A Moment-Based Bivariate Density Estimation Methodology for Large Data Sets

Hossein Zareamoghaddam

Supervisor: S.B. Provost (Western University)

SSC 2016

Outline

Moment-based density approximants via standard and orthogonal polynomial adjustments

- Density approximant
  - Univariate case
  - Bivariate case
  - Equivalence between standard and orthogonal polynomial adjustments
- Density estimate
- Examples
- Applications to regression

Univariate density Approximation

Let $\qquad f_X(x) \simeq f_{X_n}(x) = c_T\,\omega(x)\sum_{k=0}^{n}\theta_k x^k$ where

- $\omega(x)$ is the weigth function so that $\psi(x) = c_T\,\omega(x)$ is an initial (base) density function
- Objective: determining $\theta_k,\ k = 0, \ldots, n$
- $\mu(j)$ is the given j^{th} moment associated with $f_X(x)$
- $m(j)$ is the j^{th} moment associated with the initial density $\psi(x)$

By equating $\int x^j f_X(x)\mathrm{d}x$ to $\int x^j f_{X_n}(x)\mathrm{d}x$, linear equations of the form $\mu(j) = \sum_{i=j}^{j+n}\theta_j m(i)$ are obtained for $j = 0, \ldots, n$.

The density approximant is obtained by solving the linear system

$$\boldsymbol{\mu} = \mathbf{M}\,\boldsymbol{\theta}.$$

Orthogonal polynomials and density approximation

Definition: Let $\{T_i(x) = \sum_{k=0}^{i} \alpha_{ik} x^k, i = 0, 1, \ldots, n\}$ be a set of orthogonal polynomials defined on (a_0, b_0) w.r.t. the weight function $\omega(x)$ which is proportional to a given base density function $\psi(x) \equiv c_T\, \omega(x)$. Then

$$\int \omega(x) T_i(x) T_j(x) \mathrm{d}x = \theta_i\, \delta_{ij}.$$

Let the approximate pdf have the following form

$$f_{X_n}(x) = c_T\, \omega(x) \sum_{k=0}^{n} \eta_k T_k(x).$$

The coefficients η_k's are determined by equating $\int_{a_0}^{b_0} T_h(x) f_{X_n}(x) \mathrm{d}x$ to $\int_{a_0}^{b_0} T_h(x) f_X(x) \mathrm{d}x$ for $h = 0, 1, \ldots, n$:

$$c_T \int_{a_0}^{b_0} T_h(x) \omega(x) \sum_{k=0}^{n} \eta_k T_k(x) \mathrm{d}x = \int_{a_0}^{b_0} T_h(x) f_X(x) \mathrm{d}x, \quad h = 0, 1, 2, \ldots, n$$

$$\sum_{k=0}^{n} c_T\, \eta_k \int_{a_0}^{b_0} \omega(x)\, T_h(x)\, T_k(x) \mathrm{d}x = \sum_{k=0}^{h} \alpha_{hk}\, \mu(k), \quad h = 0,\, 1,\, 2,\, \ldots,\, n,$$

where α_{hk} is the coefficient of x^k in $T_h(x)$ and $\mu(k)$ is the k^{th} moment of the target distribution.

From the orthogonality of the T_i's, we have:

$$\eta_h = \frac{1}{c_T \theta_h} \Big\{ \sum_{k=0}^{h} \alpha_{hk} \mu(k) \Big\}, \quad h = 0,\, 1,\, 2,\, \ldots,\, n.$$

Then,

$$f_X(x) \simeq f_{X_n}(x) = c_T\, \omega(x) \sum_{k=0}^{n} \eta_k\, T_k(x).$$

- Hermite polynomials:

$$H_n(x) = (-1)^n e^{\frac{x^2}{2}} \frac{\mathrm{d}^n}{\mathrm{d}x^n} e^{-\frac{x^2}{2}}, \; -\infty < x < \infty, \quad n = 0, 1, 2, \ldots$$

Orthogonality: $\int_{-\infty}^{\infty} e^{-x^2/2} H_i(x)\, H_j(x)\, \mathrm{d}x = \sqrt{2\pi} j!\, \delta_{ij}$.

