The efficient evaluation of multi-normal distribution functions

Tetsuhisa Miwa
National Institute for Agro-Environmental Sciences
3-1-3 Kannondai, Tsukuba 305-8604, Japan
miwa@niaes.affrc.go.jp
(Plese call me Miwa /miːwə/ or Tetsu /tetsuː/.)

Anthony J. Hayter
Georgia Institute of Technology

Satoshi Kuriki
The Institute of Statistical Mathematics

1 Orthant & orthoscheme probabilities

1.1 Non-centred orthant probability

$$P_m(\mu, R) = \Pr\{x_1 \geq 0, \ldots, x_m \geq 0\} \tag{1}$$
$$= \int_0^\infty \cdots \int_0^\infty \phi_m(x; \mu, R)\ dx_1 \cdots dx_m$$

$$x = (x_1, \ldots, x_m)' \sim N_m(\mu, R)$$

R : Positive-definite correlation matrix

$P_m(0, R)$: centred orthant probability

- Multi-normal distribution function

$$F_m(c; \mu, R) = \Pr\{x_i \leq c_i,\ 1 \leq i \leq m\}$$
$$= \Pr\{-x_i + c_i \geq 0,\ 1 \leq i \leq m\}$$
$$= P_m(-\mu + c, R)$$

1.2 orthoscheme probability

$\tilde{R}$: tridiagonal

$$\tilde{R} = \begin{pmatrix} 1 & \rho_{12} & 0 & \cdots & 0 \\ \rho_{21} & 1 & \rho_{23} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho_{m-1,m-2} & 1 & \rho_{m-1,m} \\ 0 & \cdots & 0 & \rho_{m,m-1} & 1 \end{pmatrix}$$

$P_m(\mu, \tilde{R})$: non-centred orthoscheme probability

$P_m(0, \tilde{R})$: centred orthoscheme probability

OUTLINE OF TODAY'S TALK

1) Quick & accurate evaluation of $P_m(\mu, \tilde{R})$

2) Expressing a non-centred orthant probability as a linear combination of n-c orthoscheme probabilities

2 Recursive evaluation of non-centred orthoscheme probabilities

2.1 Cholesky decomposition

$$\tilde{R} = \begin{pmatrix} 1 & \rho_{12} & 0 & \cdots & 0 \\ \rho_{21} & 1 & \rho_{23} & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & \rho_{m-1,m-2} & 1 & \rho_{m-1,m} \\ 0 & \cdots & 0 & \rho_{m,m-1} & 1 \end{pmatrix} = BB'$$

$$= \begin{pmatrix} b_{11} & 0 & 0 & \cdots & 0 \\ b_{21} & b_{22} & 0 & \cdots & 0 \\ \vdots & \ddots & \ddots & \ddots & \vdots \\ 0 & \cdots & b_{m-1,m-2} & b_{m-1,m-1} & 0 \\ 0 & \cdots & 0 & b_{m,m-1} & b_{mm} \end{pmatrix} B'$$

$b_{11} = 1, b_{21}, b_{22}, \ldots$ are sequentially determined.

- **Variable transformation**

With $\tilde{R} = BB'$

$$x = Bz + \mu \sim N_m(\mu, \tilde{R})$$
$$z \sim N_m(0, I_m)$$

$$\begin{aligned}
x_1 &= z_1 + \mu_1 \\
x_2 &= b_{21} z_1 + b_{22} z_2 + \mu_2 \\
x_3 &= b_{32} z_2 + b_{33} z_3 + \mu_3 \\
&\vdots \\
x_i &= b_{i,i-1} z_{i-1} + b_{ii} z_i + \mu_i \\
&\vdots \\
x_m &= b_{m,m-1} z_{m-1} + b_{mm} z_m + \mu_m
\end{aligned}$$

$$
\begin{aligned}
P_m(\mu, \tilde{R}) &= \Pr\{x_1 \geq 0, \ldots, x_m \geq 0\} \\
&= \Pr\{z_1 + \mu_1 \geq 0, \\
&\qquad b_{21} z_1 + b_{22} z_2 + \mu_2 \geq 0, \ldots \\
&\qquad b_{i,i-1} z_{i-1} + b_{ii} z_i + \mu_i \geq 0, \ldots \\
&\qquad b_{m,m-1} z_{m-1} + b_{mm} z_m + \mu_m \geq 0\}
\end{aligned}
$$

$$
\begin{aligned}
&= \int_{-\mu_1}^{\infty} \phi(z_1) dz_1 \int_{(-b_{21} z_1 - \mu_2)/b_{22}}^{\infty} \phi(z_2) dz_2 \\
&\quad \cdots \int_{(-b_{i,i-1} z_{i-1} - \mu_i)/b_{ii}}^{\infty} \phi(z_i) dz_i \\
&\quad \cdots \int_{(-b_{m,m-1} z_{m-1} - \mu_m)/b_{mm}}^{\infty} \phi(z_m) dz_m
\end{aligned}
$$

