. しかし多変数では計算量が膨大. 級数展開を求める問題, また大域的計算をおこなうには接続公式が有用.
3. Z の近似を用いる. 鞍点法など.

約 10 年ほど前: holonomic gradient method (HGM) および holonomic gradient descent (HGD) を提案. 日比編, グレブナー道場 (2012), 文献表 は openxm hgm で検索.

HGM(holonomic gradient method) とは?　HGM は非正規化確率分布関数や S に台を持つ Heaviside 関数が holonomic 系を生成する場合に対する方法.

HGM の 3 段とび

1. 正規化定数の満たす holonomic 系を計算する. 後述の積分アルゴリズムを用いるか理論的に導出 (素手での導出) *. この方程式系を Pfaffian (Pfaff 方程式) に変形.
2. Pfaffian を数値的に解くための初期値ベクトルを計算. これは正規化定数とその微分をある点で数値評価することにほかならない. うまい点を見つけることが必要でここは素手.
3. 常微分方程式の数値解法アルゴリズムを用いて上記の初期値問題を解く.

*微分作用素環での Gröbner 基底を用いたアルゴリズムがある.

Example of HGM $Z(\theta) = \int_{-\infty}^{\infty} \exp(\theta_1 t + \theta_2 t^2) dt$.

1. Z is annihilated by

$$\theta_1 \partial_1 + 2\theta_2 \partial_2 + 1, \partial_1^2 - \partial_2$$

Put $F = (Z, \partial_2 \bullet Z)^T$.

$$\frac{\partial F}{\partial \theta_1} = PF, \quad \frac{\partial F}{\partial \theta_2} = QF$$

