An introduction to holonomic gradient method in statistics

Akimichi Takemura, Univ. Tokyo

September 5, 2012

Items

1. **First example: Airy-like function**
2. **Holonomic function and holonomic gradient method (HGM)**
3. **Another example: incomplete gamma function**
4. **Wishart distribution and hypergeometric function of a matrix argument**
5. **HGM for two-dimensional Wishart matrix**
6. **Pfaffian system for general dimension**
7. **Numerical experiments**

- **Coauthors on HGM: H.Hashiguchi, T.Koyama, H.Nakayama, K.Nishiyama, M.Noro, Y.Numata, K.Ohara, T.Sei, N.Takayama.**
- **HGM proposed in "Holonomic gradient descent and its application to the Fisher-Bingham integral",** *Advances in Applied Mathematics*, **47, 639–658. N^3OST^2. 2011.**
- **We have now about 7 manuscripts on HGM, mainly by Takayama group.**
- **Wishart discussed in** `arXiv:1201.0472v3`

Ex.1: Airy-like function

An exercise problem by Nobuki Takayama on Sep.16, 2009, during "Kobe Gröbner School".

Question: Let

$$A(x) = \int_0^\infty e^{-t-xt^3} dt, \qquad x > 0.$$

Derive a differential equation satisfied by $A(x)$.

Answer:

$$\begin{aligned} 1 &= (27x^3\partial_x^2 + 54x^2\partial_x + 6x + 1)A(x) \\ &= 27x^3 A''(x) + 54x^2 A'(x) + (6x+1)A(x). \end{aligned}$$

- **Actually this question is pretty hard, even if you are told the answer.**
- **I was struggling with this problem, wasting lots of papers and wondering why I was doing this exercise.**
- **After one hour, I suddenly realized that this is indeed an important problem in statistics.**

- **We try to confirm the given answer.**
- **First**

$$A'(x) = \partial_x \int_0^\infty e^{-t-xt^3}\,dt = -\int_0^\infty t^3 e^{-t-xt^3}\,dt$$
$$A''(x) = \int_0^\infty t^6 e^{-t-xt^3}\,dt.$$

- **Hence**

$$\begin{aligned}
&27x^3A''(x) + 54x^2A'(x) + (6x+1)A(x) \\
&\quad = \int_0^\infty (27x^3t^6 - 54x^2t^3 + 6x + 1)e^{-t-xt^3}\,dt \\
&\quad = 1 \quad ??
\end{aligned}$$

- **Integration by parts:**

$$\partial_t \, e^{-t-xt^3} = -(1+3xt^2)e^{-t-xt^3}.$$

Hence

$$\begin{aligned} 0 &= \left[te^{-t-xt^3} \right]_0^\infty \\ &= A(x) - \int_0^\infty t(1+3xt^2)e^{-t-xt^3}. \end{aligned}$$

- **Similarly, if you work hard and work out**

$$1 = \left[(-9x^2t^4 + 3xt^2 + 6xt - 1)e^{-t-xt^3}\right]_0^\infty$$

by integration by parts (how did you get this?), you obtain the answer:

$$\int_0^\infty (27x^3t^6 - 54x^2t^3 + 6x + 1)e^{-t-xt^3}\,dt = 1.$$

- **Believe and try**

$$\begin{aligned}
&(-9x^2t^4 + 3xt^2 + 6xt - 1)' + (9x^2t^4 - 3xt^2 - 6xt + 1)(1 + 3xt^2) \\
&\quad = (-36x^2t^3 + 6xt + 6x) + (9x^2t^4 - 3xt^2 - 6xt + 1) \\
&\qquad + (27x^3t^6 - 9x^2t^4 - 18x^2t^3 + 3xt^2) \\
&= 27x^3t^6 - 54x^2t^3 + 6x + 1 \quad (\textbf{OK!})
\end{aligned}$$

- **Actually Prof. Takayama (or one of his students) did this by computer, asking us to do this by hand(!)**

  $\leftarrow$ **Integration algorithm for** D**-modules**

  based on Gröbner basis for D**-modules**

Is this exercise related to statistics?