2.2 Recursive integration formulae

$$f_{m-1}(z) = \int_{(-b_{m,m-1}z-\mu_m)/b_{mm}}^{\infty} \phi(t)dt, \tag{2}$$

$$f_{i-1}(z) = \int_{(-b_{i,i-1}z-\mu_i)/b_{ii}}^{\infty} f_i(t)\phi(t)dt, \tag{3}$$

$$2 \le i \le m-1,$$

$$P_m(\mu, \tilde{R}) = \int_{-\mu_1}^{\infty} f_1(z)\phi(z)dz. \tag{4}$$

- Evaluate (2) over a grid of points, and repeat (3).
- We can delete $f_i(z)$ after obtaining $f_{i-1}(z)$.
- Computational time $\propto m \cdot G$ (G: #{grid points}).

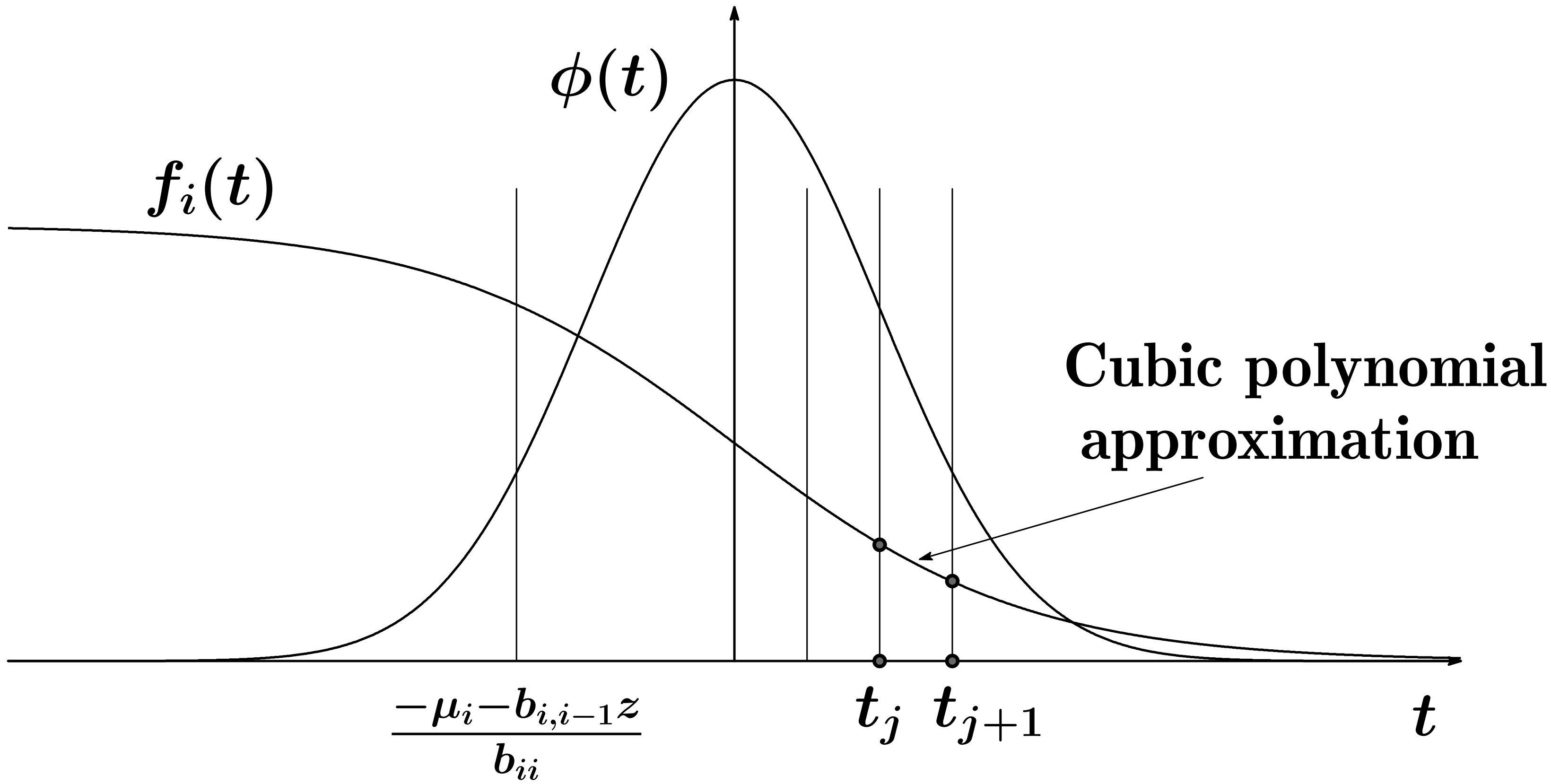

Figure 1. Recursive integration method

$$f_{i-1}(z) = \int_{(-\mu_i - b_{i,i-1}z)/b_{ii}}^{\infty} f_i(t)\phi(t)dt$$

($f_i(t)$ is defined at grid points $\ldots, t_j, t_{j+1}, \ldots$)

2.3 Comparisons with known exact values

$$\rho_{i,i-1} \equiv \rho = -1/2\text{: } P_m(0, \tilde{R}) = 1/(m+1)!$$

Table 1. Centred orthoscheme probabilities

G	$m = 5$	$m = 10$		$m = 50$	
32	0.001388885	0.2502	$\times 10^{-7}$		
64	0.0013888888	0.25051	$\times 10^{-7}$		
128	0.0013888889	0.250520	$\times 10^{-7}$	0.63	$\times 10^{-68}$
256		0.25052105	$\times 10^{-7}$	0.6440	$\times 10^{-68}$
