Estimation and Test 1 Let $T_i(t)$ be (a smooth) function in $t \in S \subset \mathbf{R}^d$, dt (the Legesgue) measure on $\mathbf{R}^d$, $\theta \in \mathbf{R}^n$.

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

$$Z(\theta) = \int_S u(\theta, t) dt \quad \text{normalizing constant}$$

Example: $A = (a_1, \ldots, a_n)$, $a_i \in \mathbf{N}_0^d$ (column vectors). Put $T_i(t) = t^{a_i} := \prod_{j=1}^{d} (t_j)^{(a_i)_j}$.

Example: $d = 1$, $A = (1, 2)$. $u(\theta, t) = \exp(\theta_1 t + \theta_2 t^2)$.

$Z(\theta) = \frac{\sqrt{\pi}}{\sqrt{-\theta_2} \exp(\theta_1^2/(4\theta_2))}$.

$u(\theta, t)/Z(\theta)$ defines a probability distribution on S with a parameter θ (exponential family). The set of $T_i(t)$'s are called the sufficient statistics of this distribution.

Fisher's maximum likelihood estimation (MLE) For given data $t(j) \in S$, $j = 1, \ldots, N$, find θ such that $\prod_{j=1}^{N} u(\theta, t(j))/Z(\theta)$ (probability) takes the maximum.

Estimation and Test 2

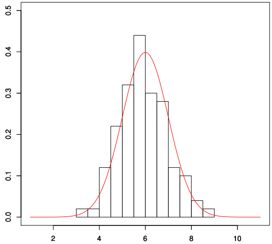
$$
\begin{aligned}
& \text{maxarg}_\theta \prod_{j=1}^{N} u(\theta, t(j))/Z(\theta) \\
= \quad & \text{maxarg}_\theta \sum_j \log u(\theta, t(j)) - N \log Z(\theta) \\
= \quad & \text{maxarg}_\theta \sum_i \theta_i \frac{1}{N} \sum_j T_i(t(j)) - \log Z(\theta) \qquad (1)
\end{aligned}
$$

$-\log Z(\theta)$ is an upper convex function. Then the θ wanted is a solution of the MLE equation

$$\frac{\partial}{\partial \theta_i} \log Z = \frac{1}{N} \sum_j T_i(t(j)) = \bar{T}_i \qquad (2)$$

Example: $A = (1, 2)$, $\theta_2 = -\frac{1}{2}$, $\frac{1}{N} \sum_j t(j) = \frac{\partial}{\partial \theta_1} \log Z = -\frac{\theta_1}{2\theta_2} = \theta_1$ (mean).

Estimation and Test 3 , Example.



$\theta_2 = -1/2$ (fixed). 100 Data as the histgram with $\bar{T}_1 = 6.0$ and MLE $\theta_1 = \bar{T}_1 = \frac{1}{100}\sum_{j=1}^{100} t_1(j) = 6.0$ (mean).

It is easy, because an explicit expression of $\partial_i \bullet \log Z$, $\partial_i = \frac{\partial}{\partial\theta_i}$ is known and there is no difficulty of solving the MLE equation.

Problem How do we solve the MLE equation?

Theorem (See, e.g., L.D.Brown (1986)*, Michalek-Sturmfels-Uhler-Zwiernik (2015) †)

1. $C = \{\theta \in \mathbf{R}^n \mid \log Z(\theta) < \infty\}$. *C is convex.*
2. $K = \mathrm{conv}\, T(S)^{\ddagger}$. *Under some conditions,* $\nabla \log Z$ *is a bijection from* C *to the interior of* K.

A is a $d \times n$ matrix with entries in $\mathbf{N}_0$ with $\mathrm{rank}\, A = d$ and with an $(1, 1, \ldots, 1)$-row. Fix $\beta \in \mathbf{N}_0^d$.

Theorem (T-Kuriki-Takemura (2018) §)

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

If the dimension of $\mathrm{New}(Z)$ *(as poly in p) is* $n - d$*, then* $\nabla \log Z$ *is a bijection from* $\mathbf{R}^n/\mathrm{Im}\, A^T$ *to* $\mathrm{relint}(\mathrm{New}(Z))$.

