Some matrix algebra problems arisen from recent developments in Statistics

J. E. Chacón
Universidad de Extremadura

Toric Geometry Seminar
November 13, 2010

Kronecker product of matrices

- If $\mathbf{A} = \left(a_{ij}\right)_{i,j} \in \mathcal{M}_{m\times n}$, $\mathbf{B} \in \mathcal{M}_{p\times q}$ their **Kronecker product** is

$$\mathbf{A}\otimes\mathbf{B} = \left(\begin{array}{c|c|c} a_{11}\mathbf{B} & \cdots & a_{1n}\mathbf{B} \\ \hline \vdots & \ddots & \vdots \\ \hline a_{m1}\mathbf{B} & \cdots & a_{mn}\mathbf{B} \end{array}\right) \in \mathcal{M}_{(mp)\times(nq)}$$

- Basic properties:
  1. $\mathbf{A}\otimes\mathbf{B} \neq \mathbf{B}\otimes\mathbf{A}$!!
  2. $\mathbf{A}\otimes\mathbf{B}\otimes\mathbf{C} = (\mathbf{A}\otimes\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes(\mathbf{B}\otimes\mathbf{C})$
  3. $(\mathbf{A}+\mathbf{B})\otimes\mathbf{C} = \mathbf{A}\otimes\mathbf{C}+\mathbf{B}\otimes\mathbf{C}$
  4. $(\mathbf{A}\otimes\mathbf{B})(\mathbf{C}\otimes\mathbf{D}) = (\mathbf{AC})\otimes(\mathbf{BD})$
  5. $\alpha\otimes\mathbf{A} = \alpha\mathbf{A} = \mathbf{A}\alpha = \mathbf{A}\otimes\alpha$, for $\alpha\in\mathbb{R}$
  6. $(\mathbf{A}\otimes\mathbf{B})^T = \mathbf{A}^T\otimes\mathbf{B}^T$
  7. $(\mathbf{A}\otimes\mathbf{B})^{-1} = \mathbf{A}^{-1}\otimes\mathbf{B}^{-1}$
  8. $\operatorname{tr}(\mathbf{A}\otimes\mathbf{B}) = (\operatorname{tr}\mathbf{A})(\operatorname{tr}\mathbf{B})$
  9. $|\mathbf{A}\otimes\mathbf{B}| = |\mathbf{A}|^p|\mathbf{B}|^m$, for $\mathbf{A}\in\mathcal{M}_{m\times m}$, $\mathbf{B}\in\mathcal{M}_{p\times p}$

Kronecker product of vectors

- If $\mathbf{a} = (a_1, \ldots, a_p) \in \mathbb{R}^p$, $\mathbf{b} = (b_1, \ldots, b_q) \in \mathbb{R}^q$, then

$$\mathbf{a} \otimes \mathbf{b} = \left(\begin{array}{c} a_1 \mathbf{b} \\ \hline \vdots \\ \hline a_p \mathbf{b} \end{array} \right) = \left(\begin{array}{c} a_1 b_1 \\ \vdots \\ a_1 b_q \\ \hline \vdots \\ \hline a_p b_1 \\ \vdots \\ a_p b_q \end{array} \right) \in \mathbb{R}^{pq}$$

- The r^{th} **Kronecker power** of $\mathbf{A} \in \mathcal{M}_{m \times n}$ is

$$\mathbf{A}^{\otimes r} = \bigotimes_{i=1}^{r} \mathbf{A} = \mathbf{A} \otimes \mathbf{A} \otimes \cdots \otimes \mathbf{A} \in \mathcal{M}_{m^r \times n^r}$$

The vec operator

- If $\mathbf{A} = (a_{ij})_{i,j} \in \mathcal{M}_{m\times n}$ then $\operatorname{vec}\mathbf{A} \in \mathbb{R}^{mn}$ is the vector constructed by stacking the columns of $\mathbf{A}$ one underneath the other

$$\operatorname{vec}\mathbf{A} = \begin{pmatrix} a_{11} \\ \vdots \\ a_{m1} \\ \hline \vdots \\ \hline a_{1n} \\ \vdots \\ a_{mn} \end{pmatrix} \in \mathbb{R}^{mn}$$

- Basic properties:
  1. $\operatorname{vec}\mathbf{a} = \operatorname{vec}\mathbf{a}^T = \mathbf{a}$, for $\mathbf{a} \in \mathbb{R}^p$
  2. $\operatorname{vec}(\mathbf{ABC}) = (\mathbf{C}^T \otimes \mathbf{A})\operatorname{vec}\mathbf{B}$
  3. $\operatorname{vec}(\mathbf{ab}^T) = \operatorname{vec}(\mathbf{b}^T \otimes \mathbf{a}) = \operatorname{vec}(\mathbf{b} \otimes \mathbf{a}^T) = \mathbf{b} \otimes \mathbf{a}$
  4. $\operatorname{tr}(\mathbf{A}^T\mathbf{B}) = (\operatorname{vec}\mathbf{A})^T \operatorname{vec}\mathbf{B} = \sum_{i,j} a_{ij}b_{ij}$

The commutation matrix

- If $\mathbf{A} \in \mathcal{M}_{m\times n}$, the **commutation matrix** $\mathbf{K}_{m,n} \in \mathcal{M}_{(mn)\times(mn)}$ is the only one such that

$$\mathbf{K}_{m,n} \operatorname{vec} \mathbf{A} = \operatorname{vec}(\mathbf{A}^T)$$