512		0.25052108	$\times 10^{-7}$	0.64466	$\times 10^{-68}$
1024				0.644694	$\times 10^{-68}$
2048				0.6446958	$\times 10^{-68}$
4096				0.64469596	$\times 10^{-68}$
exact	0.0013888889	0.25052108	$\times 10^{-7}$	0.64469596	$\times 10^{-68}$

† G: number of grid points

Computational time is negligible.

3 Expressing an orthant probability as a linear combination of orthoscheme probabilities

3.1 Theorem

Any non-centred orthant probability can be expressed as a linear combination of at most $(m-1)!$ non-centred orthoscheme probabilities:

$$P_m(\mu, R) = \sum_{s=1}^{(m-1)!} c_s P_m(\mu^{(s)}, \tilde{R}^{(s)}) \tag{5}$$

$\tilde{R}^{(s)}$: **tridiagonal**

$c_s = +1$ **or** -1 **or** 0.

Outline of the proof

- **Variable transformation from x to z**

$$R = A'A, \quad A = (a_1, \ldots, a_m)$$
$$a_i{}'a_i = 1, \quad a_i{}'a_j = \rho_{ij}\ (i \neq j)$$

$$x = A'z \sim N_m(\mu, R)$$
$$z \sim N_m(\mu_z, I_m), \quad \mu_z = (A')^{-1}\mu$$

$$\begin{aligned} P_m(\mu, R) &= \Pr\{x_i \geq 0,\ 1 \leq i \leq m\} \\ &= \Pr\{a_i{}'z \geq 0,\ 1 \leq i \leq m\} \\ &= \Pr\{z \in Q\}. \end{aligned} \tag{6}$$