Set $\omega(x) = e^{-\frac{x^2}{2}}$ and $c_T = \frac{1}{\sqrt{2\pi}}$, so that the base density is

$$\psi_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, \; -\infty < x < \infty,$$

its j^{th} moment being

$$m(j) = \frac{2^{\frac{1}{2}(j-1)}(1 + (-1)^j)\Gamma(\frac{j+1}{2})}{\sqrt{2\pi}}, \quad j = 0, 1, \ldots .$$

Hermite polynomials are suitable for approximating normal-type distributions

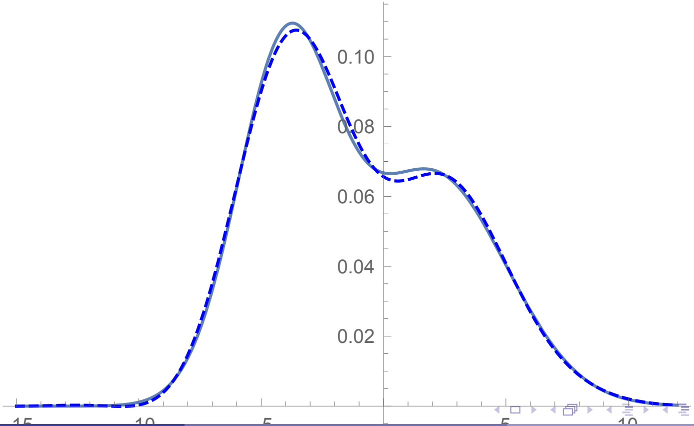
- Legendre, Laguerre and Jacobi polynomials

Example 1:

Suppose $Y_1 \sim \mathcal{N}(-4, 2)$ & $Y_2 \sim \mathcal{N}(2, 3)$ and

$$f_Y(y) = \frac{1}{2}(f_{Y_1}(y) + f_{Y_2}(y))$$

Figure: Univariate Mixture of Normal Densities



Bivariate density approximation

Let $f_{\mathbf{Z}}(\mathbf{z}) = \Psi(\mathbf{z})p(\mathbf{z})$, where

- $\mathbf{Z} = (X, Y)'$
- $\Psi(\mathbf{z})$: bivariate base density
- $p(\mathbf{z}) = \sum_{i=0}^{\infty}\sum_{j=0}^{\infty} c_{i,j}x^i y^j$ is the polynomial adjustment

Then,

$$f_{\mathbf{Z}}(\mathbf{z}) \simeq f_{\mathbf{Z_n}}(\mathbf{z}) = \Psi(\mathbf{z})\sum_{i=0}^{n}\sum_{j=0}^{n} c_{i,j}x^i y^j = \Psi(\mathbf{z})\,p_{n,n}(\mathbf{z}).$$

Letting
$$\begin{aligned}\mu(k,\ell) &\equiv \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell f_{\mathbf{Z}}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell\,\Psi(\mathbf{z})\,p_{n,n}(\mathbf{z})\,\mathrm{d}\mathbf{z}\\ &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} x^k y^\ell \Psi(\mathbf{z})\sum_{i=0}^{n}\sum_{j=0}^{n} c_{i,j}x^i y^j \mathrm{d}\mathbf{z}\\ &= \sum_{i=0}^{n}\sum_{j=0}^{n} c_{i,j}m(k+i,\,\ell+j)\end{aligned}$$

Bivariate density approximation

Thus,

$$\mu(k,\ell) = \sum_{i=0}^{n}\sum_{j=0}^{n} c_{i,j} m(k+i,\, \ell+j), \qquad k,\ell = 0,\, 1,\, \ldots,\, n \tag{1}$$

where $m(i,j)$ denotes the $(i,j)^{\text{th}}$ joint moments associated with the base density.

The $c_{i,j}$'s, $i,j = 0,\, 1,\, \ldots,\, n$, are obtained by solving these $(n+1)^2$ linear equations which can be expressed in matrix form as

$$M_{(n+1)^2\times(n+1)^2}\,\mathbf{c} = \boldsymbol{\mu}_{(n+1)^2\times 1}\,.$$

Example 2:

Let $f_{\mathbf{Z}}(\mathbf{z}) = \frac{1}{2}(f_{\mathbf{X}}(\mathbf{z}) + f_{\mathbf{Y}}(\mathbf{z}))$ where $\mathbf{z} = (x, y)'$ with