- Basic properties:
  1. $\mathbf{K}_{m,n}^T = \mathbf{K}_{m,n}^{-1} = \mathbf{K}_{n,m}$
  2. $\mathbf{K}_{p,m}(\mathbf{A} \otimes \mathbf{B})\mathbf{K}_{n,q} = \mathbf{B} \otimes \mathbf{A}$
  3. $\mathbf{K}_{p,m}(\mathbf{A} \otimes \mathbf{a}) = \mathbf{a} \otimes \mathbf{A}$, for $\mathbf{a} \in \mathbb{R}^p$
  4. $\mathbf{K}_{m,p}(\mathbf{a} \otimes \mathbf{A}) = \mathbf{a} \otimes \mathbf{A}$, for $\mathbf{a} \in \mathbb{R}^p$
  5. $\mathbf{K}_{q,p}(\mathbf{a} \otimes \mathbf{b}) = \mathbf{b} \otimes \mathbf{a}$, for $\mathbf{a} \in \mathbb{R}^p$, $\mathbf{b} \in \mathbb{R}^q$
  6. $\operatorname{vec}(\mathbf{A} \otimes \mathbf{B}) = (\mathbf{I}_n \otimes \mathbf{K}_{q,m} \otimes \mathbf{I}_p)(\operatorname{vec} \mathbf{A} \otimes \operatorname{vec} \mathbf{B})$, where $\mathbf{I}_d$ is the $d \times d$ identity matrix

Multivariate Taylor expansion

- Suppose we have a real d-variate function $f\colon \mathbb{R}^d \to \mathbb{R}$ and two points $\mathbf{x} = (x_1, \ldots, x_d),\ \mathbf{h} = (h_1, \ldots, h_d) \in \mathbb{R}^d$.
- When $d = 1$, the Taylor expansion of f around x goes

$$f(x+h) = f(x) + f'(x)h + \tfrac{1}{2!}f''(x)h^2 + \tfrac{1}{3!}f'''(x)h^3 + \cdots$$

Multivariate Taylor expansion

- Suppose we have a real d-variate function $f\colon \mathbb{R}^d \to \mathbb{R}$ and two points $\mathbf{x} = (x_1, \ldots, x_d),\ \mathbf{h} = (h_1, \ldots, h_d) \in \mathbb{R}^d$.
- When $d = 1$, the Taylor expansion of f around x goes

$$f(x+h) = f(x) + f'(x)h + \tfrac{1}{2!}f''(x)h^2 + \tfrac{1}{3!}f'''(x)h^3 + \cdots$$

- Denote

$$\mathsf{D}f(\mathbf{x}) = \Big(\frac{\partial f}{\partial x_1}(\mathbf{x}), \ldots, \frac{\partial f}{\partial x_d}(\mathbf{x})\Big) \in \mathbb{R}^d,$$

$$\mathsf{H}f(\mathbf{x}) = \begin{pmatrix} \frac{\partial^2 f}{\partial x_1^2}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_d \partial x_1}(\mathbf{x}) \\ \vdots & \ddots & \vdots \\ \frac{\partial^2 f}{\partial x_1 \partial x_d}(\mathbf{x}) & \cdots & \frac{\partial^2 f}{\partial x_d^2}(\mathbf{x}) \end{pmatrix} \in \mathcal{M}_{d\times d}$$

the gradient vector and Hessian matrix of f, respectively.

Multivariate Taylor expansion

- For $d > 1$ the Taylor expansion of f around $\mathbf{x}$ goes

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \mathsf{D} f(\mathbf{x}) + \frac{1}{2!}\mathbf{h}^T \mathsf{H} f(\mathbf{x})\mathbf{h}$$

Multivariate Taylor expansion

- For $d > 1$ the Taylor expansion of f around $\mathbf{x}$ goes

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \mathsf{D} f(\mathbf{x}) + \frac{1}{2!}\mathbf{h}^T \mathsf{H} f(\mathbf{x})\mathbf{h} + \frac{1}{3!}\sum_{|\boldsymbol{\alpha}|=3} \mathsf{D}^{\boldsymbol{\alpha}} f(\mathbf{x})\mathbf{h}^{\boldsymbol{\alpha}} + \cdots$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ is a multiindex, $|\boldsymbol{\alpha}| = \sum_i \alpha_i$, the power is $\mathbf{h}^{\boldsymbol{\alpha}} = (h_1^{\alpha_1}, \ldots, h_d^{\alpha_d})$ and

$$\mathsf{D}^{\boldsymbol{\alpha}} f = \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$$

Multivariate Taylor expansion

- For $d > 1$ the Taylor expansion of f around $\mathbf{x}$ goes

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \mathsf{D} f(\mathbf{x}) + \tfrac{1}{2!}\mathbf{h}^T \mathsf{H} f(\mathbf{x})\mathbf{h} + \tfrac{1}{3!}\sum_{|\boldsymbol{\alpha}|=3} \mathsf{D}^{\boldsymbol{\alpha}} f(\mathbf{x})\mathbf{h}^{\boldsymbol{\alpha}} + \cdots$$

where $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_d) \in \mathbb{N}_0^d$ is a multiindex, $|\boldsymbol{\alpha}| = \sum_i \alpha_i$, the power is $\mathbf{h}^{\boldsymbol{\alpha}} = (h_1^{\alpha_1}, \ldots, h_d^{\alpha_d})$ and

$$\mathsf{D}^{\boldsymbol{\alpha}} f = \frac{\partial^{|\boldsymbol{\alpha}|} f}{\partial x_1^{\alpha_1} \cdots \partial x_d^{\alpha_d}}$$

- How ugly! It's just all 3^rd order partials times the corresponding h's !!