- **Change the notation and let**

$$A(\theta) = \int_0^\infty e^{-x-\theta x^3} dx.$$

- **Let**

$$f(x;\theta) = \frac{1}{A(\theta)} e^{-x-\theta x^3}, \qquad x, \theta > 0.$$

- **This is an exponential family with the sufficient statistic $T(x) = x^3$. (We can absorb e^{-x} into the base measure dx.)**

- **Therefore we are evaluating the normalizing constant and its derivatives of an exponential family.**

- **We now know**

$$1 = 27\theta^3 A''(\theta) + 54\theta^2 A'(\theta) + (6\theta + 1)A(\theta).$$

- **Hence the Fisher information $A''(\theta)$ is automatically obtained from $A(\theta)$ and $A'(\theta)$.**

- **Suppose that we have independent observations $x_1, \ldots, x_n$ from $f(x;\theta)$.**
- **Then the log likelihood $\ell(\theta)$ is written as**

$$\ell(\theta) = -\theta \sum_{i=1}^{n} x_i^3 - n \log A(\theta)$$

and the likelihood equation is

$$0 = -\sum_{i=1}^{n} x_i^3 - n \frac{A'(\theta)}{A(\theta)}.$$

- **Can we numerically evaluate $A(\theta)$ and $A'(\theta)$? (Also $A''(\theta)$ for Newton-Raphson?)**

- **For illustration, we use simple linear approximation (Euler method for numerical integration).**

$$A(\theta + \Delta\theta) \doteq A(\theta) + \Delta\theta A'(\theta)$$
$$A'(\theta + \Delta\theta) \doteq A'(\theta) + \Delta\theta A''(\theta).$$

- **But from the differential equation we know**

$$A''(\theta) = \frac{1}{27\theta^3}(1 - (6\theta + 1)A(\theta) - 54\theta^2 A'(\theta)).$$

- **Punch line: if you keep numerical values of $A(\theta), A'(\theta)$ at one point θ, then you can compute these values at nearby $\theta + \Delta\theta$.**

- **At each point, higher-order derivatives $A''(\theta)$, $A'''(\theta), \ldots,$ can be computed as needed.**

- **Hence by numerically solving ODE, you can compute $A(\theta)$ and its derivatives at any point**

  $\rightarrow$ "Holonomic Gradient Method"

- **For explanation we used Euler method, but in our actual implementation we use Runge-Kutta method to solve ODE.**

Holonomic function and holonomic gradient method (HGM)

Univariate homogeneous case:

- **A smooth function f is holonomic if f satisfies the following ODE**

$$0 = h_k(x) f^{(k)}(x) + \cdots + h_1(x) f'(x) + h_0(x) f(x),$$

where $h_k(x), \ldots, h_1(x), h_0(x)$ are rational functions of x.

- **Approximation**

$$\begin{pmatrix} f(x+\Delta x) \\ f'(x+\Delta x) \\ \vdots \\ f^{(k-1)}(x+\Delta x) \end{pmatrix}$$

$$\doteq \left[I_k + \Delta x \begin{pmatrix} 0 & 1 & 0 & \dots & 0 \\ 0 & 0 & 1 & \dots & 0 \\ & & & \dots & \\ 0 & 0 & \dots & 0 & 1 \\ -\frac{h_0(x)}{h_k(x)} & -\frac{h_1(x)}{h_k(x)} & \dots & & -\frac{h_{k-1}(x)}{h_k(x)} \end{pmatrix} \right] \begin{pmatrix} f(x) \\ f'(x) \\ \vdots \\ f^{(k-1)}(x) \end{pmatrix}$$

Multivariate case (for simplicity let $\dim = 2$)

- **A smooth $f(x, y)$ is holonomic if for each of x and y, fixing the other variable arbitrarily, we have a holonomic function in x and y.**