- **Polyhedral cone**

$$Q = \{z:\ a_i{}'z \geq 0,\ 1 \leq i \leq m\} \tag{7}$$

- **Hyper plane**

$$H_i = \{z : \; {a_i}' z = 0\}, \quad 1 \le i \le m$$

a_i : Normal vector to H_i

${a_i}' a_j = \rho_{ij} \; (i \ne j)$

- **Edge vector**

$$V = (v_1, \ldots, v_m) = (A')^{-1}$$

$${a_i}' v_k = 0 \; (k \ne i)$$

$$\Downarrow$$

$$v_k \in H_i$$

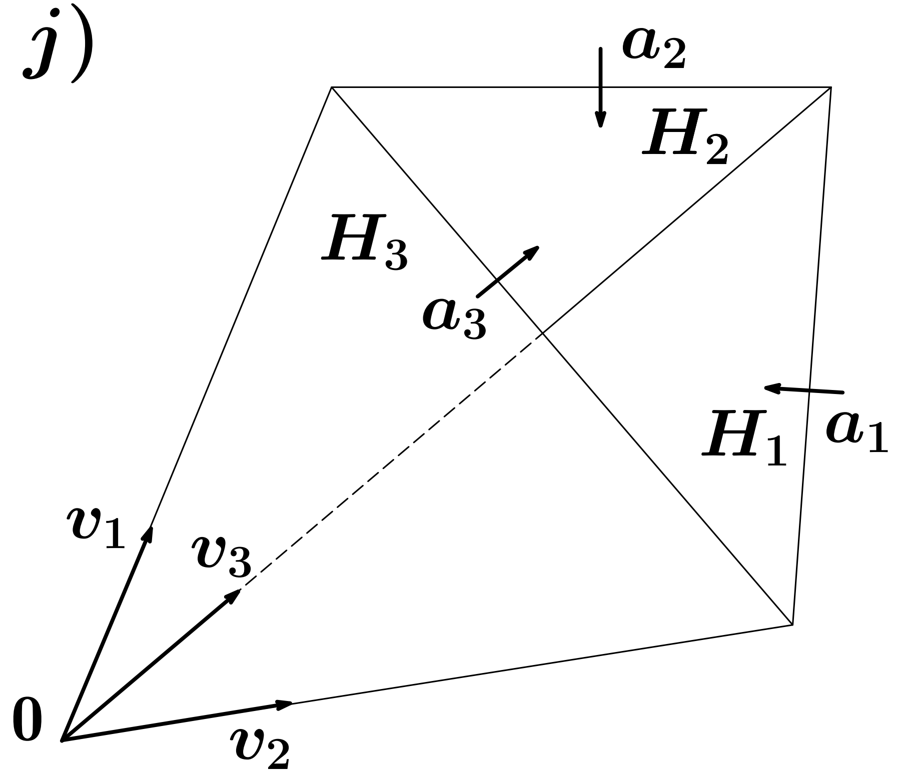

Figure 2. Example ($m = 3$)
A trihedral cone

- **Orthoscheme cone**

$$a_i' a_j = \rho_{ij} = 0, \quad |i-j| > 1 \tag{8}$$

$P_m(\mu, R)$ corresponding to these orthoscheme cones can be easily evaluated by the recursive procedure.

- **The theorem is proved by showing that any polyhedral cone can be dissected into at most $(m-1)!$ orthoscheme cones.**

- **For the detail, see**

  Miwa, T., Hayter, A. J. and Kuriki, S. (2003?). The evaluation of general non-centred orthant probabilities. *Submitted for publication.*

3.2 Example: Dissectioning a trihedral cone

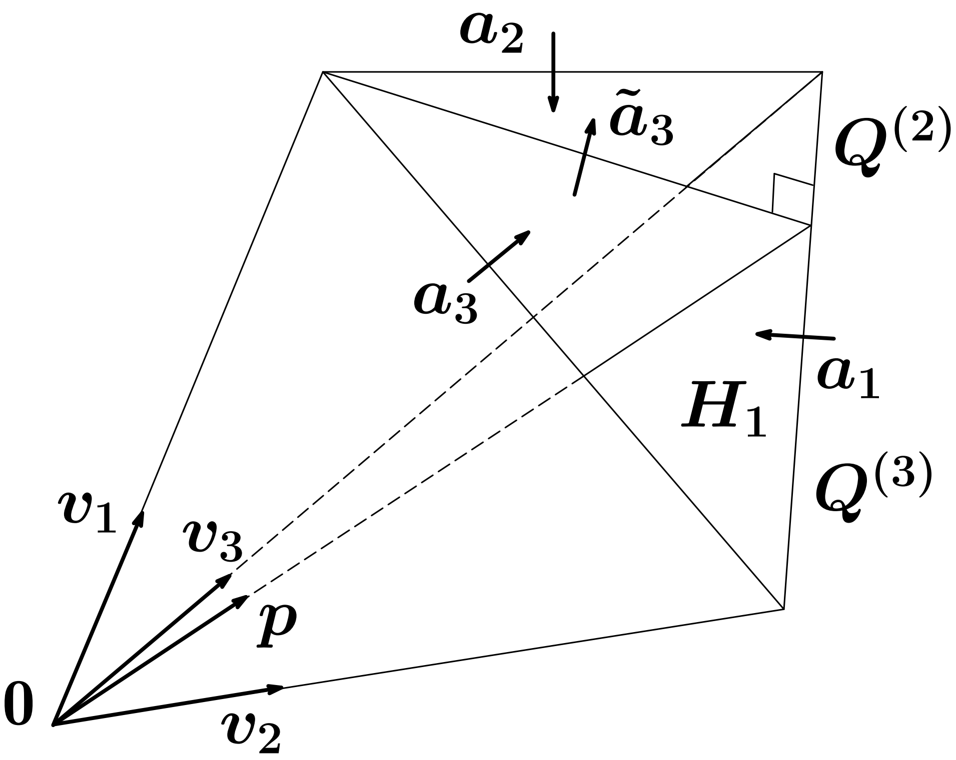

Figure 3. $Q = Q^{(2)} \cup Q^{(3)}$

Define a new edge vector p:

$$p = -\rho_{12} v_2 - \rho_{13} v_3 \in H_1$$

$$\tilde{a}_3 = \text{const} \cdot \left(a_3 - \frac{\rho_{13}}{\rho_{12}} a_2 \right)$$

$$p' \tilde{a}_3 = 0, \quad a_1{}' \tilde{a}_3 = 0$$

$Q^{(2)}$: edge vectors $v_1,\ p,\ v_3$
normal vectors $a_1,\ a_2,\ \tilde{a}_3$

$Q^{(3)}$: edge vectors $v_1,\ v_2,\ p$
normal vectors $a_1,\ \tilde{a}_2,\ a_3$

3.3 Comparisons with known exact values

Known exact values: $P_m(0, R) = 1/(m+1)$

(1) $\rho_{ij} \equiv \rho = 1/2$ ($\#\{\text{cones}\} = 40320 = 8!$)

(2) $R^{-1} = \{\rho^{ij}\}$ is tridiagonal; $\rho^{ii} = 1,\ \rho^{i,i+1} = -1/2.$

($\#\{\text{cones}\} = 323 < 8!$)

Table 2. Centred orthant probabilities ($m = 9$)

G	$\rho_{ij} = 1/2$	$\rho^{i,i+1} = -1/2$
32	0.0999751100 (12.61)	0.1000142803 (0.08)
64	0.0999990004 (23.01)	0.1000006159 (0.16)
128	0.0999999489 (44.02)	0.1000000321 (0.30)
256	0.0999999968 (87.49)	0.1000000020 (0.59)
512	0.0999999998 (172.12)	0.1000000001 (1.18)
exact	0.1000000000	0.1000000000

† G: #{grid points}. CPU time in (sec).
Run on a PC with Pentium® 4 (1.5 GHz).

4 Discussions

4.1 Applications: many

4.2 Remaining problems

- In the worst case where $(m-1)!$ cones are needed,

$$\text{Computational time} \propto m \cdot G \times (m-1)! = m! \cdot G.$$

Table 3. Computational time for various m
($\rho_{ij} \equiv \rho = 1/2$; $G = 128$ grid points)

m	5	6	7	8	9	10
CPU time (sec)	0.01	0.08	0.58	4.76	44.02	450.46

- If the correlation matrix R is singular, we have to evaluate $\Pr\{z \in \mathbb{Q}\}$ where $\mathbb{Q}$ is a polyhedron.

Thank you.

5 Historical remarks

- **Schläfli (1858):**

$$P_m(0, R) = \sum_{s=1}^{m!} c_s P_m(0, \tilde{R}^{(s)})$$

- **Abrahamson (1964):**

$$P_4(0, R) = \sum_{s=1}^{6=3!} c_s P_4(0, \tilde{R}^{(s)})$$

- **Miwa *et al.* (2003?):**

$$P_m(\mu, R) = \sum_{s=1}^{(m-1)!} c_s P_m(\mu^{(s)}, \tilde{R}^{(s)})$$

$\tilde{R}^{(s)}$: tridiagonal

6 Comparison with Monte Carlo methods

Table 4. Centred orthant probabilities with known exact probabilities ($\rho_{ij} \equiv \rho = 0.5$, $\mu_i = 0.0$)

	Our procedure				Genz's procedure			
	$G = 16$		$G = 128$		$\epsilon = 0.001$		$\epsilon = 0.0001$	
m	CPU time (sec)	abs error	CPU time (sec)	abs error	CPU time (sec)	abs error	CPU time (sec)	abs error
5	0.00	2.1e-6	0.01	1.4e-9	0.68	2.1e-4	59.67	3.5e-5
6	0.01	3.5e-6	0.08	2.0e-10	0.74	2.7e-4	97.86	2.2e-5
7	0.07	9.2e-6	0.58	5.8e-9	1.14	3.5e-4	100.22	2.7e-5
8	0.58	5.7e-5	4.76	2.1e-8	1.10	2.5e-4	97.70	3.3e-5
9	5.35	1.6e-4	44.02	5.1e-8	1.55	2.8e-4	137.62	3.6e-5
10	53.99	3.5e-4	450.46	1.0e-7	1.39	2.1e-4	124.86	2.3e-5
20					1.85	2.8e-4	198.90	3.3e-5

Genz, A. (1992). Numerical computation of multivariate normal probabilities. *J. Comput. Graph. Statist.*, 1, 141–149.