$$\tfrac{1}{3!} \sum_{i,j,k=1}^{d} \frac{\partial^3 f}{\partial x_i \partial x_j \partial x_k}(\mathbf{x}) h_i h_j h_k$$

Multivariate Taylor expansion

- If I want to expand $f(\mathbf{x}+\mathbf{H}\mathbf{z})$ for a matrix $\mathbf{H} \in \mathcal{M}_{d\times d}$ and $\mathbf{z} \in \mathbb{R}^d$, the expression for $(\mathbf{H}\mathbf{z})^{\boldsymbol{\alpha}}$ becomes even uglier!
- Is it possible to go beyond order 2 in a neat way?
- Even if we are satisfied with order 2, how do we expand $\mathsf{D}f(\mathbf{x}+\mathbf{H}\mathbf{z})$?
- The answer is: use the Kronecker product!
  With the usual convention $\frac{\partial}{\partial x_i}\frac{\partial}{\partial x_j} = \frac{\partial^2}{\partial x_i \partial x_j}$ we formally define

$$\mathsf{D}^{\otimes r} f = \frac{\partial^r f}{(\partial \boldsymbol{x})^{\otimes r}} \in \mathbb{R}^{d^r}$$

  the **vector** containing all the partial derivatives of order r.

Multivariate Taylor expansion

- Example: for $r = 2$, we have $\mathsf{D} = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d})$ so that

$$\mathsf{D}^{\otimes 2} = \mathsf{D} \otimes \mathsf{D} = \begin{pmatrix} \frac{\partial}{\partial x_1}\mathsf{D} \\ \hline \vdots \\ \hline \frac{\partial}{\partial x_d}\mathsf{D} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_d} \\ \hline \vdots \\ \hline \frac{\partial}{\partial x_d}\frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_d}\frac{\partial}{\partial x_d} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} \\ \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_d} \\ \hline \vdots \\ \hline \frac{\partial^2}{\partial x_d \partial x_1} \\ \vdots \\ \frac{\partial^2}{\partial x_d^2} \end{pmatrix}$$

Multivariate Taylor expansion

- Example: for $r=2$, we have $\mathsf{D} = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d})$ so that

$$\mathsf{D}^{\otimes 2} = \mathsf{D} \otimes \mathsf{D} = \begin{pmatrix} \frac{\partial}{\partial x_1}\mathsf{D} \\ \hline \vdots \\ \hline \frac{\partial}{\partial x_d}\mathsf{D} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_d} \\ \hline \vdots \\ \hline \frac{\partial}{\partial x_d}\frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_d}\frac{\partial}{\partial x_d} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} \\ \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_d} \\ \hline \vdots \\ \hline \frac{\partial^2}{\partial x_d \partial x_1} \\ \vdots \\ \frac{\partial^2}{\partial x_d^2} \end{pmatrix}$$

- Notice that $\mathsf{D}^{\otimes 2} f = \operatorname{vec} \mathsf{H} f$

Multivariate Taylor expansion

- Example: for $r = 2$, we have $\mathsf{D} = (\frac{\partial}{\partial x_1}, \ldots, \frac{\partial}{\partial x_d})$ so that

$$\mathsf{D}^{\otimes 2} = \mathsf{D} \otimes \mathsf{D} = \begin{pmatrix} \frac{\partial}{\partial x_1}\mathsf{D} \\ \hline \vdots \\ \hline \frac{\partial}{\partial x_d}\mathsf{D} \end{pmatrix} = \begin{pmatrix} \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_1}\frac{\partial}{\partial x_d} \\ \hline \vdots \\ \hline \frac{\partial}{\partial x_d}\frac{\partial}{\partial x_1} \\ \vdots \\ \frac{\partial}{\partial x_d}\frac{\partial}{\partial x_d} \end{pmatrix} = \begin{pmatrix} \frac{\partial^2}{\partial x_1^2} \\ \vdots \\ \frac{\partial^2}{\partial x_1 \partial x_d} \\ \hline \vdots \\ \hline \frac{\partial^2}{\partial x_d \partial x_1} \\ \vdots \\ \frac{\partial^2}{\partial x_d^2} \end{pmatrix}$$

- Notice that $\mathsf{D}^{\otimes 2} f = \operatorname{vec} \mathsf{H} f$
- But we already knew that! Because the Hessian matrix is $\mathsf{H} = \mathsf{D}\mathsf{D}^T$, so that

$$\operatorname{vec} \mathsf{H} = \operatorname{vec}(\mathsf{D}\mathsf{D}^T) = \mathsf{D} \otimes \mathsf{D}$$

Multivariate Taylor expansion

- Now we can write our Taylor expansion in a unified way:

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \mathsf{D} f(\mathbf{x}) + \tfrac{1}{2!}(\mathbf{h}^{\otimes 2})^T \mathsf{D}^{\otimes 2} f(\mathbf{x}) + \tfrac{1}{3!}(\mathbf{h}^{\otimes 3})^T \mathsf{D}^{\otimes 3} f(\mathbf{x}) + \cdots$$

Multivariate Taylor expansion

- Now we can write our Taylor expansion in a unified way:

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \mathsf{D} f(\mathbf{x}) + \tfrac{1}{2!}(\mathbf{h}^{\otimes 2})^T \mathsf{D}^{\otimes 2} f(\mathbf{x}) + \tfrac{1}{3!}(\mathbf{h}^{\otimes 3})^T \mathsf{D}^{\otimes 3} f(\mathbf{x}) + \cdots$$