- **Namely, if there exist rational functions $h_0^1(x, y), \ldots, h_{k_1}^1(x, y)$, $h_0^2(x, y), \ldots, h_{k_2}^2(x, y)$ in x, y such that**

$$\sum_{i=0}^{k_1} h_i^1(x, y) \partial_x^i f(x, y) = 0,$$

$$\sum_{i=0}^{k_2} h_i^2(x, y) \partial_y^i f(x, y) = 0.$$

- **Consider** $\partial_x^{r_1}\partial_y^{r_2} f(x,y)$. **If** $r_1 \geq k_1$ **or** $r_2 \geq k_2$, **we can always compute this by recursively applying the differential equations.**

- **As in the univariate case, if we keep numerical values of** $\partial_x^i \partial_y^j f(x,y)$ **in the range** $i = 0, \ldots, k_1 - 1$, $j = 0, \ldots, k_2 - 1$, **then we can always compute other higher-order derivatives.**

- **We can also approximate**

$$\partial_x^i \partial_y^j f(x + \Delta x, y + \Delta y)$$

by the values $\{\partial_x^i \partial_y^j f(x,y)\}_{i=0,\ldots,k_1-1, j=0,\ldots,k_2-1}$.

- **We usually only need to keep a subset of** $\{\partial_x^i \partial_y^j f(x,y)\}_{i=0,\ldots,k_1-1,j=0,\ldots,k_2-1}$ **in memory. The subset is given by the set of standard monomials obtained by the division algorithm based on a Gröbner basis.**

Which functions are holonomic (Zeilberger(1990))?

- **Polynomials and rational functions are holonomic.**
- $\exp($**rational**$)$, $\log($**rational**$)$ **are holonomic.**
- f, g : **holonomic** $\Rightarrow f + g, f \times g$: **holonomic**
- $f(x_1, \ldots, x_m)$: **holonomic**

  $\Rightarrow \int f(x_1, \ldots, x_m) dx_m$: **holonomic**
- **Restriction of a holonomic** f **to an affine subspace is holonomic.**

- **Holonomocity is also defined for generalized functions.**

- **From the above properties, it is often easy to tell that a given function (such as the Airy-like function) is holonomic, i.e., it <u>must</u> satisfy a differential equation with rational function coefficients.**

- **The problem is to find the explicit form of the differential equation.**

Ex.2: Incomplete Gamma function

- **Consider incomplete Gamma function**

$$G(x) = \int_0^x y^{\alpha-1} e^{-y} dy, \qquad \alpha, x > 0.$$

- **From general theory, $G(x)$ is holonomic.**

- **We integrate in the opposite direction and change scale:**

$$\int_0^x (x-y)^{\alpha-1} e^{-(x-y)} dy = x^{\alpha-1} e^{-x} \int_0^x (1 - \frac{y}{x})^{\alpha-1} e^{y} dy$$
$$= x^{\alpha} e^{-x} \int_0^1 (1-z)^{\alpha-1} e^{xz} dz$$

- **Expand e^{xz} into Taylor series** **(not clever?)**

$$G(x) = x^\alpha e^{-x} \sum_{k=0}^{\infty} \frac{1}{k!} x^k \int_0^1 (1-z)^{\alpha-1} z^k dz$$

$$= x^\alpha e^{-x} \sum_{k=0}^{\infty} \frac{\Gamma(\alpha)\Gamma(k+1)}{k!\Gamma(\alpha+k+1)} x^k$$

$$= \frac{1}{\alpha} x^\alpha e^{-x} \sum_{k=0}^{\infty} \frac{(1)_k}{(\alpha+1)_k k!} x^k$$

$$(a)_k = a(a+1)\dots(a+k-1)$$

$$= \frac{1}{\alpha} x^\alpha e^{-x} \, {}_1F_1(1; \alpha+1; x)$$

- **Confluent hypergeometric function** ${}_1F_1$

$$F = {}_1F_1(a; c; x) = \sum_{k=0}^{\infty} \frac{(a)_k}{(c)_k k!} x^k$$

- **Differential equation (ODE) satisfied by** F **:**

$$xF''(x) + (c - x)F'(x) - aF = 0$$

(**Comparison of coefficients:** $\frac{a+k}{c+k}k + c\frac{a+k}{c+k} - k - a = 0$)