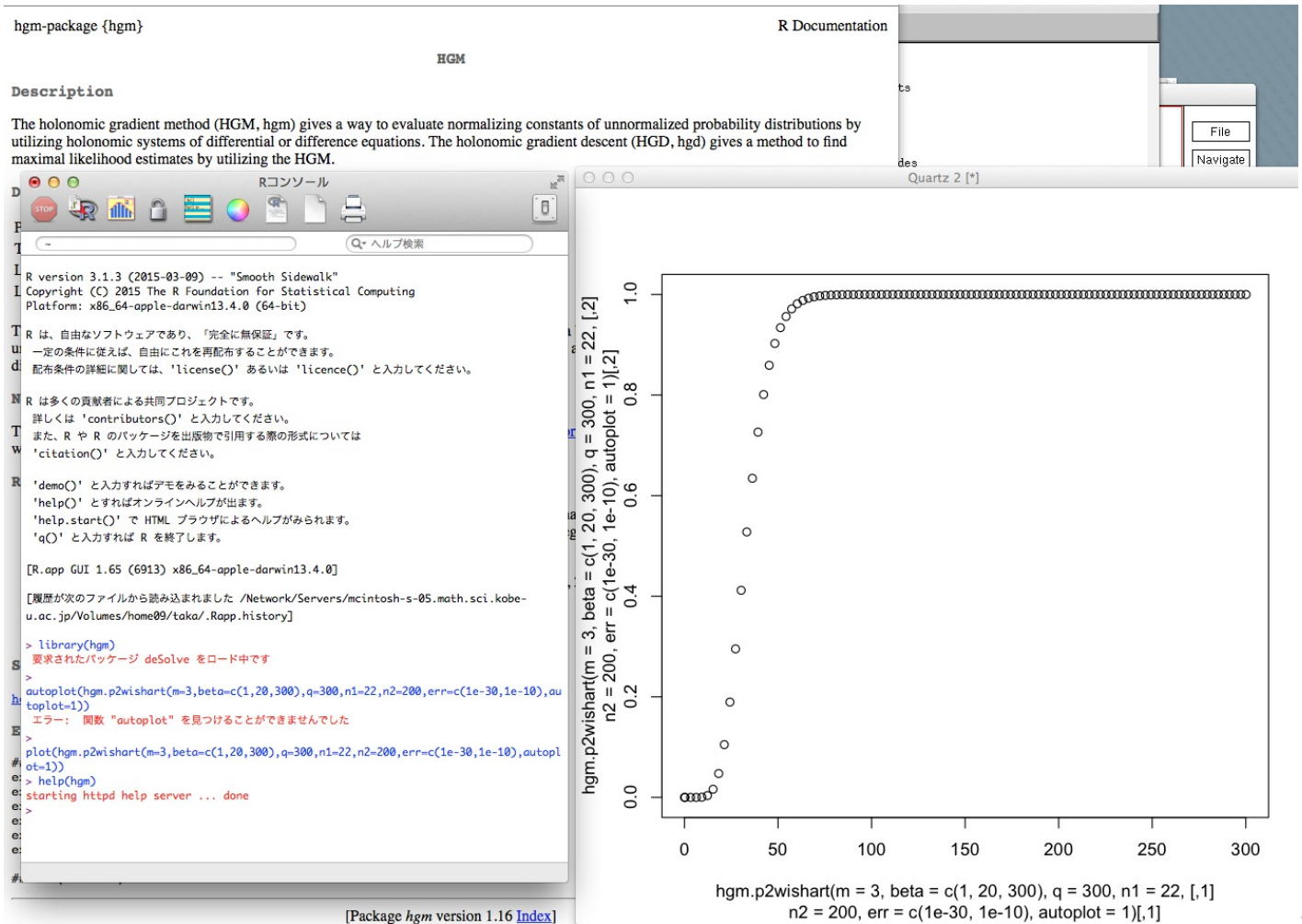
,

$$P = \begin{pmatrix} \frac{-1}{\theta_1} & \frac{-2\theta_2}{\theta_1} \\ \frac{1}{\theta_1\theta_2} & \frac{2\theta_2 - 1/2\theta_1^2}{\theta_1\theta_2} \end{pmatrix}, \; Q = \begin{pmatrix} 0 & 1 \\ \frac{\frac{-1}{2}}{\theta_2^2} & \frac{-5/2\theta_2 + 1/4\theta_1^2}{\theta_2^2} \end{pmatrix}$$

2. $Z(0, -1/2) = \sqrt{\pi}$.
3. $F(\theta_1 + h_1, \theta_2 + h_2) \sim F(\theta) + h_1 P(\theta) F(\theta) + h_2 Q(\theta) F(\theta)$, h_i's are small numbers (1st order Runge-Kutta method).

HGD, HGM を提案した文献

1. H.Nakayama, K.Nishiyama, M.Noro, K.Ohara, T.Sei, N.Takayama, A.Takemura , Holonomic Gradient Descent and its Application to Fisher-Bingham Integral, Advances in Applied Mathematics 47 (2011), 639–658.
2. H. Hashiguchi, Y. Numata, N. Takayama, A. Takemura, Holonomic gradient method for the distribution function of the largest root of a Wishart matrix, Journal of Multivariate Analysis, 117 (2013), 296–312.



S^1 上の退化した Fisher-Bingham 分布: Von Mises 分布.

$$
\begin{aligned}
Z(\theta) &= \int_0^{2\pi} \exp(\theta_1 \cos u + \theta_2 \sin u) du \\
&= \int_{\mathbf{R}^2} \exp(\theta_1 t_1 + \theta_2 t_2) \delta_{S^1}(t)
\end{aligned}
$$

$t_1 = \cos u$, $t_2 = \sin u$: S^1 のパラメータ表示. δ_{S^1} : 円周に support を持つ δ measure.

$$
\frac{\exp(\theta_1 \cos u + \theta_2 \sin u)}{Z(\theta)}
$$

MLE デモ 1: winf2.rr, `demo1();`

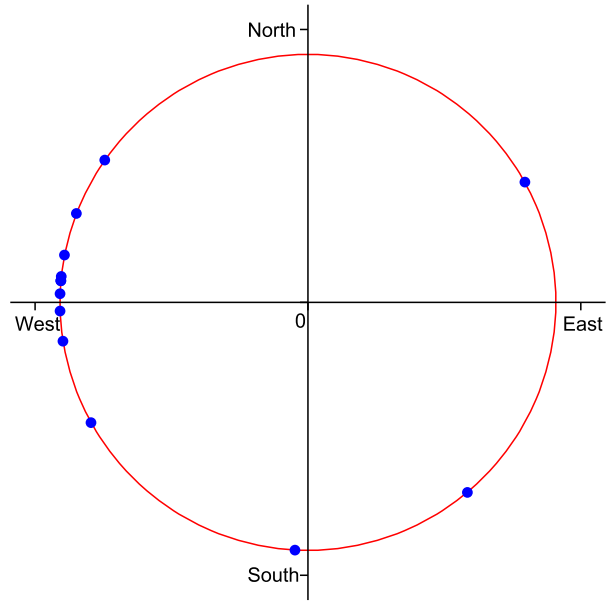


Figure: 札幌上空約 1 万メートル地点における午前 9 時の風向きの 13 日間のデータ (2011 年 1 月 1 日から 14 日 (11 日は欠損): 気象庁の気象統計情報から)