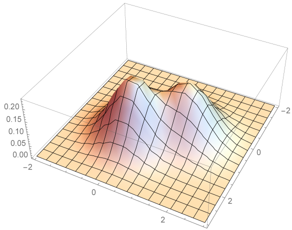
$$\mathbf{X} \sim \mathcal{N}_2\left(\begin{pmatrix} 1.1 \\ -0.1 \end{pmatrix}, \begin{pmatrix} 0.33 & 0.03 \\ 0.03 & 0.33 \end{pmatrix}\right)$$

and

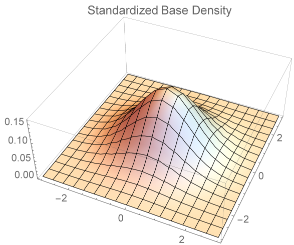
$$\mathbf{Y} \sim \mathcal{N}_2\left(\begin{pmatrix} 0.2 \\ 1.2 \end{pmatrix}, \begin{pmatrix} 0.4 & 0.04 \\ 0.04 & 0.4 \end{pmatrix}\right)$$

Let the distribution of the base density function be

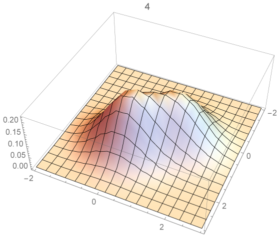
$$\mathcal{N}_2\left(\begin{pmatrix} 0.65 \\ 0.55 \end{pmatrix}, \begin{pmatrix} 0.5675 & -0.2575 \\ -0.2575 & 0.7875 \end{pmatrix}\right).$$



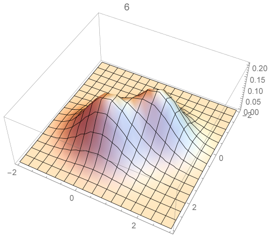
(a) Exact density



(b) Base density



(c) $n = 4$



(d) $n = 6$

Figure: Exact, base and approximate densities

Bivariate Hermite polynomials (BHP's)

Let $\mathbf{z} = (x,\ y)'$ and $\Lambda_{2\times 2}$ be a positive definite matrix. The bivariate orthogonal Hermite polynomials of orders r_1 and r_2 $\ (r = r_1 + r_2)$ associated with a $\mathcal{N}_2(\mathbf{0},\ A)$ distribution where $A = \Lambda^{-1}$ (see (Willink, 2005) and (C.S. Withers and S. Nadarajah, 2010)) is given by

$$H_{r_1,\,r_2}(\mathbf{z},\ \Lambda) \equiv (-1)^r exp(\frac{1}{2}\mathbf{z}' A\,\mathbf{z})\frac{\partial^r}{\partial x^{r_1}\partial y^{r_2}} exp(-\frac{1}{2}\mathbf{z}' A\,\mathbf{z}).$$

For example, when $\Lambda = \begin{pmatrix} 2 & 1 \\ 1 & 3 \end{pmatrix}$,

$H_{1,1}(\mathbf{z},\ \Lambda) = \frac{1}{5} - \frac{3x^2}{25} + \frac{7xy}{25} - \frac{2y^2}{25}$

$H_{1,2}(\mathbf{z},\ \Lambda) = -\frac{8x}{25} + \frac{3x^3}{125} + \frac{6y}{25} - \frac{13x^2y}{125} + \frac{16xy^2}{125} - \frac{4y^3}{125}$

$H_{2,2}(\mathbf{z},\ \Lambda) = \frac{8}{25} - \frac{33x^2}{125} + \frac{9x^4}{625} + \frac{52xy}{125} - \frac{42x^3y}{625} - \frac{22y^2}{125} + \frac{61x^2y^2}{625} - \frac{28xy^3}{625} + \frac{4y^4}{625}$