- **"Holonomic gradient method": We obtain the value of $F(x+\Delta x)$ from $F(x)$ by solving the ODE.**
- **If we know $F'(x)$, then we can do the following**

$$F(x+\Delta x) \doteq F(x) + \Delta x F'(x)$$

- **For the next step we also need to update $F'(x+\Delta x)$:**

$$F'(x+\Delta x) \doteq F'(x) + \Delta x F''(x)$$

- **Further we need $F''(x)$, ... But note that at any x we have**

$$F''(x) = -((c-x)F'(x) - aF(x))\frac{1}{x}.$$

Hence we can keep only $(F(x), F'(x))$ in memory. We can always compute higher-order derivatives from these two.

- **In summary, we can use the updating procedure**

$$\begin{bmatrix} F(x) \\ F'(x) \end{bmatrix} \quad \to \quad \begin{bmatrix} F(x) + \Delta x F'(x) \\ F'(x) + \Delta x F''(x) \end{bmatrix}.$$

- **In matrix form**

$$\begin{bmatrix} F(x) \\ F'(x) \end{bmatrix} + \ \Delta x \begin{bmatrix} 0 & 1 \\ \frac{a}{x} & -\frac{c-x}{x} \end{bmatrix} \begin{bmatrix} F(x) \\ F'(x) \end{bmatrix}.$$

- **How about the initial value? Use the series expansion at $x \doteq 0$.**

- **At $x = 0$ our ODE has singularity. Hence we have to step away from $x = 0$ by tiny amount.**

- **Generalization to matrix argument?**

Wishart distribution and hypergeometric function of a matrix argument

- W: $m \times m$ **symmetric positive definite** ($W > 0$)
- **Density of Wishart distribution with d.f.** n **and covariance matrix** $\Sigma > 0$:

$$f(W) = C \times \frac{|W|^{\frac{n-m-1}{2}}}{|\Sigma|^{\frac{n}{2}}} \exp(-\frac{1}{2}\mathrm{tr} W \Sigma^{-1})$$

- C **is known (containing gamma functions).**

- ℓ_1 **: the largest root of** W
- **We want to evaluate the probability** $\Pr(\ell_1 < x)$**.**

$$\ell_1 < x \;\Leftrightarrow\; W < xI_m,$$

where $I_m : m \times m$ **is the identity matrix**

- **Hence the probability is given in the incomplete gamma form:**

$$\Pr(\ell_1 < x) = C \int_{0<W<xI_m} \frac{|W|^{\frac{n-m-1}{2}}}{|\Sigma|^{\frac{n}{2}}} \exp(-\frac{1}{2}\mathrm{tr}W\Sigma^{-1})dW$$

- **From general theory** $\Pr(\ell_1 < x)$ **is holonomic.**

- **Just as in dim=1,** $\Pr(\ell_1 < x)$ **is written as**

$$C' \exp\left(-\frac{x}{2}\mathrm{tr}\Sigma^{-1}\right) x^{\frac{1}{2}nm} {}_1F_1\left(\frac{m+1}{2}; \frac{n+m+1}{2}; \frac{x}{2}\Sigma^{-1}\right)$$

- **Hypergeometric function of a matrix argument (Herz(1955)):**

$$\begin{aligned} {}_1F_1(a;c;Y) = \frac{\Gamma_m(c)}{\Gamma_m(a)\Gamma_m(c-a)} \int_{0<X<I_m} & \exp(\mathrm{tr}XY) \\ & \times |X|^{a-(m+1)/2} |I_m - X|^{c-a-(m+1)/2} dX, \end{aligned}$$

where

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

- **$_1F_1(a;c;Y)$ is a symmetric function of characteristic roots of Y $\Rightarrow$ its series expression is written in terms of symmetric polynomials.**

- **Zonal polynomials (A.T.James)**

$$\mathcal{C}_\kappa(Y), \quad \kappa \vdash k$$

  homogeneous symmetric polynomial of degree k in the characteristic roots of Y.

