HAL open science

Higher Order variance and Gauss Jacobi Quadrature

René Blacher

▶ To cite this version:

René Blacher. Higher Order variance and Gauss Jacobi Quadrature. [Research Report] LJK. 2012. hal-00708152

HAL Id: hal-00708152

https://hal.science/hal-00708152v1

Submitted on 14 Jun 2012

HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.

Higher Order variance and Gauss Jacobi Quadrature

René BLACHER

Laboratory LJK
Université Joseph Fourier
Grenoble
France

Summary : In this report, we study in a detailed way higher order variances and quadrature Gauss Jacobi. Recall that the variance of order j measures the concentration of a probability close to j points $x_{j,s}$ with weight $\lambda_{j,s}$ which are determined by the parameters of the quadrature Gauss Jacobi. We shall study many example in which these measures specify adequately the distribution of probabilities. We shall also study their estimation and their asymptotic distributions under very wide assumptions. In particular we look what happens when the probabilities are a mixture of points with measures nonzero and of continuous densities. We will see that the Gauss Jacobi Quadrature can be used in order to detect these points of nonzero measures. We apply these results to the decomposition of Gaussian mixtures. Moreover, in the case of regression we can apply these results to estimate higher order regression.

Summary : Dans ce rapport, on etudie de facon détaillée les variance d'ordre supérieur et la quadrature de Gauss Jacobi. On rappelle que la variance d'ordre j mesure la concentration d'une probabilité autour de j points $x_{j,s}$ avec des poids $\lambda_{j,s}$ qui sont déterminés par les paramêtres de la quadrature de Gauss Jacobi. On étudiera de nombreux exemples pour détailler différents cas où ces mesures précisent suffisament bien la répartition des probabilités. On étudiera aussi leur estimation et leurs lois asymptotiques sous des hypothèses très larges. On regarde en particulier ce qui se passe lorsque les probabilités sont un mélange de points de mesures non nulles et de densités continues. On verra que la Quadrature de Gauss Jacobi peut permettre de détecter ces points de mesures non nulles. On appliquera ces résultats à la décomposition de mélanges gaussiens. De plus dans le cas de régression on peut appliquer ces résultats à l'estimation de régression d'ordre supérieur.

Key Words : Higher order variance, Gauss Jacobi quadrature, Central limit theorem, Higher order regression. Gaussian mixtures.

Contents

1 Higher Order Variances **6**
1.1 Introduction . . . 6
1.1.1 Some examples . . . 8
1.1.2 Some properties of Gauss Jacobi Quadrature . . . 15
1.1.3 Other results . . . 20
1.1.4 Theoretical Examples . . . 21

2 Estimation **22**
2.1 Empirical Orthogonal functions . . . 22
2.1.1 Notations . . . 22
2.1.2 Proofs . . . 24
2.1.3 Asymptotic distribution . . . 26
2.2 Estimation of higher order variances . . . 26

3 Detection of points of concentration **34**
3.1 Introduction . . . 34
3.1.1 Complement of the results of section 1.1.2 . . . 34
3.1.2 Example 1 . . . 35
3.1.3 Example 2 : Gaussian standard case . . . 37
3.1.4 Example 3 . . . 38
3.1.5 Example 4 . . . 39
3.1.6 Example 5 . . . 39
3.1.7 Example 6 . . . 40
3.1.8 Example 7 . . . 41
3.1.9 Example 8 . . . 42
3.1.10 Conclusion . . . 42

4 Application : mixtures **44**
4.1 Some properties . . . 44
4.2 First application to mixtures . . . 47
4.2.1 Method . . . 47
4.2.2 Examples . . . 48
4.3 Second application to mixtures . . . 50
4.3.1 Presentation . . . 50
4.3.2 Example . . . 50
4.3.3 Calculation of the first standard deviation . . . 50
4.3.4 Suppression of the first Gaussian component . . . 55
4.3.5 Estimation of the second Gaussian component . . . 57

5 Higher Order Regression **62**
5.1 Notations and theorems . . . 62
5.1.1 Notations . . . 62
5.1.2 Properties . . . 63
5.1.3 Method of computation . . . 63
5.2 Examples : regression of order 2 . . . 64

A Variance of order 3 **68**
A.1 Elementary calculations . . . 68
A.1.1 Some formulas . . . 68
A.1.2 Polynomials . . . 68
A.1.3 Weights . . . 69
A.1.4 Variance of order 3 . . . 69

Chapter 1

Higher Order Variances

1.1 Introduction

Orthogonal polynomials have many interesting applications in Probability and Statistics. So they have introduced higher order correlation coefficients and higher order variances (cf [1], [2], [4], [5], [7], [6], [3]). They also have introduced new hypotheses for the central limit theorem (cf [3]). One can also obtain the distributions of quadratic forms, Gaussian or not Gaussian, and simple methods of calculation of these laws (cf [8]).

Higher order variances have been introduced in [6] and [7]. They generalize the classical variance. Thus, variance of order 1 measures of concentration of a probability close to a point : the expectation. Variance of order j measures the concentration close to j points which are the roots of the j-th orthogonal polynomial.

Notations 1.1.1 *let X be a random variable defined on (Ω, A, P). Let m be the distribution of X. Let $\tilde{P}_j$ be the j-th orthogonal polynomial associated to X such that $\tilde{P}_j(x) = \sum_{t=0}^{j} a_{j,t} x^t$ with $a_{j,j} = 1$.*

We set $n_0^m = dim\{L^2(R, m)\}$. Let $\Theta \subset \mathbb{N}$ such that $\tilde{P}_j(x)$ exists. We denote by P_j the j-th orthonormal polynomial associated to X if there exists.

Remark that if m is concentrated close to n_0^m points where $n_0^m < \infty$, $\Theta = \{0, 1, ..., n_0^m\}$.

If not, $\Theta = \mathbb{N}$ if all moments exists, and $\Theta = \{0, 1, ..., d\}$ if $\int |x|^{2d-1}.m(dx) < \infty$ and $\int |x|^{2d+1}.m(dx) = \infty$. In this case, P_j exist if $\int |x|^{2d}.m(dx) < \infty$.

For example, $\tilde{P}_0 \equiv 1$, $\tilde{P}_1(x) = x - \mathbb{E}(X)$ where $\mathbb{E}(.)$ is the expectation,

$$\tilde{P}_2(x) = x^2 - \frac{M_3 - M_1 M_2}{M_2 - M_1^2}(x - M_1) - M_2 \ ,$$

where $M_s = \mathbb{E}(X^s)$.

Now we know that the zeros of $\tilde{P}_j$ are real (cf th 5-2 page 27 [10])

Proposition 1.1.1 *Let $j \in \Theta$. Then, the zeros of $\tilde{P}_j$ are distincts and real. We denote them by $x_{j,s}$, s=1,2,....,j.*

For example, if j=1, $x_{1,1} = \mathbb{E}(X)$. If j=2,

$$x_{2,s} = \frac{M_3 - M_1 M_2}{2(M_2 - M_1^2)} \ \pm \ \frac{1}{2}\sqrt{\left(\frac{M_3 - M_1 M_2}{M_2 - M_1^2}\right)^2 - 4M_2} \ .$$

We recall theorem 5.3 of [10].