- Let $\{H_{i,j}(\mathbf{z}) = \sum_{\ell=0}^{j} \sum_{k=0}^{i} \alpha_{ijk\ell} x^k y^\ell\}$ be a BHP. $H^*_{u,v}(\mathbf{z}) = \sum_{n=0}^{v} \sum_{m=0}^{u} \alpha^*_{uvnm} x^m y^n$ called the dual BHP of orders u and v with respect to the bivariate function $\Psi(\mathbf{x})$ defined on $\mathbb{R}^2$ satisfies the equation,

$$\int_{\mathbb{R}^2}\int \Psi(\mathbf{z}) H_{i,j}(\mathbf{z}) H^*_{u,v}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \begin{cases} \theta_{u,v} & \text{if } (i,j) = (u,v) \\ 0 & \text{otherwise} \end{cases}$$

  where $\Psi(\mathbf{z})$ is the PDF of a $\mathcal{N}_2(\mathbf{0}, A)$ random vector.
- The dual BHP can be obtained as

$$H^*_{u,v}(x, y) = E[(x + iX)^u (y + iY)^v].$$

- Let $f(\mathbf{z}) \simeq f_{p,q}(\mathbf{z}) = \Psi(\mathbf{z}) \sum_{j=0}^{q} \sum_{i=0}^{p} \eta_{i,j} H_{i,j}(\mathbf{z})$.

  The coefficients $\eta_{i,j}$ are determined as follows: $\int_{\mathbb{R}^2}\int H^*_{u,v}(\mathbf{z}) f_{p,\,q}(\mathbf{z})\mathrm{d}\mathbf{z} = \int_{\mathbb{R}^2}\int H^*_{u,v}(\mathbf{z}) f(\mathbf{z})\mathrm{d}\mathbf{z}$ for $u = 0, 1, \ldots, p$ & $v = 0, 1, \ldots, q$.

Then,
$\int_{\mathbb{R}^2}\int \Psi(\mathbf{z})H^*_{u,v}(\mathbf{z})\sum_{j=0}^{q}\sum_{i=0}^{p}\eta_{i,j}H_{i,j}(\mathbf{z})\,\mathrm{d}\mathbf{z} = \sum_{n=0}^{v}\sum_{m=0}^{u}\alpha^*_{uvnm}\,\mu(m,\,n)$

where LHS $= \sum_{j=0}^{q}\sum_{i=0}^{p}\eta_{i,j}\int\int \Psi(\mathbf{z})H^*_{u,v}(\mathbf{z})H_{i,j}(\mathbf{z})\mathrm{d}\mathbf{z} = \eta_{u,\,v}\theta_{u,\,v}$,
so that

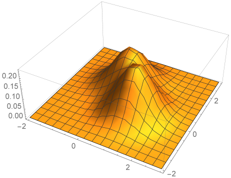
$$\eta_{u,\,v} = \frac{\sum_{n=0}^{v}\sum_{m=0}^{u}\alpha^*_{uvnm}\,\mu(m,\,n)}{\theta_{u,\,v}},\ \ u = 0,\ \ldots, p\,\&\, v = 0,\ \ldots,\ q.$$

Given that $f_{p,q}(\mathbf{z}) = \Psi(\mathbf{z})\sum_{j=0}^{q}\sum_{i=0}^{p}\eta_{i,j}H_{i,j}(\mathbf{z})$, and
$\sum_{i=0}^{p}\sum_{k=0}^{i} \equiv \sum_{k=0}^{p}\sum_{i=k}^{p}$ & $\sum_{j=0}^{q}\sum_{\ell=0}^{j} \equiv \sum_{\ell=0}^{q}\sum_{j=\ell}^{q}$, we have

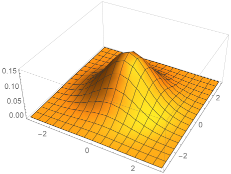
$$f_{p,\,q}(x,y) = \Psi(x,y)\sum_{\ell=0}^{q}\sum_{k=0}^{p}\Big\{\sum_{j=\ell}^{q}\sum_{i=k}^{p}\eta_{k,\,\ell}\,\alpha_{ijk\ell}\,x^k y^\ell\Big\}.$$

Example 3:

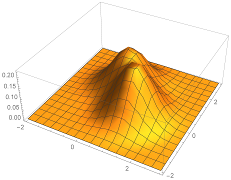
A mixture bivariate normal densities



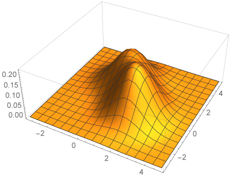
(a) Exact density



(b) base density



(c) Standard ($n = 6$)



(d) BHP ($n = 6$)

Density estimation

- The underlying density function is unknown.
- A bivariate dataset is available
- The sample moments are used instead of the exact moments

Let $\mu^{\star}(i,j)$ is the $(i,j)^{\text{th}}$ sample moment of $f(x,y)$, which is calculated as

$$\mu^{\star}(i,j) = \frac{1}{N}\sum_{k=1}^{N} x_k^i y_k^j, \quad i,\, j = 0,\, 1,\, 2, \ldots,$$

where N is the number of observations.