- **Pochhammer symbol:**

$$(a)_\kappa = \prod_{i=1}^{l} \left(a - \frac{i-1}{2} \right)_{k_i}, \quad (a)_{k_i} = \prod_{j=1}^{k_i} (a+j-1)$$

- **Series expansion of $_1F_1$ (Constantine(1963))**

$$ {}_1F_1(a;c;Y) = \sum_{k=0}^{\infty} \sum_{\kappa \vdash k} \frac{(a)_\kappa}{(c)_\kappa} \frac{\mathcal{C}_\kappa(Y)}{k!}. $$

- **This is a beautiful mathematical result. However for numerical computation, zonal polynomials have enormous combinatorial difficulties and statisticians pretty much forgot zonal polynomials.**

- **The partial differential equation satisfied by** $F(Y) = {}_1F_1(a; c; y_1, \ldots, y_m)$ **was obtained by Muirhead(1970).**

$$g_i F = 0, \quad i = 1, \ldots, m,$$

where

$$g_i = y_i \partial_i^2 + (c - y_i)\partial_i + \frac{1}{2} \sum_{j \neq i} \frac{y_j}{y_i - y_j} (\partial_i - \partial_j) - a.$$

- **Can we use this PDE for numerical computation? (People never tried this for 40 years).**

- **Works!** **works very well up to dimension** $m = 10$**!**

like Columbus's Egg

Holonomic gradient method for dimension two

- **Two partial differential equations**

$$\left[y_1\partial_1^2 + (c - y_1)\partial_1 + \frac{1}{2}\frac{y_2}{y_1 - y_2}(\partial_1 - \partial_2) - a\right]F = 0,$$

$$\left[y_2\partial_2^2 + (c - y_2)\partial_2 + \frac{1}{2}\frac{y_1}{y_2 - y_1}(\partial_2 - \partial_1) - a\right]F = 0.$$

- **Let us compute higher-order derivative from these equations.**

- **Divide the second equation by y_2 and write**

$$\partial_1^{n_1}\partial_2^{n_2}F = \partial_1^{n_1}\partial_2^{n_2-2}\Big(- \frac{c}{y_2}\partial_2 + \partial_2 - \frac{1}{2}\frac{y_1}{y_2(y_2-y_1)}(\partial_2 - \partial_1) + \frac{a}{y_2}\Big)F.$$

- **The RHS becomes messy, but an important fact is that the number of differentiations is reduced by 1.**

- **We can reduce the number of differentiations as long as there are more than 1 differentiations with respect to each variable.**

- **This implies that all higher-order derivatives can be written as a rational function combination of the following 4 square-free mixed derivatives:**

$$F(Y), \partial_1 F(Y), \partial_2 F(Y), \partial_1 \partial_2 F(Y).$$

- **In algebraic terminology, let** $K = \mathbb{C}(y_1, y_2)$ **the field of rational functions and let**

$$R = K\langle \partial_1, \partial_2 \rangle = \mathbb{C}(y_1, y_2)\langle \partial_1, \partial_2 \rangle$$

  be the ring of differential operators.

- **We take graded lexicographic term order as the term order.**

Theorem 1 $\{g_1, g_2\}$ is a Gröbner basis and $\{1, \partial_1, \partial_2, \partial_1\partial_2\}$ is the set of standard monomials.

(This theorem hold for general dimension.)

- **[Division algorithm] by repeating the differentiations, for each n_1, n_2 there exist rational functions $h_{00}^{(n_1,n_2)}, h_{10}^{(n_1,n_2)}, h_{01}^{(n_1,n_2)}, h_{11}^{(n_1,n_2)}$ in y_1, y_2, such that**

$$\partial_1^{n_1}\partial_2^{n_2}F = h_{00}^{(n_1,n_2)}F + h_{10}^{(n_1,n_2)}\partial_1 F + h_{01}^{(n_1,n_2)}\partial_2 F + h_{11}^{(n_1,n_2)}\partial_1\partial_2 F.$$

- **Hence we only keep** $F(Y), \partial_1 F(Y), \partial_2 F(Y), \partial_1\partial_2 F(Y)$ **in memory. We can always compute higher-order derivatives from these 4 values. (For dimension** m**, we need to keep** 2^m **square-free mixed derivatives in memory.)**

