ホロノミック系の数値解析の統計への応用

高山信毅 (神戸大理)

2014.12.21

$p \times q$ (2元) 分割表 ($p \times q$ (2 way) contingency table)

$p \times q$ 行列 $u = (u_{ij})$, $u_{ij} \in \mathbf{N}_0$.

例: $u = \begin{pmatrix} 4 & 7 \\ 32 & 5 \end{pmatrix}$

	アセトアミノフェン	ジクロフェナクナトリウム
死亡	4	7
生存	32	5

(i,j) の実現確率 θ_{ij}, $\sum \theta_{ij} = 1$. $|u| = \sum u_{ij}$ とおく.

$|u|$ 一定の時

$$u \text{ が起きる確率} = \frac{|u|}{u!}\theta^u$$

ここで $\theta^u = \prod \theta_{ij}^{u_{ij}}$, $u! = \prod u_{ij}!$.

行和列和一定の条件付き確率

	アセトアミノフェン	ジクロフェナクナトリウム	
死亡	4	7	$\beta^1 = 11$
生存	32	5	$\beta^2 = 37$
	$\beta_1 = 36$	$\beta_2 = 12$	

行和 β^i, 列和 β_j が一定の時,

$$u \text{ が起きる確率} = \frac{|u|}{u!}\theta^u / Z(\beta, \theta)$$

$$Z(\beta, \theta) = \sum_{i \text{ 行の和}=\beta^i,\ j \text{ 列の和}=\beta_j \text{ となる } u} \frac{|u|}{u!}\theta^u$$

$$u = \begin{pmatrix} 11 & 0 \\ 25 & 12 \end{pmatrix}, \dots, u = \begin{pmatrix} 4 & 7 \\ 32 & 5 \end{pmatrix}, \dots, u = \begin{pmatrix} 0 & 11 \\ 36 & 1 \end{pmatrix}.$$

i 行の和$=\beta^i$, j 列の和$=\beta_j$ となる u の満たす条件.

$$\sum_{j=1}^{q} u_{ij} = \beta^i, i = 1, \dots, p, \ \sum_{i=1}^{p} u_{ij} = \beta_j, j = 1, \dots, q$$

A 分布へ

$$2 \times 2 \text{ の場合} \quad A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}, \; A \begin{pmatrix} u_{11} \\ u_{12} \\ u_{21} \\ u_{22} \end{pmatrix} = \begin{pmatrix} \beta^2 \\ \beta_1 \\ \beta_2 \end{pmatrix}.$$

例. $\beta = (37, 36, 12)^T$.

$A = (a_{ij})$ を $d \times n$ 行列. $a_{ij} \in \mathbf{N}_0$. $\mathrm{rank}\, A = d$. A の行は $(1, 1, \ldots, 1)$ を含む. $\beta \in \mathbf{N}_0^d$. $u \in \mathbf{N}_0^n$, $\theta \in \mathbf{R}_0^n$.

$$Z(\beta, \theta) = \sum_{Au=\beta, u \in \mathbf{N}_0^n} \frac{\theta^u}{u!}$$

(β 条件付き) A-分布の正規化定数 (normalizing constant), A-超幾何多項式.

$Au = \beta$ なる条件のもとで u がおきる確率は $\frac{\theta^u}{u!} / Z(\beta, \theta)$.

以後 $u_1 = u_{11}, u_2 = u_{12}, u_3 = u_{21}, u_4 = u_{22}$, $\theta_1 = \theta_{11}, \theta_2 = \theta_{12}, \theta_3 = \theta_{21}, \theta_4 = u_{22}$,

$\beta_1 = \beta^2$, $\beta_2 =$ (分割表の記号の β_1), $\beta_3 =$ (分割表の記号の β_2) 記号法は状況に応じて使い分ける.

Fisher's MLE

正規化定数 Z の計算は統計の基本問題.
問題: "データ u から θ を推定せよ". Fisher の最尤推定 (Fisher's maximal likelihood estimation, MLE) は, $\frac{\theta^u}{u!}/Z(\beta, \theta)$ を最大化する θ を θ の推定値とする.