- Moreover, since $(\mathbf{H}\mathbf{z})^{\otimes r} = \mathbf{H}^{\otimes r}\mathbf{z}^{\otimes r}$ we have

$$f(\mathbf{x}+\mathbf{H}\mathbf{z}) = f(\mathbf{x}) + \mathbf{z}^T\mathbf{H}^T \mathsf{D} f(\mathbf{x}) + \tfrac{1}{2!}(\mathbf{z}^{\otimes 2})^T(\mathbf{H}^{\otimes 2})^T \mathsf{D}^{\otimes 2} f(\mathbf{x}) + \tfrac{1}{3!}(\mathbf{z}^{\otimes 3})^T(\mathbf{H}^{\otimes 3})^T \mathsf{D}^{\otimes 3} f(\mathbf{x}) + \cdots$$

Multivariate Taylor expansion

- Now we can write our Taylor expansion in a unified way:

$$f(\mathbf{x}+\mathbf{h}) = f(\mathbf{x}) + \mathbf{h}^T \mathsf{D} f(\mathbf{x}) + \tfrac{1}{2!}(\mathbf{h}^{\otimes 2})^T \mathsf{D}^{\otimes 2} f(\mathbf{x}) \\ + \tfrac{1}{3!}(\mathbf{h}^{\otimes 3})^T \mathsf{D}^{\otimes 3} f(\mathbf{x}) + \cdots$$

- Moreover, since $(\mathbf{H}\mathbf{z})^{\otimes r} = \mathbf{H}^{\otimes r}\mathbf{z}^{\otimes r}$ we have

$$f(\mathbf{x}+\mathbf{H}\mathbf{z}) = f(\mathbf{x}) + \mathbf{z}^T\mathbf{H}^T \mathsf{D} f(\mathbf{x}) + \tfrac{1}{2!}(\mathbf{z}^{\otimes 2})^T(\mathbf{H}^{\otimes 2})^T \mathsf{D}^{\otimes 2} f(\mathbf{x}) \\ + \tfrac{1}{3!}(\mathbf{z}^{\otimes 3})^T(\mathbf{H}^{\otimes 3})^T \mathsf{D}^{\otimes 3} f(\mathbf{x}) + \cdots$$

- Furthermore, we can write

$$\mathsf{D}^{\otimes r} f(\mathbf{x}+\boldsymbol{h}) = \mathsf{D}^{\otimes r} f(\mathbf{x}) + \left(\mathbf{I}_{d^r} \otimes \boldsymbol{h}^T\right)\mathsf{D}^{\otimes r+1} f(\mathbf{x}) \\ + \tfrac{1}{2!}\left[\mathbf{I}_{d^r} \otimes (\boldsymbol{h}^{\otimes 2})^T\right]\mathsf{D}^{\otimes r+2} f(\mathbf{x}) + \cdots$$

Derivatives of the multivariate normal density

- We denote the d-variate normal density by

$$\phi(\mathbf{x}) = (2\pi)^{-d/2} \exp\{-\tfrac{1}{2}\mathbf{x}^T\mathbf{x}\} \quad \text{for } \mathbf{x} \in \mathbb{R}^d$$

- If $d = 1$ it is well known that $\phi^{(r)}(x) = (-1)^r \phi(x) \mathcal{H}_r(x)$, where

$$\mathcal{H}_r(x) = r! \sum_{j=0}^{[r/2]} \frac{(-1)^j}{(r-2j)!j!2^j} x^{r-2j}$$

denotes the rth **Hermite polynomial**

Derivatives of the multivariate normal density

- We denote the d-variate normal density by

$$\phi(\mathbf{x}) = (2\pi)^{-d/2} \exp\{-\tfrac{1}{2}\mathbf{x}^T\mathbf{x}\} \quad \text{for } \mathbf{x} \in \mathbb{R}^d$$

- If $d = 1$ it is well known that $\phi^{(r)}(x) = (-1)^r \phi(x) \mathcal{H}_r(x)$, where

$$\mathcal{H}_r(x) = r! \sum_{j=0}^{[r/2]} \frac{(-1)^j}{(r-2j)!j!2^j} x^{r-2j}$$

  denotes the rth **Hermite polynomial**

- Concise expressions for $d > 1$ were not obtained until Holmquist (1996), who showed that $\mathsf{D}^{\otimes r}\phi(\mathbf{x}) = (-1)^r \phi(\mathbf{x}) \boldsymbol{\mathcal{H}}_r(\mathbf{x})$, where

$$\boldsymbol{\mathcal{H}}_r(\mathbf{x}) = r! \boldsymbol{\mathcal{S}}_{d,r} \sum_{j=0}^{[r/2]} \frac{(-1)^j}{(r-2j)!j!2^j} \left\{\mathbf{x}^{\otimes(r-2j)} \otimes (\operatorname{vec} \mathbf{I}_d)^{\otimes j}\right\}$$

  denotes the rth **vector Hermite polynomial**

Derivatives of the multivariate normal density

- The **symmetrizer matrix** $\boldsymbol{\mathcal{S}}_{d,r} \in \mathcal{M}_{d^r \times d^r}$ is defined by

$$\boldsymbol{\mathcal{S}}_{d,r} \bigotimes_{i=1}^{r} \boldsymbol{v}_i = \frac{1}{r!} \sum_{\sigma \in \mathcal{P}_r} \bigotimes_{i=1}^{r} \boldsymbol{v}_{\sigma(i)},$$

  where $\mathcal{P}_r$ denotes the set of permutations of r elements

Derivatives of the multivariate normal density

- The **symmetrizer matrix** $\boldsymbol{\mathcal{S}}_{d,r} \in \mathcal{M}_{d^r \times d^r}$ is defined by