- **Actually for $m = 2$ we only need $\partial_1 \partial_2^2 F$.**

$$\partial_1 \partial_2^2 F = \partial_1 \Big(-\frac{c - y_2}{y_2} \partial_2 - \frac{1}{2} \frac{y_1}{y_2(y_2 - y_1)} (\partial_2 - \partial_1) + \frac{a}{y_2} \Big) F$$

$$= \Big(-\frac{c - y_2}{y_2} \partial_1 \partial_2 - \frac{1}{2} \frac{1}{(y_2 - y_1)^2} (\partial_2 - \partial_1)$$

$$- \frac{1}{2} \frac{y_1}{y_2(y_2 - y_1)} (\partial_1 \partial_2 - \partial_1^2) + \frac{a}{y_2} \partial_1 \Big) F.$$

(By symmetry, $\partial_1^2 \partial F$ is obtained by the interchange $y_1 \leftrightarrow y_2$.)

- **We further substitute ∂_1^2 to the RHS**

- **Result:**

$$
\begin{aligned}
\partial_1\partial_2^2 F = & \frac{a}{2y_2(y_2-y_1)} F \\
& + \Big(\frac{3}{4}\frac{1}{(y_2-y_1)^2} + \frac{a}{y_2} - \frac{c-y_1}{2y_2(y_2-y_1)}\Big)\partial_1 F \\
& - \frac{3}{4}\frac{1}{(y_2-y_1)^2}\partial_2 F \\
& - \Big(\frac{c-y_2}{y_2} + \frac{1}{2}\frac{y_1}{y_2(y_2-y_1)}\Big)\partial_1\partial_2 F \\
= & h_{00}^{(1,2)} F + h_{10}^{(1,2)}\partial_1 F + h_{01}^{(1,2)}\partial_2 F + h_{11}^{(1,2)}\partial_1\partial_2 F.
\end{aligned}
$$

- **Then as in the one-dimensional case, we can update:**

$$
\begin{aligned}
&F(y_1 + \Delta y_1, y_2 + \Delta y_2) \\
&\qquad \doteq F(y_1, y_2) + \Delta y_1 \partial_1 F(y_1, y_2) + \Delta y_2 \partial_2 F(y_1, y_2), \\
&\partial_1 F(y_1 + \Delta y_1, y_2 + \Delta y_2) \\
&\qquad \doteq \partial_1 F(y_1, y_2) + \Delta y_1 \partial_1^2 F(y_1, y_2) + \Delta y_2 \partial_1 \partial_2 F(y_1, y_2), \\
&[\partial_2 F \ \textbf{is similar}] \\
&\partial_1 \partial_2 F(y_1 + \Delta y_1, y_2 + \Delta y_2) \\
&\quad \doteq \partial_1 \partial_2 F(y_1, y_2) + \Delta y_1 \partial_1^2 \partial_2 F(y_1, y_2) + \Delta y_2 \partial_1 \partial_2^2 F(y_1, y_2).
\end{aligned}
$$

- **Let** $\vec{F} = (F, \partial_1 F, \partial_2 F, \partial_1\partial_2 F)^t$
- **Our updating formula is**

  $$\vec{F}(Y+\Delta Y) \doteq \vec{F}(Y)+\Delta y_1\times P_1(Y)\vec{F}(Y)+\Delta y_2\times P_2(Y)\vec{F}(Y),$$

  where P_1, P_2 **are** 4×4 **matrices consisting of rational functions in** Y **and called Pfaffian system.**
- **This simple form works for dimension two very well.**
- **For general dimension** m**, the matrices in the Pfaffian system are** $2^m \times 2^m$**. The size is** $2^m = 1024$ **for** $m = 10$**.**

Pfaffian system for general dimension

- **Theorem 1 on the non-diagonal region holds for general dimension.**
- **For Pfaffian system, we should leave the recursive form as it is, without really expanding the recursive form.**

- **For** $J \subset \{1, \ldots, m\}$**, denote the square-free derivative w.r.t. variables in** J **as**

$$\partial_J = \prod_{j \in J} \partial_j .$$

- **We need** $\partial_J \partial_i^2 F = \partial_i^2 \partial_J F$ **for** $i \notin J$**.**