Proposition 1.1.2 *Suppose that, for all $j \in \Theta$, $x_{j,s} < x_{j,s+1}$ for each s=1,2,...,j-1. Then, for all $j+1 \in \Theta$, $x_{j+1,s} < x_{j,s} < x_{j+1,s+1}$ for each s=1,2,...,j.*

Now the roots of orthogonal polynomials have stronger properties : the Gauss-Jacobi Quadrature.

Theorem 1 *Let $j \in \Theta$. There exists a single probability m^j concentrated over j distincts points such that $\int x^q.m(dx) = \int x^q.m^j(dx)$ for q=0,1,...,2j-1.*

Moreover, the j points of concentration of m^j are the j zeros of $\tilde{P}_j$: $x_{j,s}$, s=1,...,j, and the probabilities $\lambda_{j,s} = m^j\big(\{x_{j,s}\}\big)$ check $\lambda_{j,s} = \int \ell_s^j(x).m(dx)$, where $\ell_s^j(x) = \frac{\tilde{P}_j(x)}{(x-x_{j,s})\tilde{P}'_j(x_{j,s})}$ when $\tilde{P}'_j$ is the derivative of $\tilde{P}_j$.

Proof The most simple way in order to prove this theorem is to use proof of [7]. It shows that the $\lambda_{j,t}$'s are the only solution of the system of Cramer $\sum_{t=1}^{j} \lambda_t P_q(x_{j,t}) = \delta_{q,0}$ for q=0,1,...,j-1. The proof is more complicated than the classic proofs. But it has the advantage of treating also the case $j = n_0^m$.

If we do not suppose $j = n_0^m$, one can use classical proofs : they are in paragraph 6 page 31 of [10] or in theorem 3-2 and formula 3-8, page 19-23 of [11]. Then, if $j = n_0^m$, one can use the proof of theorem 2. ■

For example $\ell_1^j(x) = \frac{(x-x_{j,2})(x-x_{j,3})...(x-x_{j,j})}{(x_{j,1}-x_{j,2})(x_{j,1}-x_{j,3})...(x_{j,1}-x_{j,j})}$. In particular, if j=2, $\ell_1^2(x) = \frac{x-x_{2,2}}{x_{2,1}-x_{2,2}}$ and $\ell_2^2(x) = \frac{x-x_{2,1}}{x_{2,2}-x_{2,1}}$. Therefore, $\lambda_{2,1} = \frac{M_1-x_{2,2}}{x_{2,1}-x_{2,2}}$ and $\lambda_{2,2} = \frac{M_1-x_{2,1}}{x_{2,2}-x_{2,1}}$.

Recall that the $\lambda_{j,k}$'s are called Christoffel numbers

Now, we complete the definition of Gauss Jacobi quadrature by defining higher order variances.

Definition 1.1.2 *Let $j \in \Theta$. We call variance of order j, and we note it by σ_j^2 or $\sigma_j(X)^2$ or $\sigma_j(m)^2$ the real $\sigma_j^2 = \int |\tilde{P}_j|^2.dm$.*

Remark that $\tilde{P}_j = \sigma_j P_j$. Moreover, $\sigma_1(X)^2 = M_2 - M_1^2$ is the classical variance. Now, if j=2, $\sigma_2^2 = M_4 - \frac{(M_3-M_1M_2)^2}{M_2-M_1^2} - M_2^2$.

Then, variance of order j measure the concentration close to j distinct points.

Theorem 2 *Let $j \in \Theta$. Then, $\sigma_j = 0$ if and only if m is concentrated in j distincts points which are the zeros of $\tilde{P}_j$: the $x_{j,t}$'s. Moreover the probability associated at each $x_{j,t}$ is equal at $\lambda_{j,t}$. In this case, $j = n_0^m < \infty$ and $\tilde{P}_j = 0$ in $L^2(\mathbb{R}, m)$.*

Proof we use the two following lemmas : they are proved in 4-2 and 4-3 of [4].

Lemma 1.1.1 *Let $p \in \mathbb{N}^*$. Let m' be a probability on $\mathbb{R}$. Then the two following assertions are equivalent.*

1) $dim\big(L^2(\mathbb{R}, m') = p$.

2) There exists $\Xi = \{x_1, x_2, ..., x_p\} \subset \mathbb{R}$, $Card(\Xi) = p$ such that $m'(x_s) = \lambda_s > 0$ for all $s \in \{1, 2, ..., p\}$ and $\sum_{s=1}^{p} \lambda_s = 1$, i.e. $m' = \sum_{s=1}^{p} \lambda_s \delta_{x_s}$.

Lemma 1.1.2 *Let $t \in \mathbb{N}^*$, such that $t < n_0^m$. Then, the set $\{x^j\}$, j=0,1,...,t, $x^j \in \mathbb{R}[X]$ is a set linearly independent of $L^2(\mathbb{R}, m)$.*

Proof of theorem 2 If $\sigma_j = 0$, $\tilde{P}_j = 0$ in $L^2(\mathbb{R}, m)$. Then, m is concentrated on the j roots $x_{j,s}$ of $\tilde{P}_j = 0$. Now it is not concentrated on j-h point, if not $dim\big\{L^2(\mathbb{R}, m)\big\} = j - h$ and $1, x, x^2, ..., x^{j-h-1}$ would be linearly dependent. Therefore $\sigma_{j-h} = 0$. But it is not the case : if

not σ_j would not be defined.

Now we know that $\ell_k^j(x_{j,t}) = \delta_{k,t}$. Therefore, $\lambda_{j,k} = \int \ell_k^j(x).m(dx) = m(\{x_{j,k}\})$. ■

The Bienayme-Tschbichev Inequality allows to specify more this concentration.

Proposition 1.1.3 *Let $\epsilon > 0$. Then, $P\big(|\tilde{P}_j| > \epsilon\big) \ \leq \ \frac{\sigma_j^2}{\epsilon^2}$.*

In particular assume that σ_j^2 is small enough. Let ω such that $|\tilde{P}_j(X(\omega))| \leq \epsilon$. Then, there exists s such that $X(\omega) - x_{j,s}$ is small enough. Then, the variance of order j measures the concentration of a probability close to j distincts points.

Then, they generalize the classical variance which one can call variance of order 1. Indeed, classical variance measure the concentration close to expectation. For the variance of order j, the roots of $\tilde{P}_j$ plays this role. Moreover we know the weight associated : the $\lambda_{j,t}$'s. All these properties justify well the name of variances of higher order.