二元分割表のパラメータ推定

1	1	2	3	7
1	3	1	1	6
2	4	3	4	

$=: U$ (黒の部分)

行和, 列和一定なものが事象の全体 S. A で書くと

$S = \{u \in \mathbf{N}_0^8 \,|\, Au = \beta = (6,2,4,3,4)^T\}$,

$$A = \begin{pmatrix} 0 & 0 & 0 & 0 & 1 & 1 & 1 & 1 \\ 1 & 0 & 0 & 0 & 1 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 1 & 0 & 0 & 0 & 1 \end{pmatrix}$$

$$Z(\beta; p) = \sum_{u \in S} \frac{p^u}{u!} = \sum_{u \in S} \frac{\exp\left(\sum_{i=1}^8 \theta_i u_i\right)}{u!}, \quad p_i = \exp(\theta_i)$$

$$\mathrm{maxarg}_\theta \, U \cdot \theta - \log Z(\beta; p)$$

二元分割表の Z = 青本,Gel'fand 超幾何多項式

Y. Goto, K. Matsumoto, Pfaffian equations and contiguity relations of the hypergeometric function of type$(k+1, k+n+2)$ and their applications, Funkcialaj Ekvacioj 61 (2018), 315–347.

$$F(\beta; p) = \begin{pmatrix} Z(\beta; p) \\ \cdot \\ \cdot \\ \cdot \end{pmatrix}, \quad \frac{\partial F}{\partial p_i} = Q_i F, \quad F(\beta + \mathbf{e}_i; p) = R_i F(\beta; p)$$

差分 HGM: 差分方程式で F の数値を求める.
MLE デモ 2:

```
import("gtt_ekn3.rr");
gtt_ekn3.cmle([[1,1,2,3],[1,3,1,1]]);
```

検定 (test)

$P \subset S$ として,

$$\int_P u(\theta, t)dt/Z(\theta)$$

を計算する問題.

例. 事象がどの程度まれであるか? 確率を計算することが必要となる. 例をひとつあげよう. 次のような仮定を考える. K 大学の理学部の 4 回生 100 名の中で “教室の温度を快適と感じる人の割合は 25 度を平均として標準偏差が 3 である正規分布に従っている”. さて, 20 度に暖房設定した教室でこの 100 人の学生が講義を受けている. 講義を終了したあと, 目があった一人の学生に教室の温度が快適かどうか聞いたら “とても快適!” と答えた. さて上の仮定が正しいとすると 20 度以下を快適と答える人は全体の 5 パーセント程度の 5 人のはずである. なぜなら

$$\int_{-\infty}^{20} N(25, 3; t)dt \sim 0.05, \quad N(m, s; t) = e^{-\frac{(t-m)^2}{2s^2}}/Z, Z = \sqrt{2\pi s^2}$$

一人ランダムに選んだ人が快適と答えた出来事は 5 パーセント以下の確率でおきる “まれ” な出来事ということになる.

Test and A-hypergeometric system

Let $T_i(t) = t^{a_i}$ and put

$$I(P;\theta) = \int_P \exp\left(\sum_{i=1}^n \theta_i T_i(t) + (-\beta - 1)\log t\right) dt.$$

Theorem (Nishiyama-T, 2010*)

When P is a polytope, the function $I(P;\theta)$ satisfies the following incomplete A-hypergeometric system.

We call the following system of differential equations $H_A(\beta, g)$ an *incomplete $\mathcal{A}$-hypergeometric system*:

$$\left(\sum_{j=1}^n a_{ij}\theta_j\partial_j - \beta_i\right) \bullet f = g_i, \qquad (i = 1, \ldots, d)$$

$$\left(\prod_{i=1}^n \partial_i^{u_i} - \prod_{j=1}^n \partial_j^{v_j}\right) \bullet f = 0$$

with $u, v \in \mathbf{N}_0^n$ running over all u, v such that $Au = Av$).

Here, $\mathbf{N}_0 = \{0, 1, 2, \ldots\}$, and $\beta = (\beta_1, \ldots, \beta_d) \in \mathbf{C}^d$ are parameters and $g = (g_1, \ldots, g_d)$ where g_i are given holonomic functions.