$$\boldsymbol{\mathcal{S}}_{d,r} \bigotimes_{i=1}^{r} \boldsymbol{v}_i = \frac{1}{r!} \sum_{\sigma \in \mathcal{P}_r} \bigotimes_{i=1}^{r} \boldsymbol{v}_{\sigma(i)},$$

  where $\mathcal{P}_r$ denotes the set of permutations of r elements
- For example,

$$\begin{aligned}
\boldsymbol{\mathcal{S}}_{d,1} &= \mathbf{I}_d \\
\boldsymbol{\mathcal{S}}_{d,2} &= \tfrac{1}{2}(\mathbf{I}_{d^2} + \mathbf{K}_{d,d}), \text{ where } \mathbf{K}_{d,d}(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2) = \boldsymbol{v}_2 \otimes \boldsymbol{v}_1 \\
3!\boldsymbol{\mathcal{S}}_{d,3}(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3) &= \boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3 + \boldsymbol{v}_1 \otimes \boldsymbol{v}_3 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_1 \otimes \boldsymbol{v}_3 \\
&\quad + \boldsymbol{v}_2 \otimes \boldsymbol{v}_3 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_3 \otimes \boldsymbol{v}_1 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_3 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_1
\end{aligned}$$

Derivatives of the multivariate normal density

- The **symmetrizer matrix** $\boldsymbol{\mathcal{S}}_{d,r} \in \mathcal{M}_{d^r \times d^r}$ is defined by

$$\boldsymbol{\mathcal{S}}_{d,r} \bigotimes_{i=1}^{r} \boldsymbol{v}_i = \frac{1}{r!} \sum_{\sigma \in \mathcal{P}_r} \bigotimes_{i=1}^{r} \boldsymbol{v}_{\sigma(i)},$$

  where $\mathcal{P}_r$ denotes the set of permutations of r elements
- For example,

$$\begin{aligned}
\boldsymbol{\mathcal{S}}_{d,1} &= \mathbf{I}_d \\
\boldsymbol{\mathcal{S}}_{d,2} &= \tfrac{1}{2}(\mathbf{I}_{d^2} + \mathbf{K}_{d,d}), \text{ where } \mathbf{K}_{d,d}(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2) = \boldsymbol{v}_2 \otimes \boldsymbol{v}_1 \\
3!\boldsymbol{\mathcal{S}}_{d,3}(\boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3) &= \boldsymbol{v}_1 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_3 + \boldsymbol{v}_1 \otimes \boldsymbol{v}_3 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_2 \otimes \boldsymbol{v}_1 \otimes \boldsymbol{v}_3 \\
&\quad + \boldsymbol{v}_2 \otimes \boldsymbol{v}_3 \otimes \boldsymbol{v}_1 + \boldsymbol{v}_3 \otimes \boldsymbol{v}_1 \otimes \boldsymbol{v}_2 + \boldsymbol{v}_3 \otimes \boldsymbol{v}_2 \otimes \boldsymbol{v}_1
\end{aligned}$$

- If $\{\boldsymbol{e}_1, \dots, \boldsymbol{e}_d\}$ denotes the canonical basis of $\mathbb{R}^d$ then

$$\boldsymbol{\mathcal{S}}_{d,r} = \frac{1}{r!} \sum_{k_1=1}^{d} \cdots \sum_{k_r=1}^{d} \sum_{\sigma \in \mathcal{P}_r} \bigotimes_{i=1}^{r} (\boldsymbol{e}_{k_i} \boldsymbol{e}_{k_{\sigma(i)}}^T)$$

Iterative computation of the symmetrizer matrix

- Computing the symmetrizer matrix using formula

$$\boldsymbol{\mathcal{S}}_{d,r} = \frac{1}{r!} \sum_{k_1=1}^{d} \cdots \sum_{k_r=1}^{d} \sum_{\sigma} \bigotimes_{i=1}^{r} (\boldsymbol{e}_{k_i} \boldsymbol{e}_{k_{\sigma(i)}}^T)$$

is extremely inefficient ($d = 2$, $r = 8$ takes more than 24 hours!)

Iterative computation of the symmetrizer matrix

- Computing the symmetrizer matrix using formula

$$\boldsymbol{\mathcal{S}}_{d,r} = \frac{1}{r!}\sum_{k_1=1}^{d}\cdots\sum_{k_r=1}^{d}\sum_{\sigma}\bigotimes_{i=1}^{r}(\boldsymbol{e}_{k_i}\boldsymbol{e}_{k_{\sigma(i)}}^T)$$

  is extremely inefficient ($d = 2$, $r = 8$ takes more than 24 hours!)
- We developed an iterative approach to the computation of $\boldsymbol{\mathcal{S}}_{d,r}$
- Since $\boldsymbol{\mathcal{S}}_{d,1} = \mathbf{I}_d$ our goal was to express $\boldsymbol{\mathcal{S}}_{d,r+1}$ as a function of $\boldsymbol{\mathcal{S}}_{d,r}$
- We found that $\boldsymbol{\mathcal{S}}_{d,r+1} = (\boldsymbol{\mathcal{S}}_{d,r}\otimes\mathbf{I}_d)\cdot\boldsymbol{\mathcal{T}}_{d,r+1}$, where

$$\boldsymbol{\mathcal{T}}_{d,r} = \frac{1}{r}\sum_{j=1}^{r}(\mathbf{I}_{d^j}\otimes\mathbf{K}_{d^{r-j-1},d})(\mathbf{I}_{d^{j-1}}\otimes\mathbf{K}_{d,d^{r-j}})$$