1.1.1 Some examples

We'll look at some example. We will see that the results tally what it was expected intuitively about higher order variances parameters and Gauss Jacobi quadrature.

Remark 1.1.3 *In the figures of this section, the graphs are not normalized. Indeed, we put on the same figure the densities and weights of Gauss Jacobi, which is normally impossible. Indeed if we show only the densities, the densities of the measure concentrated on the $x_{j,t}$'s should be infinite. This means that the y-axis is only there to give information on the order of size: it should not be taken into account for exact calculations.*

The x-axis is correct.

In spite of this remark, the following figures are clear enough to get an idea of density and weight $\lambda_{j,t}$'s of various probabilities.

Remark 1.1.4 *The higher order variances transformed by homothety can give very different figures since it depend on the moments which can become very large or very small. We can not properly use the higher order variances in order to know the concentration unless it is first carried out a normalization.*

For example, a normalization can may be given by considering the number $\frac{\sigma_j}{||x^j||}$ which represents the sinus of the angle formed by the polynomial x^j and the subspace spanned by polynomials of degree strictly less than j.

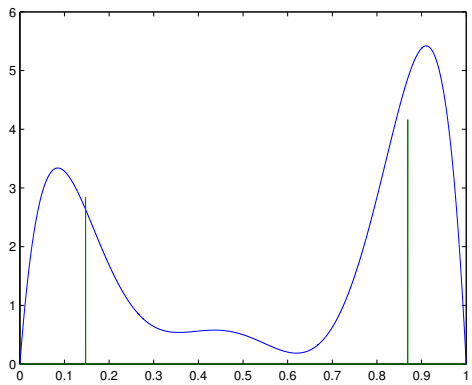

Figure 1.1: $x_{2,t}$ =0.8691, 0.1473, $\lambda_{2,t}$= 0.5944, 0.4056, σ_2^2=0.0037

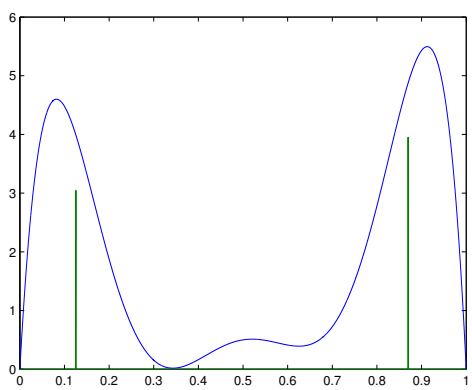

Figure 1.2: $x_{2,t}$=0.8698, 0.1257, $\lambda_{2,t}$=0.5647, 0.4353 , σ_2^2=0.0034

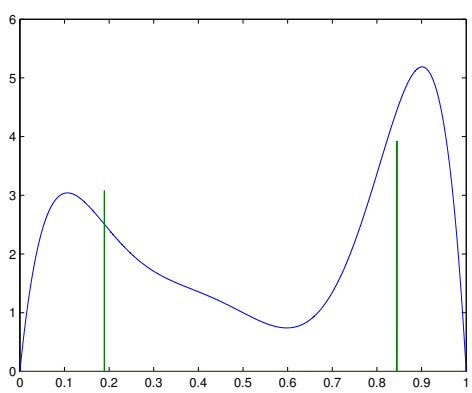

Figure 1.3: $x_{2,t}$=0.8447, 0.1893, $\lambda_{2,t}$=0.5606, 0.4394, σ_2^2=0.0044

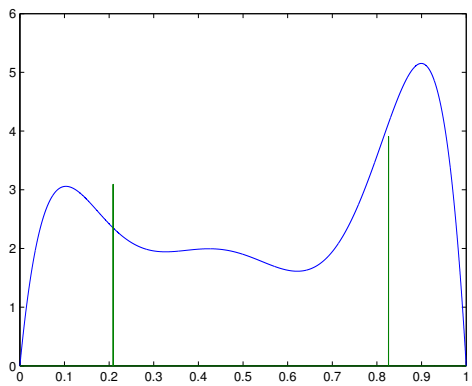

Figure 1.4: $x_{2,t}$=0.8261, 0.2090, $\lambda_{2,t}$=0.5582, 0.4418, σ_2^2= 0.0048

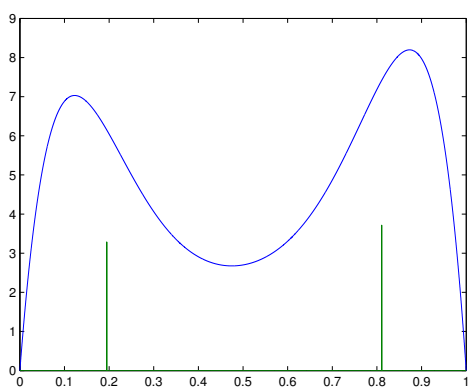

Figure 1.5: $x_{2,t}$=0.8109, 0.1948, $\lambda_{2,t}$=0.5309, 0.4691, σ_2^2=0.0044

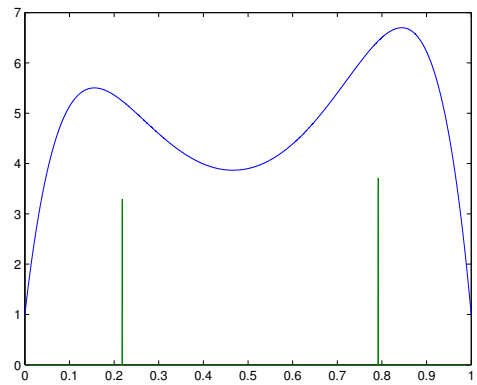

Figure 1.6: $x_{2,t}$=0.7917, 0.2183, $\lambda_{2,t}$=0.5298, 0.4702, σ_2^2=0.0045

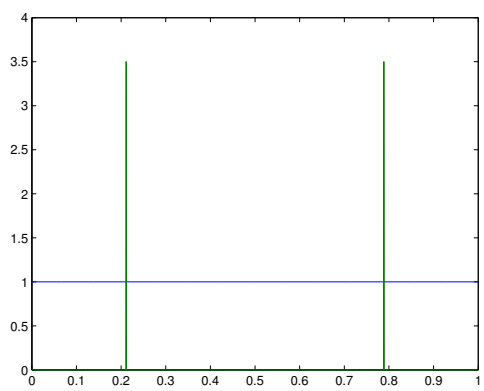

Figure 1.7: Uniform distribution, $x_{2,t}$=0.7887, 0.2114, $\lambda_{2,t}$=0.5000, 0.5000, σ_2^2=0.0056

Note again that although the variance of order j is small, σ_j^2 can measure not a good concentration close to j distinct points. For example, the classical variance of a Gaussian distribution may be small. So we have a concentration around 0. This leads that some following variances will be small. But we cannot speak about a concentration around several points.

In fact, there seems that this is the first variance σ_j^2 small when we take the sequence σ_i^2, i = 1,2, ... which may indicate a concentration around j points.