*Incomplete $\mathcal{A}$-Hypergeometric Systems

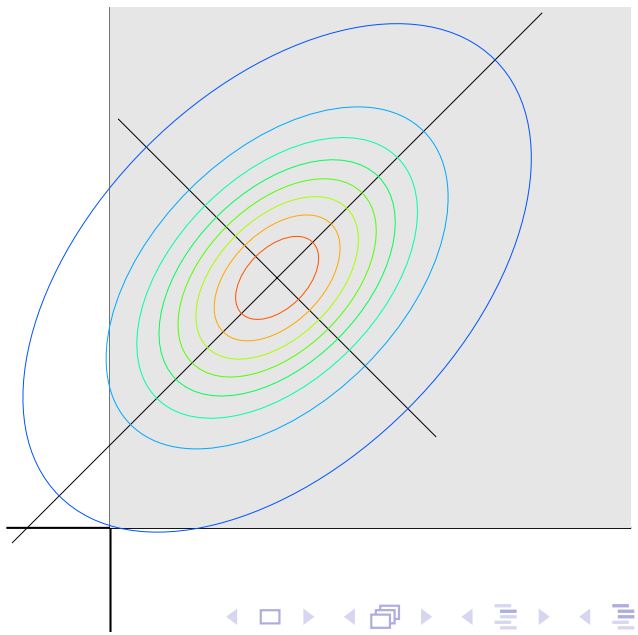
$\int_P \exp(\sum_i \theta_i t^{a_i}) dt$ は不完全 A-超幾何系をみたす. しかし一般的すぎて解析が困難. 統計的に重要でかつ実際の計算もできる場合: 小山, 竹村 (2012) や小山 (2015) による結果と実装 `http://www.github.com/tkoyama-may10`.
a_{ij}, b_j で定義される次のような polyhedron.

$$P = \{x \in \mathbf{R}^d : \sum_{i=1}^{d} a_{ij} x_i + b_j \geq 0,\ 1 \leq j \leq n\}, \tag{1}$$

$$\mathrm{Prob}(x \in P) = \frac{1}{(2\pi)^{d/2}} \int_{x \in P} \exp\left(-\frac{1}{2} \sum_{i=1}^{d} x_i^2\right) dx_1 \cdots dx_d. \tag{2}$$

Orthant probability by R
`https://cran.r-project.org/web/packages/hgm/index.html`

```
library(hgm);
hgm.ncorthant(
 matrix(c(2,1,1,2),nrow=2),
 y=c(1,1.5));
[1] 0.6930033
```



F_j を P の facet で $\sum_{i=1}^{d} a_{ij}x_i + b_j = 0$ に対応するものとする. Facet の交わりに対応する index を集めたものを $\mathcal{F}$ と置く. つまり,
$\mathcal{F} = \{J \subset \{1, \ldots, n\} \mid F_J \neq \emptyset\}$, $F_J = \cap_{j \in J} F_j$ ただし $\mathcal{F}$ は空集合も含むものとし, これは P に対応する. Pfaffian の基底は $\{g^J, J \in \mathcal{F}\}$ と $\mathcal{F}$ で index づけできる. Pfaffian は以下のようになる.

Theorem (小山, Holonomic Modules Associated with Multivariate Normal Probabilities of Polyhedra)

$$
\begin{aligned}
\partial_{a_{ij}} g^J &= \sum_{k=1}^{n} a_{ik} \partial_{b_k} \partial_{b_j} g^J \quad (1 \leq i \leq d,\ 1 \leq j \leq n,\ J \in \mathcal{F}), \\
\partial_{b_j} g^J &= g^{J \cup \{j\}} \quad (j \in J^c,\ J \in \mathcal{F}), \\
\partial_{b_j} g^J &= -\sum_{k \in J} \alpha_J^{jk}(a) \left(b_k g^J + \sum_{\ell \in J^c} \alpha_{k\ell}(a) g^{J \cup \{\ell\}} \right) \quad (j \in J,\ J \in \mathcal{F}).
\end{aligned}
$$

ここで $\alpha_F(a)$ は ij 成分が $\sum_{k=1}^{d} a_{ki}a_{kj}$ である行列. $\alpha_F(a)$ の逆行列の ij 成分を $(\alpha_F^{ij}(a))_{i,j \in F}$ とおく.

推定, 検定以外の HGM の応用: exact sampler　間野: Partitions, Hypergeometric Systems, and Dirichlet Processes in Statistics, JSS Research Series in Statistics, 2018.