Stopping criterion:

Let

$$r(n) = \sum_{i=1}^{n} (\text{ECDF}(x_i,\, y_i) - F_{n,n}(x_i,\, y_i))^2, \tag{2}$$

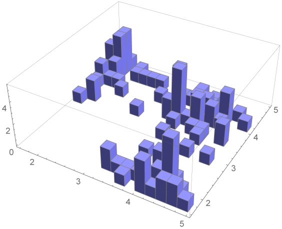
where $\text{ECDF}(x,\, y)$ is the empirical CDF corresponding to the dataset and $F_{n,n}(x,\, y) = \int_{-\infty}^{x} \int_{-\infty}^{y} f_{n,n}(x, y)\mathrm{d}y\mathrm{d}x$ is the CDF of the estimated density function $f_{n,n}(x,\, y)$.

Then we take $n^{\star}$ as the optimum degree of the polynomial adjustment which it is the value of n that minimizes $r(n)$.

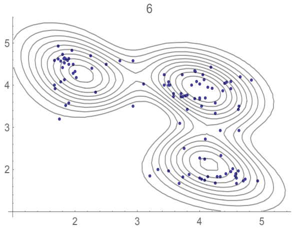
Algorithm: Moment-based density estimation by making use of standard polynomials

- Select n as the degree of the polynomial adjustment.
- Calculate the joint sample moments from the data, i.e. $\mu^\star(i,j)$ for $i,\ j = 0,\ 1,\ \ldots,\ n$.
- Assume a density estimation of the form $f_{n,n}(x,\ y) = \Psi(x,\ y)p_{n,n}(x,\ y)$, where $\Psi(x,\ y)$ is the base density and $p_{n,n}(x,\ y) = \sum_{i=0}^{n}\sum_{j=0}^{n} c_{i,j}x^i y^j$ is the polynomial adjustment of order $(n,\ n)$.
- Calculate the joint moments associated with the bivariate base density $\Psi(X,\ Y)$, i.e. $m(i,j)$ for $i,\ j = 0,\ 1,\ \ldots,\ 2n$.
- Solve the linear system resulting from $\int\int x^k y^\ell f(x,y)\mathrm{d}x\mathrm{d}y = \int\int x^k y^\ell f_{n,n}(x,y)\mathrm{d}x\mathrm{d}y$, $k,\ell = 0,\ldots,n$ to obtaine the coefficients $c_{i,j}$'s which yields the density estimate $f_{n,n}(x,\ y)$.

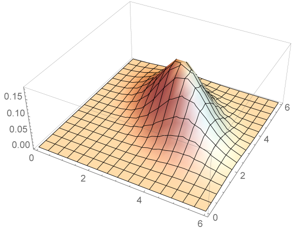
Example 4: (Old Faithful Geyser data set)



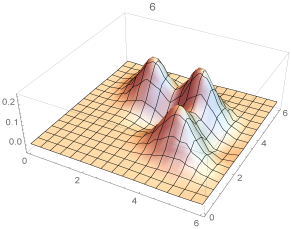
(e) Histogram of the data



(f) Contour & Scatter plots



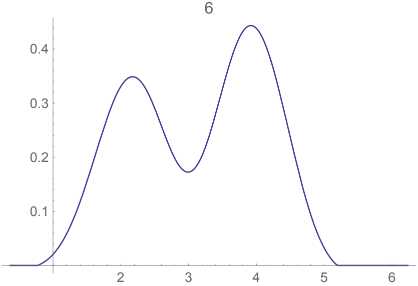
(g) Base density ($\mathcal{N}_2(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$)



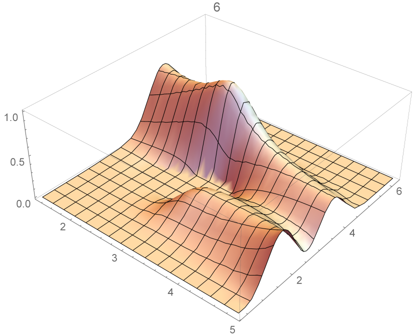
(h) Estimated ($n = 6$)

Application to regression:

Conditional density $f(y|X = x)$ obtained from $f_{6,6}(x, y)$ and $f_6(x)$.