Gaussian mixtures

Now, we have examples of Gaussian mixtures.

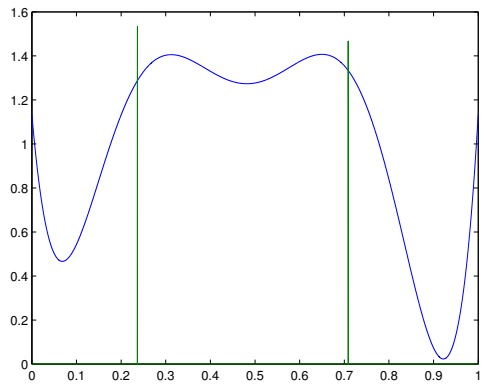

Figure 1.8: $x_{2,t}$=0.7089, 0.2365 , $\lambda_{2,t}$=0.4889, 0.5111, p_s^2, σ_2^2=0.0038

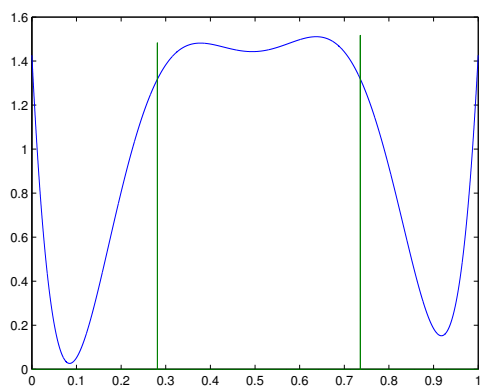

Figure 1.9: $x_{2,t}$=0.7360, 0.2813, $\lambda_{2,t}$=0.3130, 0.2126, σ_2^2=0.1554

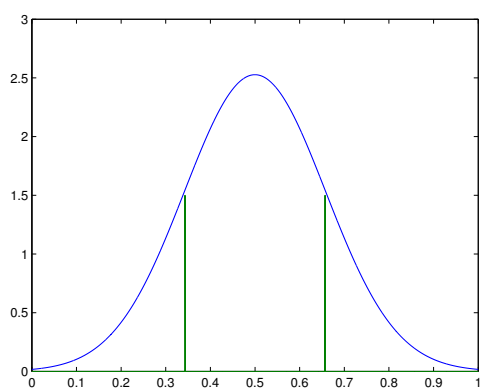

Figure 1.10: Distribution N(0,0.1), $x_{2,t}$=0.6568, 0.3433, $\lambda_{2,t}$=0.5000, 0.5000, σ_2^2=0.0011

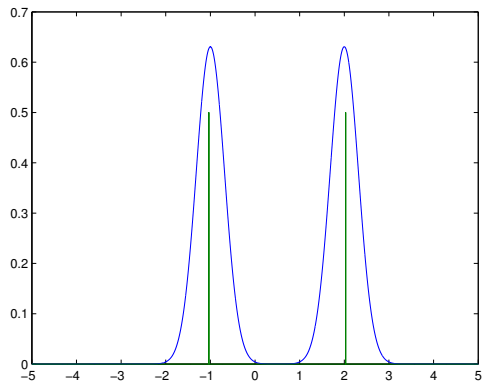

Figure 1.11: $x_{2,t}$=2.0330, -1.0330, $\lambda_{2,t}$=0.5000, 0.5000, σ_2^2=0.9200

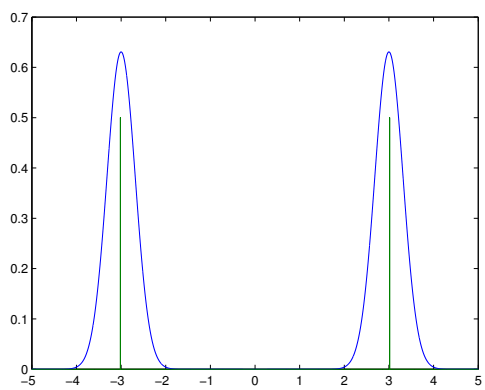

Figure 1.12: $x_{2,t}$=2.0330, -1.0330, $\lambda_{2,t}$=0.5000, 0.5000, σ_2^2=3.6200

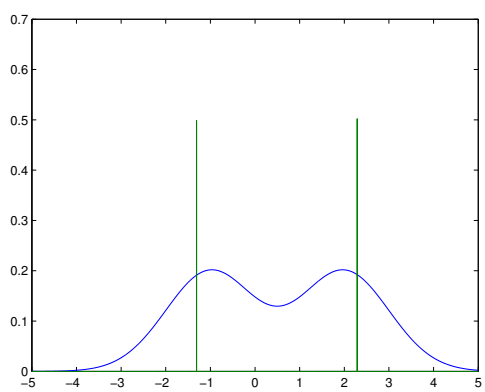

Figure 1.13: $x_{2,t}$=2.0330, -1.0330, $\lambda_{2,t}$= 0.5000, 0.5000, σ_2^2=3.6200

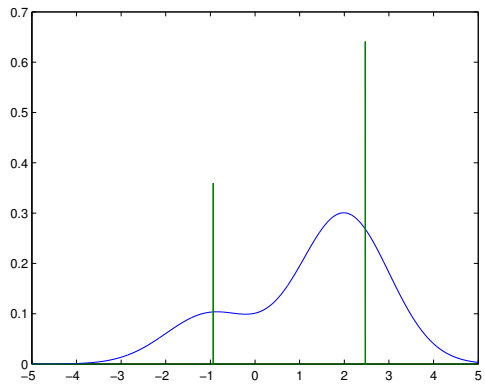

Figure 1.14: $x_{2,t}$=2.4700, -0.9381, $\lambda_{2,t}$=0.6409, 0.3591, σ_2^2=9.8473

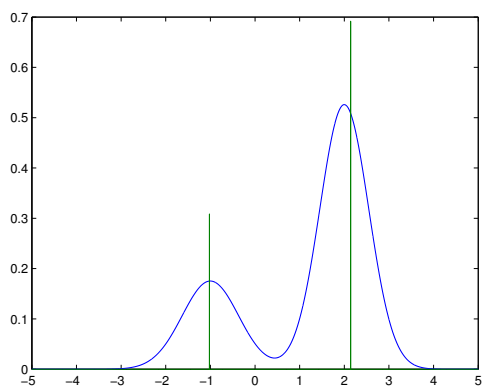

Figure 1.15: $x_{2,t}$=2.1416, -1.0216, $\lambda_{2,t}$=0.6916, 0.3084, σ_2^2=3.1403

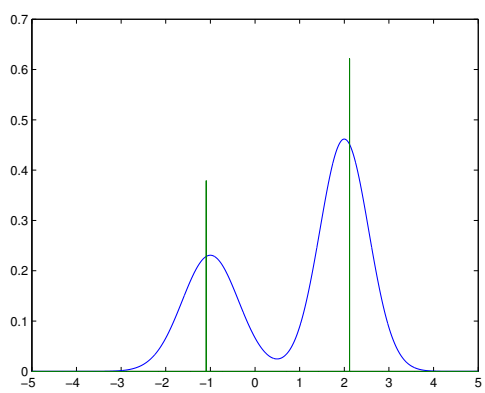

Figure 1.16: $x_{2,t}$=2.1179, -1.0924, $\lambda_{2,t}$=0.6212, 0.3788, σ_2^2=3.2614

Now we shall study the variances of order j for mixtures of j Gaussian components.