命題 (情報幾何の基礎)

上記の確率に従う random variable U_i の期待値 $E[U_i] =: E_i$ (これは θ の関数) は $\theta_i \frac{\partial Z}{\partial \theta_i}/Z$ に等しい.

証. U_i の期待値 の定義は $\sum_{Au=\beta, u \in \mathbf{N}_0^n} u_i(u$ が起きる確率).
$\partial_i = \partial/\partial\theta_i$ とおく.

$$E_i = \sum u_i \frac{\theta^u/u!}{Z} = \frac{1}{Z} \sum \theta_i \partial_i \bullet \theta^u/u! = \frac{1}{Z} \theta_i \partial_i \bullet \sum \theta^u/u!$$

期待値との関係: 期待値=観測データ u_i

命題 (情報幾何の基礎)

θ が *Fisher's MLE* ならば θ は $E_i(\theta) = \theta_i \frac{\partial Z}{\partial \theta_i} / Z u_i = u_i$ を満たす.

証. 最大化したい関数 $\frac{\theta^u}{u!}/Z(\beta,\theta)$ の対数をとると
$\sum u_i \log \theta_i - \log u! - \log Z$ これを F とおこう. 最大値をとる θ で, $\frac{\partial F}{\partial \theta_i} = \frac{u_i}{\theta_i} - \frac{\partial Z/\partial \theta_i}{Z} = 0$. よって, $u_i = \theta_i \frac{\partial Z/\partial \theta_i}{Z}$.
問題: 推定値は一意的に存在するか?
期待値写像

$$E : \mathbf{R}^n_{>0} \ni \theta \mapsto (E_i(\theta)) \in \mathbf{R}^n$$

a_i を行列 A の i 番目の列ベクトルとする.

命題

$E_i(\theta) = E_i(\theta_1 t^{a_1}, \ldots, \theta_n t^{a_n})$ が 任意の $t \in \mathbf{R}^d_{>0}$ に対して成り立つ.

$$\tilde{E} : \mathbf{R}^{n-d}_{>0} \simeq \mathbf{R}^n_{>0}/(A \text{ の作用}) \ni \theta \mapsto E(\theta) \subset P_\beta$$

像の入る多面体 P_β の定義

$$P_\beta = \{ e \in \mathbf{R}^n \mid Ae = \beta, e_i > 0 \}$$

例:

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

$$E_1 = 11 - x \frac{dF}{dx} / F, F(x) = {}_2F_1 \left(-12, -11, 26; x \right), x = \frac{\theta_1 \theta_4}{\theta_2 \theta_3}.$$

$$0 < e_1 < 11$$

e_2, e_3, e_4 は $Ae = \beta$ を用いて e_1 より決まる. つまり
$e_3 = \beta_2 - e_1, e_4 = \beta_1 - \beta_2 + e_1, e_2 = \beta_3 - \beta_1 + \beta_2 - e_1$

期待値多面体定理

Theorem ([T, 2014])

$\dim \mathrm{conv}\,\{e \in \mathbf{N}_0^n \mid Ae = \beta\} = n - d$ の時 $\tilde{E}$ は同型．さらに $u \in P_\beta$ のとき，Fisher's MLE $\theta \in \mathbf{R}^n_{>0}$ は A の作用で modulo して一意的に存在．