Holonomic distribution と計算代数

D_n を多項式係数の微分作用素のなす環. 有理式を係数とする微分作用素環 $R_n = \mathbf{C}(x)\langle\partial_1, \ldots, \partial_n\rangle$. $L = \sum_{(\alpha,\beta)\in E} a_{\alpha,\beta} x^\alpha \partial^\beta \in D_n$.

$$\mathrm{ord}_{(u,v)}(L) = \max_{(\alpha,\beta)\in E} (u\alpha + v\beta) \tag{3}$$

$$\mathrm{in}_{(u,v)}(L) = \sum_{\mathrm{ord}(x^\alpha\partial^\beta)=\mathrm{ord}(L),(\alpha,\beta)\in E} a_{\alpha,\beta} x^\alpha \xi^\beta \in \mathbf{C}[x,\xi] \tag{4}$$

ここで $u = (1, \ldots, 1)$ and $v = (1, \ldots, 1)$.
D_n の左イデアル I に対して $\mathrm{in}_{(u,v)}(I) = \langle \mathrm{in}_{(u,v)}(L) \,|\, L \in I\rangle$ を I の (u, v)-initial ideal と呼ぶ I は $\mathrm{in}_{(u,v)}(I)$ の (Krull) 次元が n であるとき holonomic イデアルとよぶ. $\dim_{\mathbf{C}(x)} R_n/R_n I$ を I の holonomic rank と定義する. holonomic イデアルに対する holonomic rank は必ず有限となり, (generic な点における) 解空間の次元に等しい.
$\mathbf{R}^n$ 上の Schwartz distribution f が holonomic ideal で零化される (holonomic ideal を連立線形偏微分方程式として見た時の解の) とき, f を holonomic distribution とよぶ. D-加群の理論より次の事実がなりたつ.

Theorem

$f(x_1, \ldots, x_n)$ が *holonomic distribution* の時, ある条件の元,
$g(x') = \int_{\mathbf{R}^{n-m}} f(x) dx_{m+1} \cdots dx_n$ も m-変数 x' の *holonomic distribution* となる.

$g(x')$ が満たす微分方程式系を $f(x)$ の満たす微分方程式系から求めるアルゴリズムは積分アルゴリズムと呼ばれている.

主な積分アルゴリズム

D_n の左イデアル I と右イデアル $\partial_1 D_n + \cdots + \partial_m D_n$ に対して

$$(I + \partial_1 D_n + \cdots + \partial_m D_n)) \cap \mathbf{C}\langle x_{m+1}, \ldots, x_n, \partial_{m+1}, \ldots, \partial_n \rangle$$

なる消去計算.

1. Creative telescoping (D.Zeilberger, 1980's – 2010's, M.Petkovsek, H.Wilf, D.Zeilberger, $A = B$, 1996),
2. D-加群の積分 (大阿久, algorithms for b-functions, restrictions, and algebraic local cohomology groups, 1997 など),
3. F.Chyzak's heuristics (2000's), C.Kouchan's heuristics (2010's) は Mathematica のパッケージ HolonomicFunctions.m `http://www.risc.jku.at/publications/download/risc_3934/hf.pdf` に実装されている.
4. Risa/Asir, Macaulay 2, Singular, などの計算代数システムの D-module パッケージ,
5. semi-algebraic set の Heaviside 関数の満たす holonomic 系 (大阿久, Algorithms for Integrals of Holonomic Functions over Domains defined by Polynomial Inequalities 2013 など),

ランダム行列と HGM

x が Polyhedron P の上にある確率 $P(x \in P)$ を計算する問題を紹介した. ランダム行列 W に対してその最大固有値が x 以上となる確率 $P(\lambda_1(W) > x)$ を求める問題も積分の計算に帰着する場合がある. A を $m \times n$ 実 random 行列. $g \in S^{m-1}$, $h \in S^{n-1}$ とする. g でパラメータ付けした $SO(m)$ の元を $(g, G(g))$, h でパラメータ付けした $SO(n)$ の元を $(h, H(h))$ として, $\sigma = g^T Ah$, $B = G^T AH$ と A を変数変換する. この時 $M_x = \{hg^T \mid g^T Ah \geq x\}$ の Euler 標数の期待値は次のような積分表示を持つ.