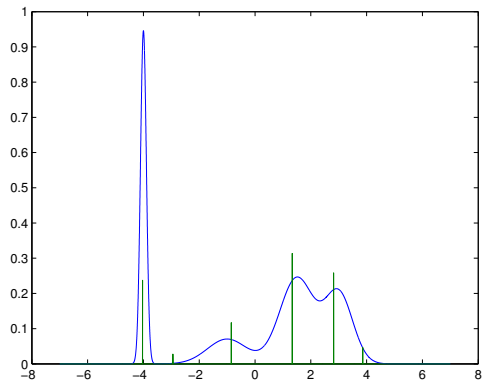

Figure 1.17: σ_6^2=1658.9

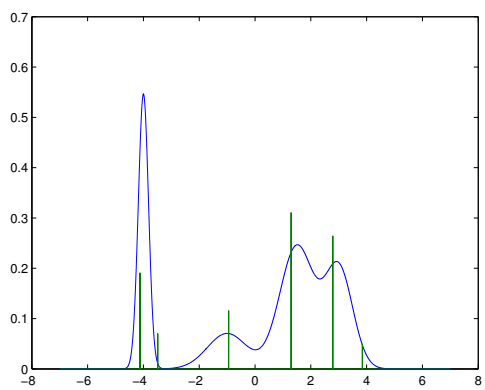

Figure 1.18: σ_6^2=2704.8

1.1.2 Some properties of Gauss Jacobi Quadrature

Concentration points of a probability can be detected using various properties of the Gauss Jacobi Quadrature. First, the most important of these properties is the Stieltjes-Markov Inequality.

Proposition 1.1.4 *Let F_X be the distribution function of X. Then, for all k ∈1,2,..,j,*

$$\sum_{x_{j,s}<x_{j,k}} \lambda_{j,s} \leq F_X(x_{j,k}-0) \quad \textit{and} \quad \sum_{x_{j,s}\leq x_{j,k}} \lambda_{j,s} \geq F_X(x_{j,k}+0) \ .$$

These results are proved pages 26-29 of [11] equation 5.4. For example, in figure 1.27, we have the distribution function of m and m^j.

This result means that if F_X has a point of discontinuity $x_{j,k} < x_0 < x_{j,k+1}$: $F_X(x_0+0) - F_X(x_0-0) = b > 0$, i.e. $m(x_0) = b$. As this discontinuity is between two roots, we thus find $\lambda_{j,k} + \lambda_{j,k+1} \geq b$ for all j.

Now we will give a condition under which we have the convergence of distributions : $m^j \xrightarrow{d} m$ (th 1.1 page 89 of [11]).