例．2×3 分割表．

	アセトアミノフェン	ジクロフェナクナトリウム	メフェナム酸
死亡	4	7	2
生存	32	5	6

$\beta = (42, 36, 12, 8)$,

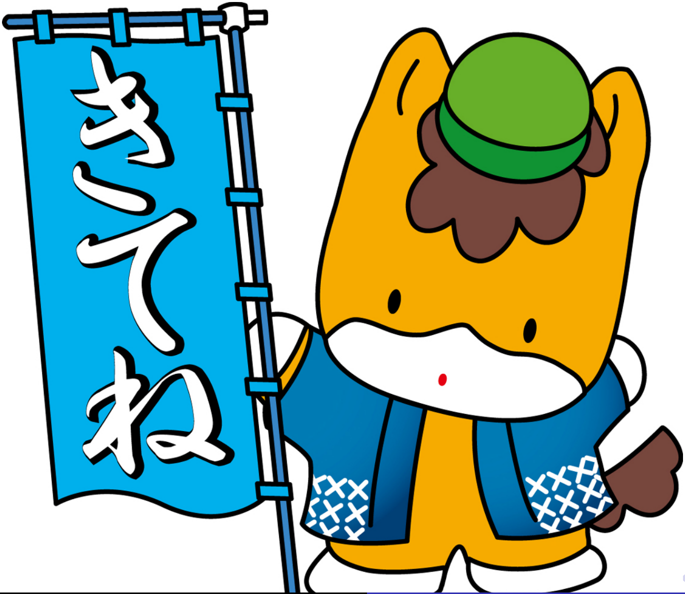
$$A = \begin{pmatrix} 0 & 0 & 0 & 1 & 1 & 1 \\ 1 & 0 & 0 & 1 & 0 & 0 \\ 0 & 1 & 0 & 0 & 1 & 0 \\ 0 & 0 & 1 & 0 & 0 & 1 \end{pmatrix}.$$

$F_D\left(-13, (-12, -8), 24; \frac{\theta_{12}\theta_{21}}{\theta_{11}\theta_{22}}, \frac{\theta_{13}\theta_{21}}{\theta_{11}\theta_{23}}\right)$ を用いて，$\tilde{E}$ が書ける．

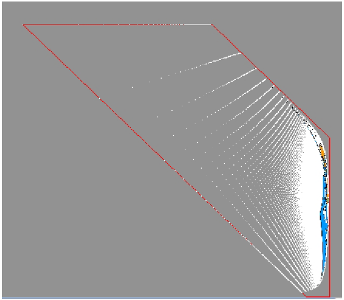
$$\tilde{E} \ : \ \mathbf{R}^2_{>0} \ni (x_1, x_2) \mapsto (E_{21}, E_{22}) \in P_\beta \cap (\mathbf{R}^6 \text{ 内の } \mathbf{R}^2)$$

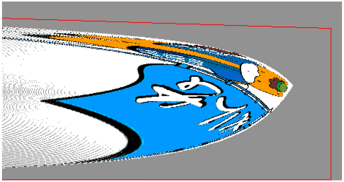
$e_{21} < 36,\ e_{22} < 12,\ e_{21} + e_{22} > 35,\ e_{21} + e_{22} < 43$

期待値写像 $\tilde{E}$



期待値写像 $\tilde{E}$ によるグンマちゃんの像





拡大図.

計算

存在と一意性はわかった．では計算は? Fisher's MLE の計算のため最適化アルゴリズムを適用するには $Z(\beta, \theta)$ やその偏微分の計算が必要．

例: x の推定値は 10.4. (条件付きでないと, 11.2). $x = \frac{\binom{A\ 生}{A\ 死}}{\binom{J\ 生}{J\ 死}}$, $x = \frac{\theta_1 \theta_4}{\theta_2 \theta_3}$.

$Z(\beta, \theta)$ やその偏微分の計算手法．

1. 多項式だから状態を全部列挙して計算．Fisher の正確法．
2. Holonomic gradient method (HGM). 正規化定数の満す holonomic 系を利用して数値計算 [新].
3. Markov chain Monte-Carlo method (MCMC). 分布に従う乱数をメトロポリスアルゴリズム等で生成して近似計算．

A (超幾何) 分布の (差分) HGM による計算

$Z(\beta,\theta)$ は holonomic な微分差分方程式系 H_A:

$$\partial_i - \prod_{k=1}^{d} S_k^{a_{ki}}, \ \sum_{j=1}^{n} a_{ij}\theta_j\partial_j - \beta_i$$

を満す．ここで $S_i : \beta_i \mapsto \beta_i - 1$, $\partial_i = \frac{\partial}{\partial\theta_i}$

Theorem (小原-高山, 2009)

$$\mathrm{rank}\, H_A = A \text{ の正規化体積 } =: r$$

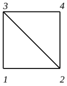
r が HGM の計算量を定める．例:

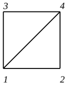
$$A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 0 & 1 & 0 & 1 \end{pmatrix}$$