- **Let $I = J \cup \{i\}$. Collect square-free terms in $\partial_J g_i$ as**

$$r(i,J;y) = -\Big[(c-y_i)\partial_I - a\partial_J + \frac{1}{2}\sum_{k\notin I}\frac{y_k}{y_i-y_k}(\partial_I - \partial_J\partial_k)$$
$$+\frac{1}{2}\sum_{k\in J}\frac{y_k}{y_i-y_k}\partial_I$$
$$+\frac{1}{2}\sum_{k\in J}\frac{y_i}{(y_i-y_k)^2}(\partial_i\partial_{J\setminus\{k\}} - \partial_J)\Big].$$

- **Separating terms, for which we need recursive differentiations in $\partial_i^2 \partial_J F$, we have**

$$y_i \partial_i^2 \partial_J F = r(i, J; y)F + \frac{1}{2} \sum_{k \in J} \frac{1}{y_i - y_k} (y_k \partial_k^2 \partial_{J \setminus \{k\}}) F.$$

- **This form is suitable for efficient programming (done by N.Takayama).**

- **When we expand the recursion to the end, we obtain**

$$y_i \partial_i^2 \partial_J F = r(i, J; y)F + \frac{1}{2} \sum_{k_1 \in J} \frac{1}{y_i - y_{k_1}} r(k_1, J \setminus \{k\}; y)F$$

$$+ \frac{1}{4} \sum_{\substack{k_1, k_2 \in J \\ k_1, k_2 : \textbf{distinct}}} \frac{1}{(y_i - y_{k_1})(y_{k_1} - y_{k_2})} r(k_2, J \setminus \{k_1, k_2\}; y)F$$

$$+ \ldots$$

$$+ \frac{1}{2^{|J|}} \sum_{\substack{k_1, \ldots, k_{|J|} \in J \\ k_1, \ldots, k_{|J|} : \textbf{distinct}}} \frac{1}{(y_i - y_{k_1}) \ldots (y_{k_{|J|-1}} - y_{k_{|J|}})} r(k_{|J|}, \emptyset; y)F.$$

- **This form is also useful for theoretical consideration of the Pfaffian system.**
- **The problem of initial values is also difficult in general dimension. N.Takayama expanded the algorithm of Koev-Edelman(2006) to handle partial derivatives.**

Numerical experiments

- **Statisticians need good numerical performance (recall zonal polynomials).**
- **We were not sure whether it works up to dimension $m = 10$.**
- **How can we check the numerical accuracy?**
  - **The initial value is taken to be close to zero. We can check the numerical convergence $\lim_{x\to\infty} \Pr(\ell_1 < x) = 1$.**

– **The following easy bounds are available:**

$$\begin{aligned}\Pr[\ell_1 < x|\mathrm{diag}(\sigma_1^2, \ldots, \sigma_1^2)] \\ \le \Pr[\ell_1 < x|\mathrm{diag}(\sigma_1^2, \ldots, \sigma_m^2)] \\ \le \Pr[\ell_1 < x|\mathrm{diag}(\sigma_1^2, 0, \ldots, 0)].\end{aligned}$$

- **Example:** $m = 10$**, d.f.** $n = 12$**,** $\Sigma^{-1} = 2\,\mathrm{diag}(1, 2, \ldots, 10)$**.**