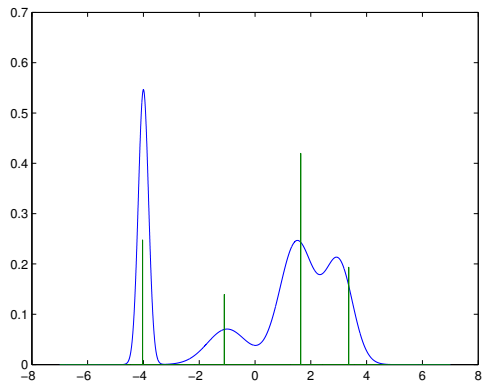

Figure 1.19: σ_5^2=2704.8

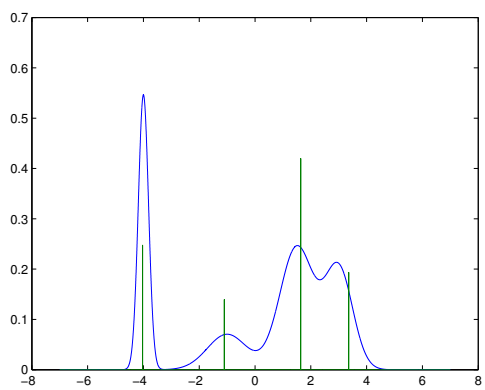

Figure 1.20: σ_4^2=92.0874

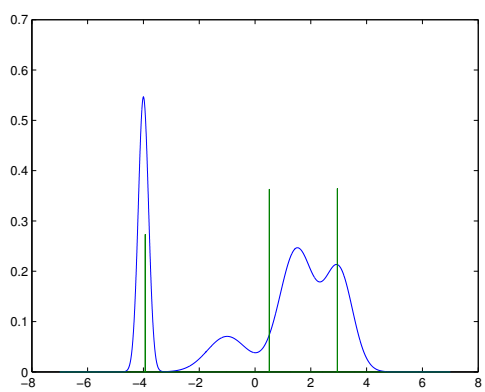

Figure 1.21: σ_3^2=26.9485

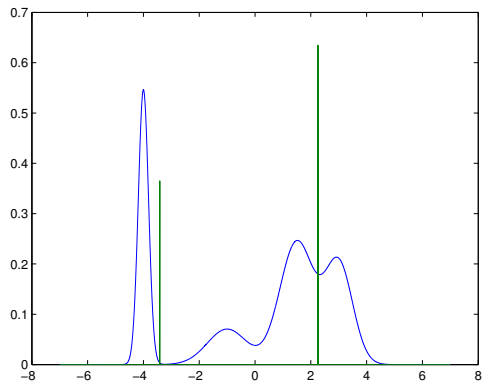

Figure 1.22: σ_2^2=7.4576

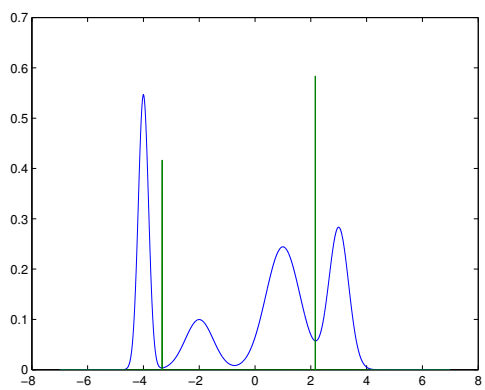

Figure 1.23: σ_2^2=7.3208

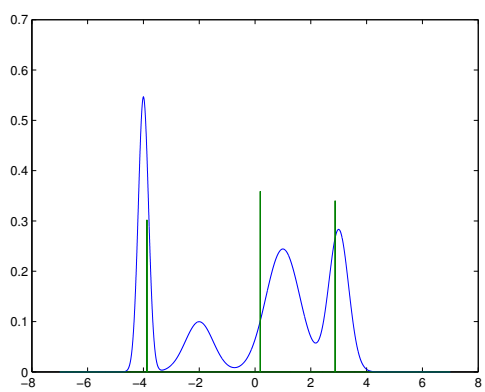

Figure 1.24: σ_3^2=27.1092

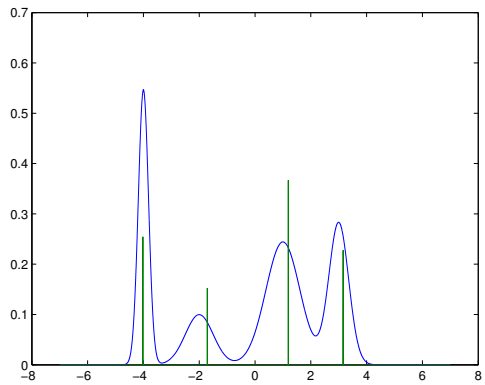

Figure 1.25: σ_4^2=93.1528

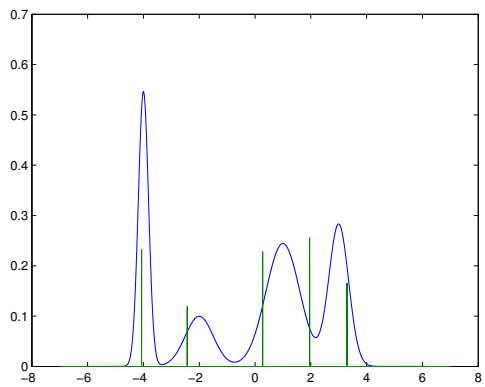

Figure 1.26: σ_5^2=306.6277

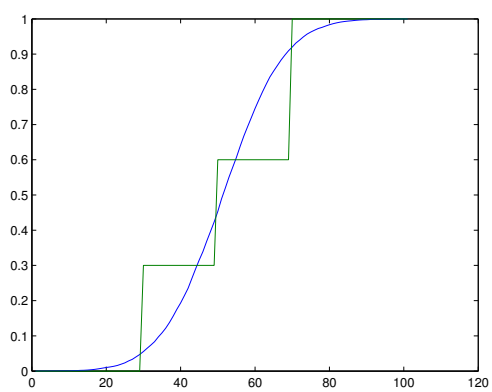

Figure 1.27: Stieljes-Markov Inequality

Theorem 3 *We suppose that there is not other random variable T , $T \neq X$ m-almost surely, such that $E\{T^n\} = E\{X^n\}$ for n=0,1,2,...*

Let $f \in L^1(\mathbb{R}, m)$. Assume that there exists $A \geq 0$, $B \geq 0$ and $s \in \mathbb{N}$ such that $|f(x)| \leq A + Bx^{2s}$. Then,

$$Lim_{j\to\infty} \int f(x)m^j(dx) = \int f(x)m(dx) \ .$$

One can specify the speed of convergence in the following way (Theorem 4.4 page 110 of [11]).

Theorem 4 *Assume that $X \in [-1, 1]$ has a absolutely continuous distribution function F_X such that $F'_X(x) \leq \frac{k_0}{\sqrt{1-x^2}}$ for all $x \in [-1, 1]$. Then, for all $-1 < x_0 < 1$,*

$$\int_{-1}^{x_0} m^j(dx) = \int_{-1}^{x_0} m(dx) + O(\frac{1}{j}) \ .$$

Now if the probability is enough regular, the weight $\lambda_{j,k}$'s converges regularly to 0 (cf Lemma 3.1 page 100 and remark page 101 of [11]).

Theorem 5 *Assume that $X \in [-1, 1]$ and that $\left|\frac{F_X(x)-F_X(y)}{x-y}\right| \leq M < \infty$. Then, $\lambda_{j,k} = O(\frac{M}{j})$.*

We can specify this result in the following way (Theorem 6.8 page 254 of [11]).

Theorem 6 *Assume that $X \in [-1, 1]$. Assume that there exists a polynomial $\tau(x)$ such that $F'_X(x) \geq \tau(x)^2$ for all $x \in [-1, 1]$. We suppose $F_X(x)$ is absolutely continuous in $[-1, +1]$ where $\tau(x)$ does not vanish. Assume that, for all $x,\ y \in [-1, 1]$,*

$$|F'_X(x) - F'_X(y)| \leq K|x - y|^\rho$$

is satisfied for a $0 < \rho \leq 1$ and for all x, $y \in [-1, 1]$. Then,

$$\frac{1}{\lambda_{j,k}} = \frac{j}{\pi} \ \frac{1}{\sqrt{1 - x_{j,k}}\ F'_X(x_{j,k})} + O(j^{1-\rho}) \quad when\ \rho < 1,$$

$$\frac{1}{\lambda_{j,k}} = \frac{j}{\pi} \ \frac{1}{\sqrt{1 - x_{j,k}}\ F'_X(x_{j,k})} + O(log(j)) \quad when\ \rho = 1.$$

Now if the distribution of X is enough regular, distances of successive roots $x_{j,k}$ converges to 0 (Theorem 5.1 page 111 of [11]).

Theorem 7 *Assume that $X \in [-1, 1]$. Assume that $0 < M' < \left|\frac{F_X(x)-F_X(y)}{x-y}\right| \leq M < \infty$ holds for x, $y \in [c, d]$. Let $x_{j,k} < x_{j,k+1}$ be two successive zeros of $P_j(x)$ such that $x_{j,k},\ x_{j,k+1} \in [c+\epsilon, d-\epsilon]$ where $\epsilon > 0$.*

Then, there exists two positive numbers $c_1(\epsilon) > 0$ and $c_2(\epsilon) > 0$ depending only on m, c, d, and ϵ such that

$$\frac{c_1(\epsilon)}{j} \leq x_{j,k+1} - x_{j,k} \leq \frac{c_2(\epsilon)}{j} \ .$$

This means that the distance of the roots is of the order of 1/j if the Lipschitz condition is checked by F_X. We can specify this result in the following way (Theorem 9.2 page 130 of [11]).

Theorem 8 *Assume that $X \in [-1, 1]$. Assume that $F'_X(x) > 0$ for all $x \in [-1, 1]$. Let us denote by $N(\Theta_1, \Theta_2)$ the number of $x_{j,k} \in [cos(\Theta_1), cos(\Theta_2)]$. Then,*

$$lim_{j\to\infty} \frac{N(\Theta_1, \Theta_2)}{j} = \frac{\Theta_2 - \Theta_1}{\pi} \ .$$

These theorems in particular means that if there is no point x_0 such that $m(\{x_0\}) > 0$, the distribution of roots and weights is enough regular. As this is not the case if $m(\{x_0\}) > 0$, it will detect the existence of those discontinuities by a way enough simple.