縦に見る $\partial_1 - S_2, \partial_2 - S_3, \partial_3 - S_1S_2, \partial_4 - S_1S_3$

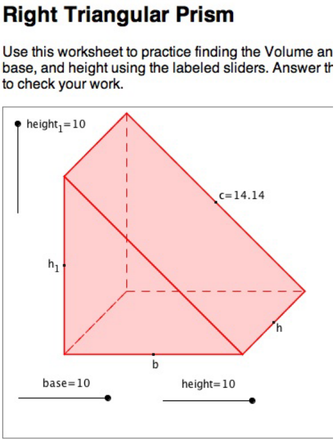
横に見る $\theta_3\partial_3 + \theta_4\partial_4 - \beta_1, \theta_1\partial_1 + \theta_3\partial_3 - \beta_2, \theta_2\partial_2 + \theta_4\partial_4 - \beta_3$

2 行目を 3 行目に足すと, $A = \begin{pmatrix} 0 & 0 & 1 & 1 \\ 1 & 0 & 1 & 0 \\ 1 & 1 & 1 & 1 \end{pmatrix}$. 2 次元の配置として図を書くと正方形．vol $(A) = 2$.





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



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

Pfaffian: 高階 ODE を連立一階 ODE へ, の多変数化

$\delta_i = \partial_i$ $i = 1, \ldots, n$, $\delta_{n+k} = S_k$, $k = 1, \ldots, d$ とおく.
$\alpha^1 = \mathbf{0} \in \mathbf{N}_0^{n+d}$, $\alpha^i \in \mathbf{N}_0^{n+d}$, $i = 2, \ldots, r$ が存在して,

$$Y = \begin{pmatrix} \delta^{\alpha^1} \bullet Z \\ \delta^{\alpha^2} \bullet Z \\ \vdots \\ \delta^{\alpha^r} \bullet Z \end{pmatrix}$$

とおくと,

$$S_i \bullet Y = Q_i Y \quad \text{差分 Pfaffian}$$

$$\partial_i \bullet Y = R_i Y \quad \text{微分 Pfaffian}$$

となる $r \times r$ 行列 $Q_i(\beta, \theta)$, $R_i(\beta, \theta)$ が存在する. この方程式を Pfaffian とよぶ.
注: H_A から Pfaffian への変換はグレブナー基底で原理的には可能. yang.rr など.

例, 2×2 分割表

例: $x_i = \theta_i$. $D = x_1x_4 - x_2x_3$, $B = \beta_1 - \beta_2 - \beta_3$,

$$S_1 - \frac{1}{x_4}\begin{pmatrix} \frac{-\beta_1 x_1 x_4 + \beta_3 D}{(B-1)x_3} & \frac{-x_2 D}{(B-1)x_3} \\ \beta_3 & -x_2 \end{pmatrix}$$

$$S_2 - \frac{1}{x_1}\begin{pmatrix} -B & -x_2 \\ \frac{\beta_3 B x_3}{D} & \frac{\beta_2 x_1 x_4 + \beta_3 x_2 x_3}{D} \end{pmatrix}$$

$Z(\beta = 0, \theta) = 1$. e_i は i 番目の成分が 1 の長さ d のベクトル. β についての漸化式で欲しい β での値を計算.

$$Y(ne_i, \theta) = Q_i(ne_i, \theta)^{-1} Q_i((n-1)e_i, \theta)^{-1} \cdots Q_i(e_i, \theta)^{-1} Y(0, \theta)$$

HGM の 3 ステップ

1. 正規化定数 Z に対する holonomic 系を求める．さらに holonomic 系を Pfaffian に変換する．理論的方法または D-加群のアルゴリズムを用いる．
2. 正規化定数とその偏微分または偏差分を必要な rank まである点で計算する．[初期値を求める.]
3. 2 で求めた値を Pfaffian で必要な点まで延長する．[初期値問題を解く.]