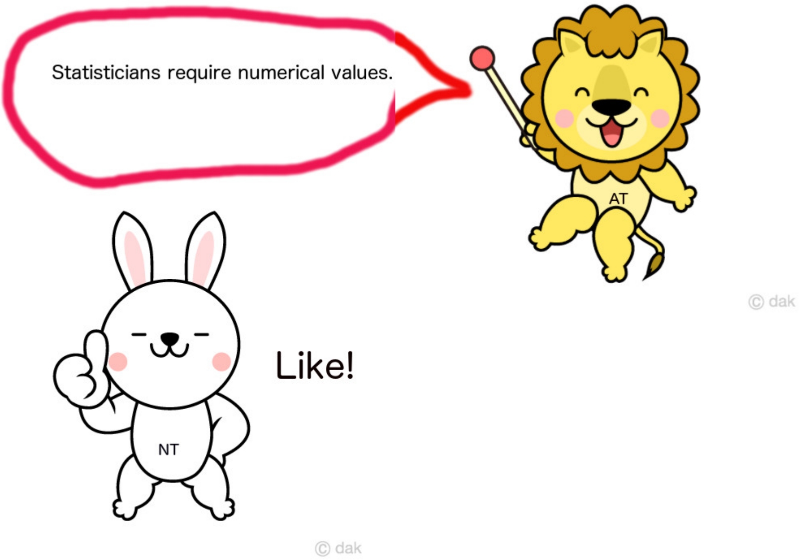
*Th 3.6 in "Fundamentals of Statistical Exponential Families with Applicatioins in Statistical Decision Theory"

†Th 2.2 in "Exponential Varieties"

‡ $T(t) = (T_1(t), \ldots, T_d(t))$

§Th 1 in "A-hypergeometric distributions and Newton polytopes"

Estimation and Test 5. There exists a solution of MLE, and ...



Estimation and Test 6 Numerical values of Z and its derivatives are needed for the gradient descent to solve the MLE equation.

1. Numerical Integration of Z and Monte-Carlo simulation: General. Need good random number/vector/matrix generators. No high precision output.
2. Series expansion of Z: Fast and high precision in the case of one variable. Slow for several variable case. How to derive series expansions and connection formulas?
3. Approximation of Z: Saddle point method, Euler characteristic method, ... No global approximation. How to derive an approximation?

We proposed the holonomic gradient method (HGM) * and the holonomic gradient descent (HGD).

* hgm openxm : search ⇒ a page of references of HGM

Some introductory references.
Hibi, eds. "Gröbner bases : statistics and software systems" (2013)
Sattelberger-Sturmfels, "D-Modules and Holonomic Functions" (2019)

What is HGM (holonomic gradient method)? It works for holonomic functions and holonomic distributions.

Three Steps of HGM

1. Finding a holonomic system satisfied by the normalizing constant. We may use computational * or theoretical methods to find it. Translating it to a Pfaffian system.
2. Finding an initial value vector for the Pfaffian system. This is equivalent to evaluating the normalizing constant and its derivatives at a point.
3. Solving the Pfaffian system numerically. We use several methods in numerical analysis.

* There are algorithms to obtain it (by Gröbner basis in the ring of differential operators).

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).

Statistical Test 1 How "rare" is the event?

(Sloppy) Example: Assume "people feels comfortable when the temperature is 25C and the distribution of feeling comfortable follows the normal distribution with the mean=25 and the standard deviation 3".

We prepare a room of 20C and make an interview to a randomly chosen person if this person feels comfortable. If this person says "comfortable", we should reject this assumption, because

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

and then it is "rare" that randomly chosen person says "comfortable" *

This example motivates, e.g., the problem of evaluating $I(P;\theta)/I(S;\theta)$ where S the whole event space, $P \subset S$ and

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

*In text books of statistics, the definition of "rare" is not rigorously given. They usually assume that the sum or integration of the tail probability is the probability of "rare".

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

Summary

1. The exponential family is an important class of distributions in statistics. When the sufficient statistics $T_i(t)$'s are monomials or logarithmic functions, we obtain an incomplete A-hypergeometric system when the domain is, e.g, a polyhedron, which are used in statistical test [‡].
2. The MLE equation has a unique solution under some conditions. In order to obtain the "exact" numerical solution, we need to evaluate the normalizing constant Z and its derivatives. So is statistical tests.
3. The HGM (holonomic graidient method) and the HGD (holonomic gradient descent) are new (algorithmic) method for the numerical evaluation for "holonomic" distributions. The methods are applied to several important distributions and software packages are provided.

[‡] In statistics, we often integrate by a delta measure. The integral is no longer A-hypergeometric.