1.1.3 Other results

At first, we have the following property.

Proposition 1.1.5 *Let $j \in \Theta$. Then $\sigma\big(\tilde{P}_j(X)\big) = \sigma_j$. Moreover, if $j < n_0^m$, $\sigma\big(P_j(X)\big) = 1$.*

Now, the variance of order j is is invariant by translation.

Proposition 1.1.6 *Let $a \in \mathbb{R}$. Let m_a the translated probability : $m_a(B) = P(X + a \in B)$. For each $j \in \theta$, the (j+1)-th orthonormal polynomial associated at m_a is $\tilde{P}_j(x - a)$. Moreover, let $x'_{j,1}, x'_{j,2}, ..., x'_{j,j}$, the zeros of $\tilde{P}_j(x - a)$, $\lambda'_{j,1}, \lambda'_{j,2}, ..., \lambda'_{j,j}$, be the weights of associated Gauss-Jacobi Quadrature, and ${\sigma'}_j^2$ be the variance of order j associated at m_a. Then, $x'_{j,s} = x_{j,s} + a$, $\lambda'_{j,s} = \lambda_{j,s}$ and ${\sigma'}_j^2 = \sigma_j^2$.*

In order to prove this result, it is enough to remark that $\int \tilde{P}_j(x - a)\tilde{P}_k(x - a).m_a(dx) = \int \tilde{P}_j(x)\tilde{P}_k(x).m(dx)$

Now recall how to calculate practically the variance of order j.

Proposition 1.1.7 *Let $j \in \Theta$. Then,*

$$\sigma_j^2 = M_{2j} - \sum_{s=0}^{j-1} \beta_{j,s}^2 \text{ where } \beta_{j,s} = \int x^j P_s(x).mdx\,.$$

Proof We have

$$\tilde{P}_j = x^j - \sum_{s=0}^{j-1} \mathbb{E}\{X^j P_s(X)\} P_s(x)\ .$$

Therefore,

$$\sigma_j^2 = \int \Big(x^j - \sum_{s=0}^{j-1} \mathbb{E}\{X^j P_s(X)\} P_s(x)\Big)^2 m(dx)$$

$$= \int x^{2j} m(dx) - 2\int x^j \Big(\sum_{s=0}^{j-1} \mathbb{E}\{X^j P_s(X)\} P_s(x)\Big) m(dx) + \int \Big(\sum_{s=0}^{j-1} \mathbb{E}\{X^j P_s(X)\} P_s(x)\Big)^2 m(dx)$$

$$= \int x^{2j} m(dx) - 2\sum_{s=0}^{j-1} \mathbb{E}\{X^j P_s(X)\}^2 + \sum_{s=0}^{j-1} \mathbb{E}\{X^j P_s(X)\}^2\ . \blacksquare$$

The following proposition results from the Gram-Schmidt Process

Proposition 1.1.8 *The real σ_j is the distance in $L^2(\mathbb{R}, m)$ of the polynomial $x \longmapsto x^j$ to the subspace of $L^2(\mathbb{R}, m)$ spanned by the polynomials of degree more little than j-1. Moreover, the minimum of $\int \big[(x-t_1)(x-t_2)...(x-t_j)\big]^2.m(dx)$ when $(t_1, t_2, ..., t_j) \in \mathbb{R}^j$ is reached for $(t_1, t_2, ..., t_j) = (x_{j,1}, x_{j,2}, ..., x_{j,j})$ and is equal to σ_j^2 .*

Now note that there cannot be more than two roots in an interval of measure zero.

Proposition 1.1.9 *It can not be three successive roots $x_{j,s} < x_{j,s+1} < x_{j,s+2}$ such that $P\{X \in [x_{j,s}, x_{j,s+2}]\} = 0$ if $\lambda_{j,s+1} > 0$.*

Proof By Stieljes Markov inequality, we know that $\sum_{x_{j,s} < x_{j,k+2}} \lambda_{j,s} \leq F_X(x_{j,k+2} - 0)$ and $\sum_{x_{j,s} \leq x_{j,k}} \lambda_{j,s} \geq F_X(x_{j,k} + 0)$. Then,

$$0 = F_X(x_{j,k+2}) - F_X(x_{j,k}) = F_X(x_{j,k+2} - 0) - F_X(x_{j,k} + 0)$$

$$\geq \sum_{x_{j,s} < x_{j,k+2}} \lambda_{j,s} - \sum_{x_{j,s} \leq x_{j,k}} \lambda_{j,s} = \lambda_{j,k+1} > 0\ . \blacksquare$$

1.1.4 Theoretical Examples

At first, we recall the results on Jacobi polynomials associated to the Beta distribution (cf page 143 [10]).

Proposition 1.1.10 *We suppose that X has the density*

$$\frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)}\ x^{a-1}(1-x)^{b-1}$$

if $0 \leq x \leq 1$. *We denote by* $\tilde{J}_j^{ab}$ *and* $\left(\sigma_j^{ab}\right)^2$ *the orthogonal polynomials and associated variances. Then,*

$$\tilde{J}_j^{ab}(x) = (-1)^j \frac{\Gamma(a+b+j-1)}{\Gamma(a+b+2j-1)} x^{1-a}(1-x)^{1-b} \frac{d^j\left(x^{a-1+j}(1-x)^{b-1+j}\right)}{dx^j}$$

$$\left(\sigma_j^{ab}\right)^2 = \frac{\Gamma(a+j)\Gamma(b+j)\Gamma(a+b+j-1)(j!)}{\beta(a,b)\Gamma(a+b+2j-1)^2(a+b+2j-1)}\ .$$

Now, we study Legendre polynomials. (cf page 143 [10]).

Proposition 1.1.11 *We suppose that X has the uniform distribution on [0,1]. We denote by* $\tilde{Le}_j$ *and* σ_j^2 *the orthogonal polynomials and associated variances. Then,*

$$\tilde{Le}_j(x) = \frac{j!}{2j!} \sum_{t=0}^{j} \frac{C_j^t(-1)^t((j+t)!)}{t!}\ x^t$$

$$\left(\sigma_j\right)^2 = \frac{(j!)^4)}{[(2j)!]^2(2j+1)}\ .$$

With the normal distribution we use the Hermite polynomials (cf page 145 [10]).

Proposition 1.1.12 *Let* $\hat{H}_j(x) = e^{x^2}\frac{d^j(e^{-x^2})}{dx}$ *the Hermite orthogonal. We suppose that X has the* $N(m,\sigma^2)$ *distribution. We denote by* $H_j^{m\sigma^2}$ *and* $\left(\sigma_j^{m\sigma^2}\right)^2$ *the orthogonal polynomials and associated variances. Then,*

$$\tilde{H}_j^{m\sigma^2}(x) = \frac{(-1)^j\sigma^j}{2^{j/2}}\hat{H}_j\Big(\frac{x-m}{\sigma\sqrt{2}}\Big)\ ,$$

$$\left(\sigma_j^{m\sigma^2}\right)^2 = j!\sigma^{2j}\ .$$

At last we have the Laguerre polynomials (cf page 144 [10]).