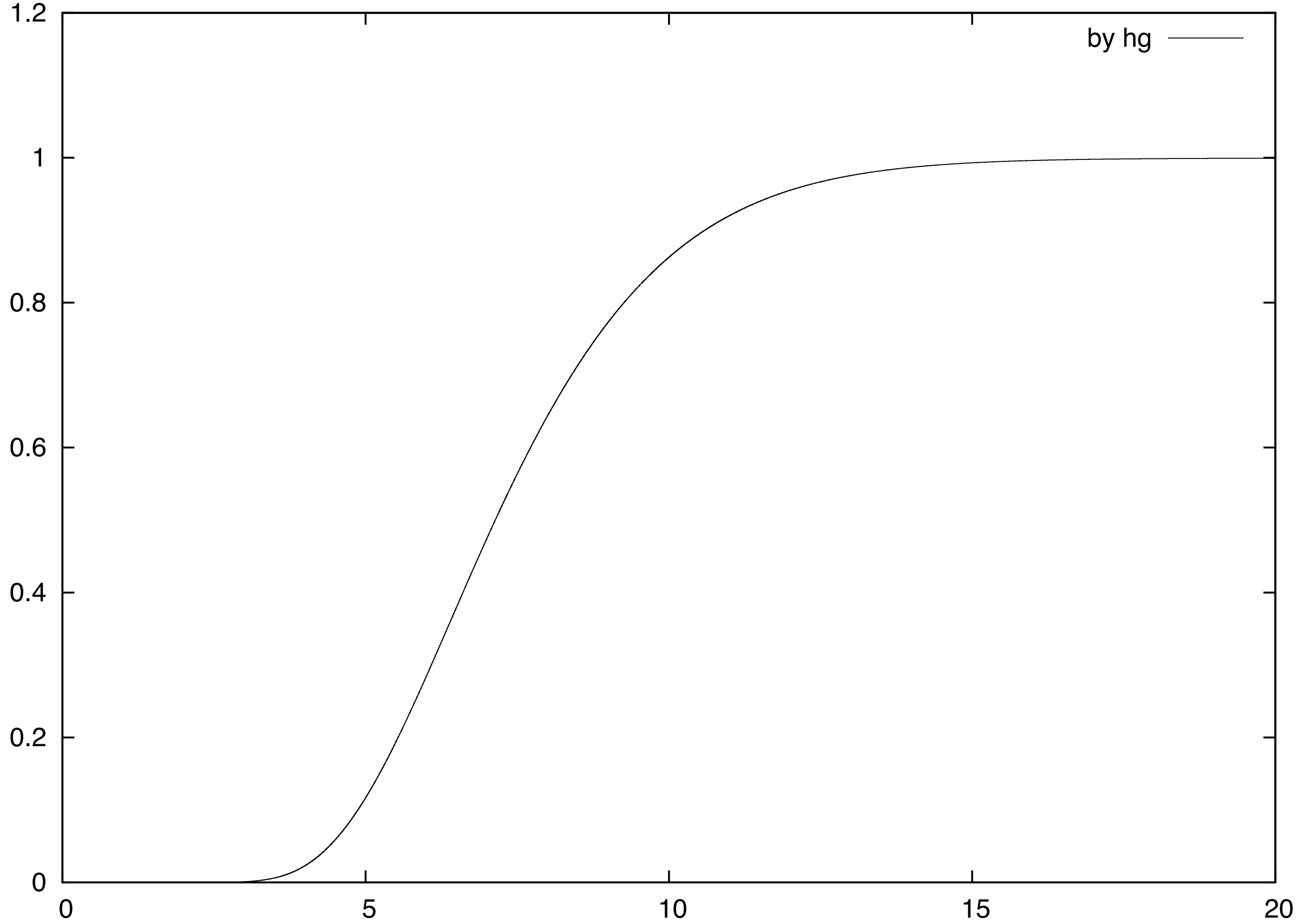
- **For $x = 30$ the bounds are $(0.99866943, 0.99999998)$.**

- **Intel Core i7 machine. The computation of the initial value at $x_0 = 0.2$ takes 20 seconds. Then with the step size 0.001, we solve the PDE up to $x = 30$, which takes 75 seconds. Output:**

$$\Pr[\ell_1 < 30] = 0.999545$$

- **This accuracy is somewhat amazing, if we consider that we updated a 1024-dimensional vector 30,000 times.**

- **Plot of the cumulative distribution**



Summary

- **Holonomic gradient method is practical if we implement it efficiently.**
- **Our approach brought a totally new approach to a longstanding difficult problem in multivariate statistics.**
- **Holonomic gradient methods is general and can be applied to many problems.**
- **We stand at the beginning of applications of** D**-module theory to statistics!**