注: rank は step 3 の計算量を決める．
理論的手法による Pfaffian の導出と応用例: $2 \times (m+1)$ 分割表. Lauricella F_D (m 変数) の方程式．微分 Pfaffian $\partial_i - R_i$ の交点数による公式 (松本, 2013). β non-generic のもと，差分 Pfaffian の $S_i - Q_i$ の公式 (後藤, Contiguity relations of Lauricella's F_D revisited, arxiv:1412.3256).
2×11 分割表の Z とその偏微分の計算．2s 以内．MCMC より早く高精度．応用: センター試験の科目選択の統計解析．小川, 竹村, 高山 (2015).

(差分) HGM のタイミングデータ

正規化定数 $Z = \sum x^e/e!$ の数値計算．ここで e は次を満たす 2×12 分割表を動く．

-A												
29+2046												
	-A+29	31	62	93	124	155	186	217	248	279	310	341

また x は $\begin{pmatrix} 1 & 1/2 & 1/3 & 1/4 & 1/5 & 1/6 & 1/7 & 1/8 & 1/9 & 1/10 & 1/11 & 1/12 \\ 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 & 1 \end{pmatrix}$.

後藤の差分方程式 (漸化式) は 1.056 秒で計算できる．この差分方程式 (漸化式) を用いて Z を計算する時間は以下のとおり．

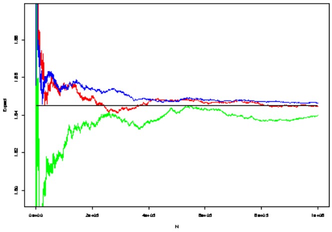
-A	Time	Z
50	0.224s	8.162 E-3964
100	0.825s	1.461 E-4026
150	2.111s	1.087 E-4105
200	4.292s	2.171 E-4195

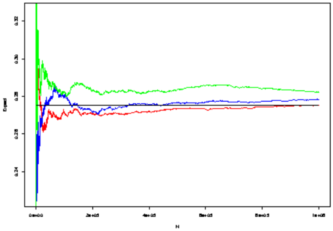
Note that the HGM generates all the normalizing constants for $\{1, 2, \ldots, -A\}$ and derivatives with these timings.

$$E(e_{ij}) = x_{ij} \frac{\partial Z}{\partial x_{ij}} \frac{1}{Z}$$

The timing data are taken on a machine with Intel Xeon E5-4650 (2.7GHz) with 256G memory and the computer algebra system Risa/Asir (20130215) and `tk_fd.rr` package.

MCMC による期待値計算





赤: 標準的な Markov basis. [[3,0,0,0],[3,3,3,3]],

[[1,1/2,1/3,1/4],[1,1,1,1]]. E_{11}

E_{14}

MCMC の妥当性を調べる.
MCMC での期待値計算. 計算に数分. $2 \times (m+1)$ 分割表では多くの場面で MCMC を HGM に置き換えうる.

HGM の広がり: Fisher-Bingham 分布

d 次元球面 $S^d(r) = \{(u_1, \ldots, u_{d+1}) \mid \sum_{i=1}^{d+1} u_i^2 = r^2, r > 0\}$ の上でパラメータ x_{ij}, y_i を持つ分布

$$\mu(u; x, y, r)\delta := \frac{1}{Z(x, y, r)} \exp\left(\sum_{1 \le i \le j \le d+1} x_{ij} u_i u_j + \sum_{i=1}^{d+1} y_i u_i \right) \delta$$

を Fisher-Bingham 分布と呼ぶ (1980 年代から始まる A.T.A. Wood 等の研究が先駆). ここで $\delta = \delta(r, u)$ は $\int_{\mathbf{R}^{d+1}} \delta(r, u) = r^d \frac{2\pi^{(d+1)/2}}{\Gamma((d+1)/2)}$ をみたす球面に台をもつデルタ関数, Z は次で定義される正規化定数

$$Z(x, y, r) = \int_{\mathbf{R}^{d+1}} \exp\left(\sum_{1 \le i \le j \le d+1} x_{ij} u_i u_j + \sum_{i=1}^{d+1} y_i u_i \right) \delta(r, u). \quad (1)$$

(清, 竹村が発端. その後, 小山民雄ら)