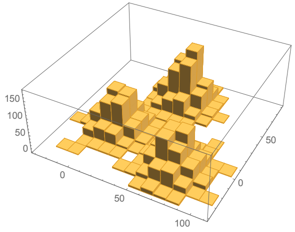
(i) Conditional density $f(y|X = 4)$



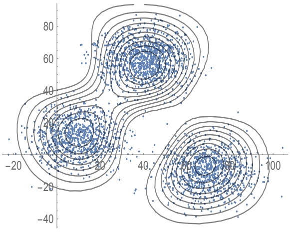
(j) Conditional density $f(y|X = x)$

Nonparametric Regression at $X = x$: $\quad E(Y|X = x)$

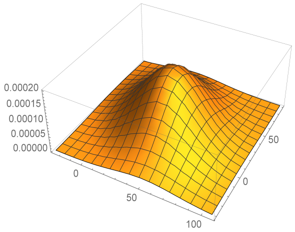
Example 5: Xclara data set



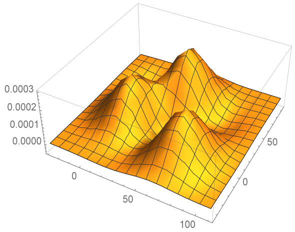
(k) Histogram of the data



(l) Contour & Scatter plots

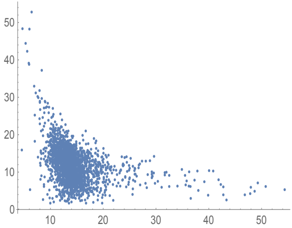


(m) Base density ($\mathcal{N}_2(\hat{\boldsymbol{\mu}}, \hat{\Sigma})$)

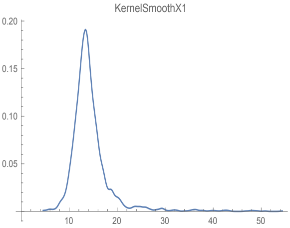


(n) Estimated density ($n = 7$)

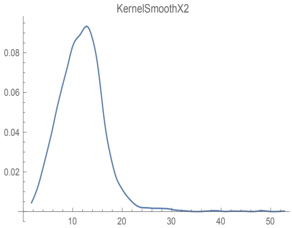
Example 6: CommViolPredUnnormalizedData



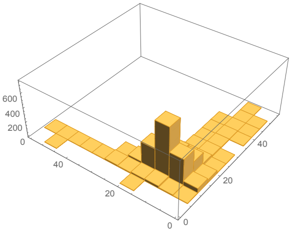
(o) Scatter plots



(p) Marginal density for X

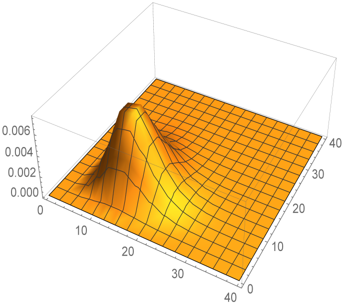


(q) Marginal density for Y

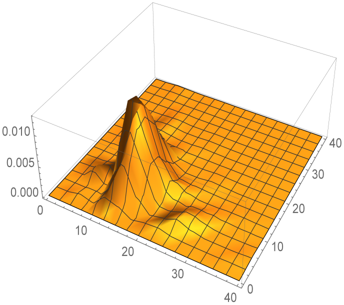


(r) Histogram of the data

Gamma distribution



(s) base density



(t) Estimated density ($n = 7$)

THANK YOU!

- S.B. Provost, (2005) “Moment-based density approximants", The Mathematica Journal, (9)727-756.
- J.C. Rayner, O. Thas, P. Pipelers, E.J. Beh, (2013) “Calculating bivariate orthonormal polynomials by recurrence", Australian & New Zealand Journal of Statistics, 1369-1473.
- R. Willink, (2005) “Normal moments and Hermite polynomials", Statistics & Probability Letters, 73(3) 271-275.
- C.S. Withers and S. Nadarajah, (2010) “Expansions for the multivariate normal", Journal of Multivariate Analysis, 101(5) 1311-1316.