Iterative computation of the symmetrizer matrix

- Computing the symmetrizer matrix using formula

$$\boldsymbol{\mathcal{S}}_{d,r} = \frac{1}{r!}\sum_{k_1=1}^{d}\cdots\sum_{k_r=1}^{d}\sum_{\sigma}\bigotimes_{i=1}^{r}(\boldsymbol{e}_{k_i}\boldsymbol{e}_{k_{\sigma(i)}}^T)$$

  is extremely inefficient ($d = 2$, $r = 8$ takes more than 24 hours!)
- We developed an iterative approach to the computation of $\boldsymbol{\mathcal{S}}_{d,r}$
- Since $\boldsymbol{\mathcal{S}}_{d,1} = \mathbf{I}_d$ our goal was to express $\boldsymbol{\mathcal{S}}_{d,r+1}$ as a function of $\boldsymbol{\mathcal{S}}_{d,r}$
- We found that $\boldsymbol{\mathcal{S}}_{d,r+1} = (\boldsymbol{\mathcal{S}}_{d,r} \otimes \mathbf{I}_d)\cdot\boldsymbol{\mathcal{T}}_{d,r+1}$, where

$$\boldsymbol{\mathcal{T}}_{d,r} = \frac{1}{r}\sum_{j=1}^{r}(\mathbf{I}_{d^j}\otimes\mathbf{K}_{d^{r-j-1},d})(\mathbf{I}_{d^{j-1}}\otimes\mathbf{K}_{d,d^{r-j}})$$

- As a consequence, the symmetrizer matrix can be factorized as

$$\boldsymbol{\mathcal{S}}_{d,r} = \prod_{i=1}^{r}(\mathbf{T}_{d,i}\otimes\mathbf{I}_{d^{r-i}}) = (\mathbf{T}_{d,1}\otimes\mathbf{I}_{d^{r-1}})(\mathbf{T}_{d,2}\otimes\mathbf{I}_{d^{r-2}})\cdots(\mathbf{T}_{d,r}\otimes\mathbf{I}_{d^0})$$

Iterative computation of the $\mathcal{T}_{d,r}$ matrix

- The matrix

$$\mathcal{T}_{d,r} = \frac{1}{r}\sum_{j=1}^{r}(\mathbf{I}_{d^j} \otimes \mathbf{K}_{d^{r-j-1},d})(\mathbf{I}_{d^{j-1}} \otimes \mathbf{K}_{d,d^{r-j}})$$

is much easier to compute than $\boldsymbol{\mathcal{S}}_{d,r}$

Iterative computation of the $\mathcal{T}_{d,r}$ matrix

- The matrix

$$\mathcal{T}_{d,r} = \frac{1}{r}\sum_{j=1}^{r}(\mathbf{I}_{d^j} \otimes \mathbf{K}_{d^{r-j-1},d})(\mathbf{I}_{d^{j-1}} \otimes \mathbf{K}_{d,d^{r-j}})$$

  is much easier to compute than $\mathcal{S}_{d,r}$
- However, it still requires some time
- Since $\mathcal{T}_{d,1} = \mathbf{I}_d$ we found a faster, recursive formula for $\mathcal{T}_{d,r+1}$ in terms of $\mathcal{T}_{d,r}$

$$\mathcal{T}_{d,r+1} = \tfrac{r}{r+1}(\mathbf{I}_{d^{r-1}} \otimes \mathbf{K}_{d,d})(\mathcal{T}_{d,r} \otimes \mathbf{I}_d + \mathbf{I}_{d^{r-1}} \otimes \mathbf{K}_{d,d})(\mathbf{I}_{d^{r-1}} \otimes \mathbf{K}_{d,d})$$

Further applications of multivariate Hermite polynomials

- If $\mathbf{X}$ is a random variable with distribution $N_d(\boldsymbol{\mu}, \boldsymbol{\Sigma})$ then

$$\mathbb{E}\big[\mathbf{X}^{\otimes r}\big] = r!\boldsymbol{\mathcal{S}}_{d,r} \sum_{j=0}^{[r/2]} \frac{\boldsymbol{\mu}^{\otimes(r-2j)} \otimes (\operatorname{vec} \boldsymbol{\Sigma})^{\otimes j}}{(r-2j)!j!2^j}$$

- Especially, when $\boldsymbol{\mu} = 0$,

$$\mathbb{E}\big[\mathbf{X}^{\otimes 2r}\big] = \frac{(2r)!}{r!2^r}\boldsymbol{\mathcal{S}}_{d,2r}(\operatorname{vec} \boldsymbol{\Sigma})^{\otimes r}$$

- If $\mathbf{A}$ is a symmetric matrix, the rth moment of the quadratic form $q = \mathbf{X}^T\mathbf{A}\mathbf{X}$ can be written as

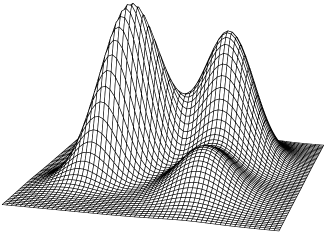
$$\mathbb{E}[q^r] = (2r)!(\operatorname{vec}^T \mathbf{A})^{\otimes r}\boldsymbol{\mathcal{S}}_{d,2r} \sum_{j=0}^{r} \frac{\boldsymbol{\mu}^{\otimes(2r-2j)} \otimes (\operatorname{vec} \boldsymbol{\Sigma})^{\otimes j}}{(2r-2j)!j!2^j}$$

What is a density?