例：神戸市の平均風向 (16 方位) のデータ

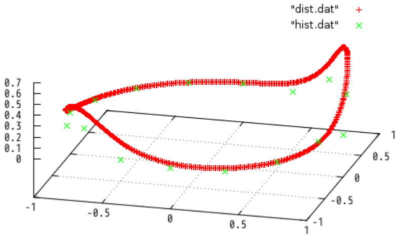
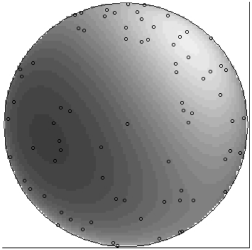


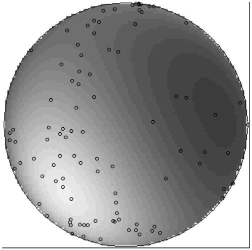
Figure: 頻度と密度関数

方位	ラジアン	頻度
北	1.570760	0.00
北北東	1.178070	0.00
北東	0.785380	0.24
東北東	0.392690	0.24
東	0.000000	0.03
東南東	5.890349	0.12
南東	5.497659	0.03
南南東	5.104969	0.00
南	4.712279	0.02
南南西	4.319589	0.00
南西	3.926899	0.19
西南西	3.534209	0.07
西	3.141519	0.03
西北西	2.748829	0.00
北西	2.356139	0.00
北北西	1.963450	0.00

出典: 小山, 中山, 西山, 高山: 高次元における FB 分布の MLE への hgd 法の適用について (スライド), 2011–2013.

天球の星, S^2 での Fisher-Bingham 分布による fitting





出典: 中山, 西山, 野呂, 小原, 清, 高山, 竹村, Holonomic Gradient Descent and its Application to Fisher-Bingham Integral, Advances in Applied Mathematics 47 (2011), 639–658,

Orthant probability

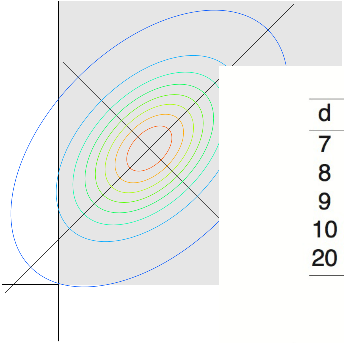
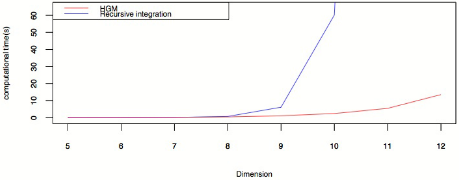


Table: Results of HGM

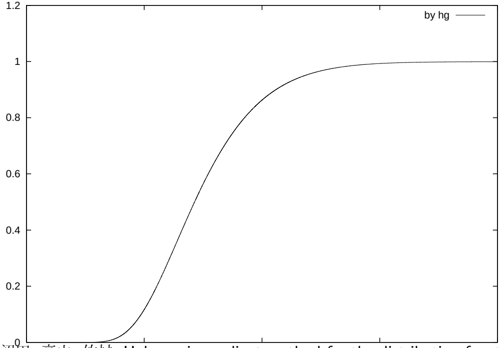
d	HGM	Exact
7	0.125000000	0.125000000
8	0.111111111	0.111111111
9	0.100000000	0.100000000
10	0.090909091	0.090909091
20	0.0476190473	0.0476190476



Orthant Probability の HGM による計算 (小山, 竹村, 2012))

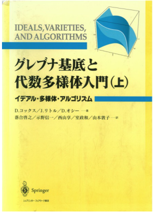
Wishart 分布の第一固有値, 行列引数超幾何系 ${}_1F_1$

累積分布関数 $\Pr[\ell_1 < x]$, $m = 10$, $n = 12$,
$\Sigma^{-1} = \mathrm{diag}(1, 2, \ldots, 10)$



出典: 橋口, 沼田, 高山, 竹村, Holonomic gradient method for the distribution function of the largest root of a Wishart matrix, Journal of Multivariate Analysis, 117, (2013) 296-312.

参考書の紹介



まもなく “グレブナー教室” が.
HGM の要素: 統計, 代数, 特殊関数 (Pfaffian, 漸近展開, ...), ホロノミック系, 数式処理, 数値解析.

$\boxed{\text{OpenXM} \quad \text{hgm}}$ 検索