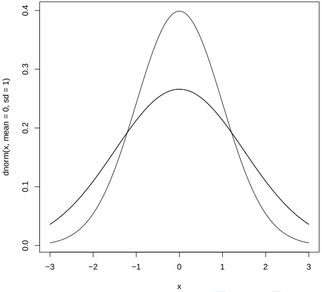
Holonomic Gradient Method and Statistics

$f(\theta, t)$: unnormalized probability distribution function with respect to $t = (t_1, \ldots, t_n)$ where $\theta = (\theta_1, \ldots, \theta_m)$ is a parameter vector.

$$z(\theta) = \int_\Omega f(\theta, t) dt$$

is the normalizing constant. $f(t, \theta)/z(\theta)$ is a probability distribution function on Ω. Evaluation of the N.C. $z(\theta)$ is a fundamental problem in statistics.

Example: $f(\theta, t) = \exp\left(\frac{-t^2}{2\theta^2}\right)$,
$\Omega = (-\infty, +\infty)$,
then $z(\theta) = \sqrt{2\pi\theta^2}$.



Holonomic Gradient Method and Statistics

An analytic function $f(x)$ is called a holonomic function when it satisifes n linear ODE's

$$\sum_{j=0}^{r_i} a_{ij}(x) \left(\frac{\partial}{\partial x_i} \right)^j f, \quad a_{ij}(x) \in \mathbf{C}[x_1, \ldots, x_n], i = 1, \ldots, n.$$

Theorem (Zeilberger, 1990)

If $f(x_1, \ldots, x_n)$ is a holonomic function in x, then the integral $\int_\Omega f(x) dx_n$ is a holonomic function in $(x_1, \ldots, x_{n-1})$ (under some conditions on the set Ω). [Use the theory of D-modules]

Holonomic Gradient Method [$N_3 OST_2$; 2011]

When $f(\theta, t)$ is a holonomic function, the N.C. satisfies a system of linear partial differential equations, which can be algorithmically constructed by Gröbner bases. Evaluate the N.C. and its derivatives by the system with methods in numerical analysis.

3 Steps of Holonomic Gradient Method

1. Construct a Pfaffian system (system of ODE's) for $z(\theta)$. [1]
2. Evaluate numerically $z(\theta)$ and its derivatives at a point $\theta = \theta_0$.
3. Apply numerical analysis methods for the Pfaffian system.

Example:

$$z(\theta) = \int_\Omega \exp(\theta t) t^{1/2} (1-t)^{1/2} dt, \ \Omega = [0,1]$$

$$\Rightarrow \quad (\theta \partial_\theta^2 + (3-\theta)\partial_\theta - 3/2) z = 0, \quad \partial_\theta = \frac{\partial}{\partial \theta}$$

$$\Rightarrow \quad \frac{\partial}{\partial \theta} Z = PZ,$$

$$Z = \begin{pmatrix} z \\ \frac{\partial}{\partial \theta} z \end{pmatrix}, P = \begin{pmatrix} 0 & 1 \\ \frac{3}{2\theta} & -\frac{3-\theta}{\theta} \end{pmatrix}$$

[1] Oaku's algorithm (1997) constructs it in general, but the complexity is high.

Example from directional statistics

Von-Mises distribution. $f(\theta, t) = \exp(\theta_1 t_1 + \theta_2 t_2)$, $(t_1, t_2) \in S^1$.

$$z(\theta) = \int_{t_1^2 + t_2^2 = 1} \exp(\theta_1 t_1 + \theta_2 t_2) dt$$

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

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

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

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

Application to Fisher's maximal likelihood estimates

T_i's are data. Find θ such that

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

is maximal.

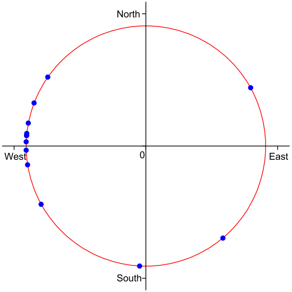


Figure : Wind directions at 9:00 AM from 2011-01-01 to 2011-01-14 (NA on 11). Data is taken at the height of 10000m of Sapporo.

References

Zeilberger Doron Zeilberger, A Holonomic Systems Approache to Speical Function Identities, Journal of Computational and Applied Mathematics 32 (1990), 321–368.

Oaku Toshinori Oaku, Algorithms for *b*-functions, restrictions, and algebraic local cohomology groups of *D*-modules, Advances in Applied Mathematics 19 (1997), 61–105.

N_3OST_2 Hiromasa Nakayama, Kenta Nishiyama, Masayuki Noro, Katsuyoshi Ohara, Tomonari Sei, Nobuki Takayama, Akimichi Takemura, Holonomic Gradient Descent and its Application to the Fisher-Bingham Integral, Advances in Applied Mathematics 47 (2011), 639–658.

HNTT Hiroki Hashiguchi, Yasuhide Numata, Nobuki Takayama, Akimichi Takemura, Holonomic gradient method for the distribution function of the largest root of a Wishart matrix, Journal of Multivariate Analysis 117 (2013), 296–312.