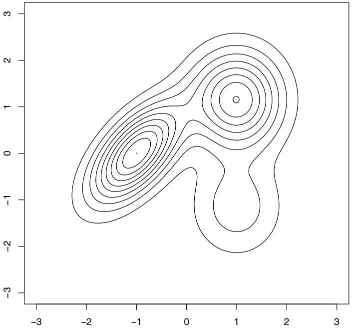
- Density: positive function $f\colon \mathbb{R}^d \to \mathbb{R}$ such that $\int_{\mathbb{R}^d} f = 1$
- Most graphical way of representing the distribution of a random variable $\mathbf{X}$, $\mathbb{P}(\mathbf{X} \in A) = \int_A f$, for any Borel set $A \subset \mathbb{R}^d$

What is a density?

- Density: positive function $f\colon \mathbb{R}^d \to \mathbb{R}$ such that $\int_{\mathbb{R}^d} f = 1$
- Most graphical way of representing the distribution of a random variable $\mathbf{X}$, $\mathbb{P}(\mathbf{X} \in A) = \int_A f$, for any Borel set $A \subset \mathbb{R}^d$



Perspective plot



Contour plot

Kernel density estimation

$\mathbf{X}_1, \ldots, \mathbf{X}_n$ iid with density $f\colon \mathbb{R}^d \to \mathbb{R} \rightsquigarrow$ **Kernel density estimator**:

$$\hat{f}_{n\mathbf{H}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$$

Kernel density estimation

$\mathbf{X}_1, \ldots, \mathbf{X}_n$ iid with density $f\colon \mathbb{R}^d \to \mathbb{R} \rightsquigarrow$ **Kernel density estimator**:

$$\hat{f}_{n\mathbf{H}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$$

- K kernel: $K \geq 0$, $\int_{\mathbb{R}^d} K = 1$ (usually K symmetric)
- $\mathbf{H}$ bandwidth matrix:
  $\mathbf{H} \in \mathcal{F} = \{$symmetric positive definite $d \times d$ matrices$\}$
- $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x})$
  $\mathbf{H}^{1/2} \equiv$ matrix square root of $\mathbf{H}$, $\mathbf{H}^{-1/2} = (\mathbf{H}^{1/2})^{-1} = (\mathbf{H}^{-1})^{1/2}$

Kernel density estimation

$\mathbf{X}_1, \ldots, \mathbf{X}_n$ iid with density $f\colon \mathbb{R}^d \to \mathbb{R} \rightsquigarrow$ **Kernel density estimator**:

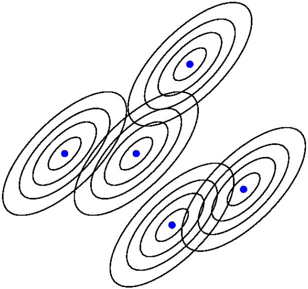
$$\hat{f}_{n\mathbf{H}}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{\mathbf{H}}(\mathbf{x} - \mathbf{X}_i)$$

- K kernel: $K \geq 0$, $\int_{\mathbb{R}^d} K = 1$ (usually K symmetric)
- $\mathbf{H}$ bandwidth matrix:
  $\mathbf{H} \in \mathcal{F} = \{$symmetric positive definite $d \times d$ matrices$\}$
- $K_{\mathbf{H}}(\mathbf{x}) = |\mathbf{H}|^{-1/2} K(\mathbf{H}^{-1/2}\mathbf{x})$
  $\mathbf{H}^{1/2} \equiv$ matrix square root of $\mathbf{H}$, $\mathbf{H}^{-1/2} = (\mathbf{H}^{1/2})^{-1} = (\mathbf{H}^{-1})^{1/2}$

– Ex.: K density of $N(\mathbf{0}, \mathbf{I}_d) \rightsquigarrow K_{\mathbf{H}}$ density of $N(\mathbf{0}, \mathbf{H})$

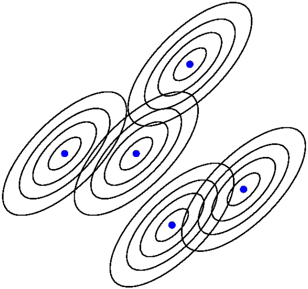
Kernel density estimation

Individual kernels

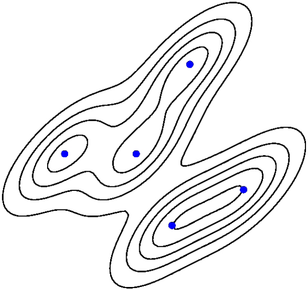


Kernel density estimation

Individual kernels



Kernel density estimator



Performance of the kernel estimator

- The choice of the bandwidth matrix $\mathbf{H}$ is crucial for the performance of the kernel estimator in practice
- We use the **Mean Integrated Squared Error** criterion

$$\mathrm{MISE}(\mathbf{H}) = \mathbb{E} \int (\hat{f}_{n\mathbf{H}} - f)^2$$

  to measure the accuracy of $\hat{f}_{n\mathbf{H}}$ as an estimator of f
- It can be shown that asymptotically (when $n \to \infty$) this MISE function is well approximated by

$$A(\mathbf{H}) = \alpha |\mathbf{H}|^{-1/2} + \beta (\operatorname{vec} \mathbf{H})^T \mathbf{Q} \operatorname{vec} \mathbf{H}$$

  where $\alpha, \beta > 0$ and $\mathbf{Q}$ is a symmetric positive semi-definite matrix

$$\mathbf{Q} = \int \mathsf{D}^{\otimes 2} f(\mathbf{x}) \mathsf{D}^{\otimes 2} f(\mathbf{x})^T d\mathbf{x} \in \mathcal{M}_{d^2 \times d^2}$$

  containing all possible combinations of the real quantities

$$\int \frac{\partial^2 f}{\partial x_i \partial x_j}(\mathbf{x}) \frac{\partial^2 f}{\partial x_k \partial x_l}(\mathbf{x}) d\mathbf{x} \in \mathbb{R}$$