Theorem (T, L.Jiu, S.Kuriki, Y.Zhang, 2020)

行列 A は分布 $p(A)$ に従うとする. ある条件のもと, $\mathrm{E}[\chi(M_x)]$ は

$$\frac{1}{2}\int_x^\infty \sigma^{n-m} d\sigma \int_{\mathbf{R}^{(m-1)(n-1)}} dB \int_{S^{m-1}} G^\top dg \int_{S^{n-1}} H^\top dh \, \det(\sigma^2 I_{m-1} - BB^\top) p(A). \tag{5}$$

に等しい. ここで $G^\top dg = \wedge_{i=1}^{m-1} G_i^\top dg$, $H^\top dh = \wedge_{i=1}^{n-1} H_i^\top dh$, であり, G_i と H_i はそれぞれ G と H の i 番目の列ベクトルである. また $dg = (dg_1, \ldots, dg_m)^\top$, $dh = (dh_1, \ldots, dh_n)^\top$.

Euler 標数法の一般論 (Adler-Taylor 2007, 栗木-竹村 2009 参照) によるとこれは AA^T の最大固有値の分布を $x \to +\infty$ で近似している.

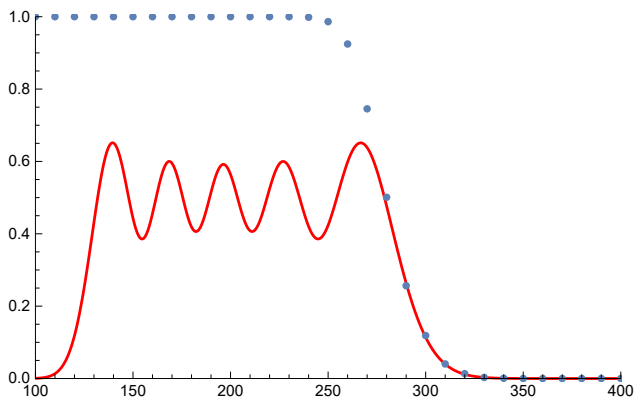
$$p(A)dA = \frac{1}{(2\pi)^{mn/2}\det(\Sigma)^{n/2}} \exp\Big\{-\frac{1}{2}\mathrm{Tr}\,(A-M)^\top \Sigma^{-1}(A-M)\Big\}dA. \tag{6}$$

なる場合. A の最大特異値は non-central Wishart matrix $W_m(n, \Sigma, \Sigma^{-1}MM^\top)$ の最大固有値の平方根.
理論的な結果: $M = 0$ となる central かつ共分散行列がスカラー行列 I_m/s となる場合 (cf. 青本, 金子: Selberg 型積分)

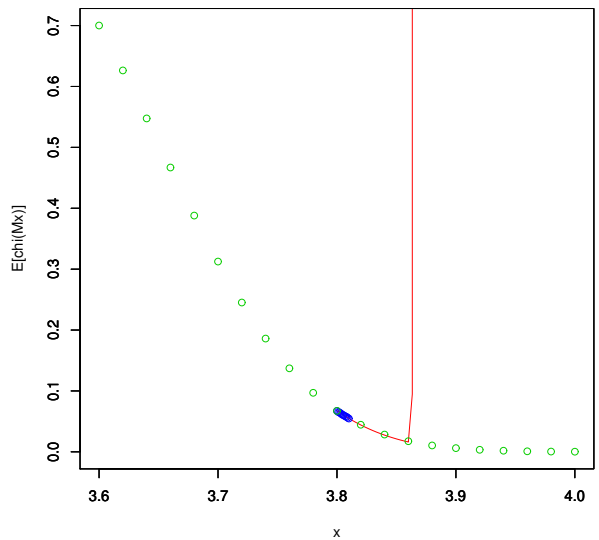
Theorem (T, L.Jiu, S.Kuriki, Y.Zhang, 2020)

$$\mathrm{E}[\chi(M_x(s))] = \prod_{i=1}^{5} c_i \int_x^{+\infty} \exp\left(-\frac{s}{2}\sigma^2\right)\sigma^{n-m}{}_1F_1(-(m-1), 1+n-m; s\sigma^2)d\sigma,$$

ここで c_1, c_2, c_3, c_4, c_5 は Γ 関数を用いて書ける定数.
計算による結果: 最新の積分アルゴリズムを用いると $n = m = 2$ で x についての 11 階の線形常微分方程式.



$m = 10,\ n = 200,\ M = 0,\ \Sigma = I_m.$



$$m = n = 2,\ M = \begin{pmatrix} 1 & 3 \\ 2 & 0 \end{pmatrix},$$
$$\Sigma = \mathrm{diag}(10^{-3}, 10^{-2}).$$