Optimal bandwidth choice

- We would like to have an explicit expression for the **optimal bandwidth** $\mathbf{H}_0 = \underset{\mathbf{H}\in\mathcal{F}}{\operatorname{argmin}} A(\mathbf{H})$ where

$$A(\mathbf{H}) = \alpha|\mathbf{H}|^{-1/2} + \beta(\operatorname{vec}\mathbf{H})^T\mathbf{Q}\operatorname{vec}\mathbf{H}$$

  with α, $\beta > 0$ and $\mathbf{Q}$ a symmetric positive semi-definite matrix
- If we restrict $\mathbf{H}$ to be in $\mathcal{S} = \{h^2\mathbf{I}_d \colon h > 0\}$ then

$$A(h^2\mathbf{I}_d) = \alpha h^{-d} + \beta[(\operatorname{vec}\mathbf{I}_d)^T\mathbf{Q}\operatorname{vec}\mathbf{I}_d]h^4$$

  so we easily obtain $\mathbf{H}_0 = h_0^2\mathbf{I}_d$ with

$$h_0 = \left(\frac{d\alpha}{4\beta(\operatorname{vec}\mathbf{I}_d)^T\mathbf{Q}\operatorname{vec}\mathbf{I}_d}\right)^{1/(d+4)}$$

Optimal bandwidth choice

- We would like to have an explicit expression for the **optimal bandwidth** $\mathbf{H}_0 = \operatorname*{argmin}_{\mathbf{H}\in\mathcal{F}} A(\mathbf{H})$ where

$$A(\mathbf{H}) = \alpha|\mathbf{H}|^{-1/2} + \beta(\operatorname{vec}\mathbf{H})^T\mathbf{Q}\operatorname{vec}\mathbf{H}$$

  with $\alpha,\ \beta > 0$ and $\mathbf{Q}$ a symmetric positive semi-definite matrix
- If we restrict $\mathbf{H}$ to be in $\mathcal{D} = \{\operatorname{diag}(h_1^2,\dots,h_d^2)\colon h_i > 0,\ \forall i\}$ then a explicit solution exists for $d = 2$:

$$A(h_1,h_2) = \alpha h_1^{-1}h_2^{-1} + \beta\big(q_{11}h_1^4 + 2q_{41}h_1^2h_2^2 + q_{44}h_2^4\big)$$

  and we obtain $\mathbf{H}_0 = \operatorname{diag}(h_{10}^2, h_{20}^2)$ with $h_{20} = c_0h_{10}$, $c_0 = (q_{11}/q_{44})^{1/4}$ and

$$h_{10} = \left(\frac{\alpha}{4\beta(q_{11}c + q_{41}c^3)}\right)^{1/6}$$

Optimal bandwidth choice

- In general, we can write $\mathbf{H} = |\mathbf{H}|^{1/d}\mathbf{M}$, with $|\mathbf{M}| = 1$
- Then
$$A(\mathbf{H}) = \alpha|\mathbf{H}|^{-1/2} + \beta|\mathbf{H}|^{2/d}(\operatorname{vec}\mathbf{M})^T\mathbf{Q}\operatorname{vec}\mathbf{M}$$
so that it is enough to minimize
$$(\operatorname{vec}\mathbf{M})^T\mathbf{Q}\operatorname{vec}\mathbf{M} \quad \text{subject to } |\mathbf{M}| = 1$$
- If we could find a "Kronecker square root" $\mathbf{R} = \mathbf{Q}^{\otimes 1/2}$ then we could write
$$(\operatorname{vec}\mathbf{M})^T\mathbf{Q}\operatorname{vec}\mathbf{M} = (\operatorname{vec}\mathbf{M})^T(\mathbf{R}\otimes\mathbf{R})\operatorname{vec}\mathbf{M} = \operatorname{tr}\{\mathbf{RMRM}\}$$
and it is possible to show then that the optimal $\mathbf{M}$ has explicit form
$$\mathbf{M}_0 = |\mathbf{R}|^{1/d}\mathbf{R}^{-1}$$

Optimal bandwidth choice

- In general, we can write $\mathbf{H} = |\mathbf{H}|^{1/d}\mathbf{M}$, with $|\mathbf{M}| = 1$
- Then
$$A(\mathbf{H}) = \alpha|\mathbf{H}|^{-1/2} + \beta|\mathbf{H}|^{2/d}(\operatorname{vec}\mathbf{M})^T\mathbf{Q}\operatorname{vec}\mathbf{M}$$
so that it is enough to minimize
$$(\operatorname{vec}\mathbf{M})^T\mathbf{Q}\operatorname{vec}\mathbf{M} \quad \text{subject to } |\mathbf{M}| = 1$$
- If we could find a “Kronecker square root” $\mathbf{R} = \mathbf{Q}^{\otimes 1/2}$ then we could write
$$(\operatorname{vec}\mathbf{M})^T\mathbf{Q}\operatorname{vec}\mathbf{M} = (\operatorname{vec}\mathbf{M})^T(\mathbf{R}\otimes\mathbf{R})\operatorname{vec}\mathbf{M} = \operatorname{tr}\{\mathbf{RMRM}\}$$
and it is possible to show then that the optimal $\mathbf{M}$ has explicit form
$$\mathbf{M}_0 = |\mathbf{R}|^{1/d}\mathbf{R}^{-1}$$
- Necessary conditions for $\mathbf{R}$ to exist? Practical calculation?