Proposition 1.1.13 *We suppose that X has* $\gamma(a,p)$ *distribution* $(a>0)$*, i.e. X has the density*

$$\frac{p^a}{\Gamma(a)}\ e^{-px}x^{a-1}\ if\ x \geq 0.$$

We denote by $\tilde{L}_j^{ap}$ *and* $\left(\sigma_j^{ap}\right)^2$ *the orthogonal polynomials and associated variances. Then,*

$$\tilde{L}_j^{ap}(x) = \frac{(-1)^j}{p^j}x^{1-a}e^{px}\frac{d^j\left(x^{a-1+j}e^{-px}\right)}{dx^j}$$

$$\left(\sigma_j^{ap}\right)^2 = \frac{j!\Gamma(a+j)}{\Gamma(a)p^{2j}}\ .$$

Chapter 2

Estimation

We will see that one can easily estimate the higher order variances and the Gauss Jacobi quadrature. We can also obtain their asymptotic distributions. We will study this problem under the weakest possible assumptions. For this reason, we first recall some properties of empirical orthogonal functions.

2.1 Empirical Orthogonal functions

In order to define empirical orthogonal functions in the general case, at first we need to define orthogonal functions. We do this under the most general assumptions possible.

2.1.1 Notations

Notations 2.1.1 *Let (Ω, A, P) be a probability space. Let $h \in \mathbb{N}^*$ and $\Lambda = (\Lambda_0, \Lambda_1, ..., \Lambda_h) \in \mathbb{R}^{h+1}$ be a random vector defined on (Ω, A, P) . We assume that $\mathbb{E}(\Lambda_j^2) < +\infty$ for all $j \in 0, 1, ..., h$. We assume that $\Lambda_0, \Lambda_1, ..., \Lambda_h$ are lineraly independent in $L^2(\Omega, A, P)$.*

Under the previous assumptions, the Λ_j's can be orthogonalized by using the process of Gram-Schmidt.

Theorem 9 *Let μ be the distribution of Λ. Let $<,>$ and $|.|$ be the scalar product and the norm of $L^2(\mathbb{R}^{h+1}, \mu)$. Let $\chi_0, \chi_1, ..., \chi_h$ be h+1 real variables. We set $\chi = (\chi_0, \chi_1, ..., \chi_h)$ and we identify χ_j and the function $\chi \longmapsto \chi_j$. For all $\chi \in \mathbb{R}^{h+1}$, we set $\tilde{A}_{-1}(\chi) = A_{-1}(\chi) = 0$,*

$$and, for \quad h \geq j \geq 0, \ \tilde{A}_j(\chi) = \chi_j - \sum_{s=-1}^{j-1} < \chi_j, A_s > A_s(\chi),$$

$$A_j(\chi) = \frac{\tilde{A}_j(\chi)}{||\tilde{A}_j||}.$$

Then, for all $(j, j') \in \{0, 1, ..., h\}^2$, $\int A_j A_{j'} d\mu = \delta_{j,j'}$ where $\delta_{j,j'}$ is the Kronecker Delta.

For example, if $\Lambda_0 \equiv 1$, then $A_0 \equiv 1$, and $A_1(\chi) = \frac{\chi_1 - \mathbb{E}(\chi_1)}{\sigma(\chi_1)}$ where $\sigma^2(.)$ is the variance.

Now the function $\tilde{A}_j$ are completly defined by the matrix of variances covariances.

Lemma 2.1.1 *For all $j \in \{0, 1, ..., h\}$, we set $\tilde{A}_j(\chi) = \sum_{t=0}^{j} \tilde{a}_{j,t} \chi_t$. Then, there exists rational functions $\psi_{j,t}$ and η_j such that, for all random vector Λ , and for all (j,t), $\tilde{a}_{j,t} = \psi_{j,t}\big(\{\tau_{r,s}\}\big)$ and $||\tilde{A}_j||^2 = \eta_j\Big(\{\tau_{r,s}\}\Big)$, $0 \leq r \leq s \leq j$, when $\tau_{r,s} = \mathbb{E}\{\Lambda_r \Lambda_s\}$, $0 \leq r \leq s \leq j$.*

In particular,orthogonal polynomials are completly defined by the moments.

Now, one can estimate the $\tilde{A}_j$ under weak assumptions.

Proposition 2.1.1 *Let $\{\Lambda_{\ell.}\}_{\ell\in\mathbb{N}}$, $\Lambda_{\ell.} = \big(\Lambda_{\ell,0}, \Lambda_{\ell,1}, ..., \Lambda_{\ell,h}\big) \in \mathbb{R}^{h+1}$, be a sequence of random vectors such that $(1/n)\sum_{\ell=1}^{n} \Lambda'_{\ell.}\Lambda_{\ell.} \stackrel{p}{\to} \mathbb{E}\{\Lambda'\Lambda\}$ where $\mathcal{M}'$ is the transpose of the matrix $\mathcal{M}$.*

For all $n \in \mathbb{N}^*$, we denote by μ_n the empirical measure associated at the sample $\{\Lambda_{\ell.}\}_{\ell=1,2,..,n}$. We denote by $<,>_n$ and $||\,||_n$ the scalar product and the norm of $L^2(\mathbb{R}^{h+1}, \mu_n)$. For all $n \in \mathbb{N}^*$ and for all $\chi \in \mathbb{R}^{h+1}$, we set $\tilde{A}^n_{-1}(\chi) = A^n_{-1}(\chi) = 0$,

$$\textit{and, for } h \geq j \geq 0,\ \tilde{A}^n_j(\chi) = \chi_j - \sum_{s=-1}^{j-1} <\chi_j, A^n_s>_n A^n_s(\chi),$$

$$A^n_j(\chi) = \begin{pmatrix} \frac{\tilde{A}^n_j(\chi)}{||\tilde{A}^n_j||_n} & if ||\tilde{A}^n_j||_n \neq 0, \\ 0 & if ||\tilde{A}^n_j||_n = 0. \end{pmatrix}..$$

Then, for all $(j, j') \in \{0, 1, ..., h\}^2$, $\int A^n_j A^n_{j'}.d\mu_n = \delta_{j,j'}$ if $||\tilde{A}^n_s||_n \neq 0$ for s=0,1,...,max(j,j').

Notations 2.1.2 *For all $j \in \{0, 1, ..., h\}$, we set $\tilde{A}^n_j = \tilde{A}_j + \sum_{s=0}^{j} \tilde{\alpha}^n_{j,s} A_s$ and $A^n_j = A_j + \sum_{s=0}^{j} \alpha^n_{j,s} A_s$ and we define the matrices $\tilde{\alpha}^n$ and α^n by $\tilde{\alpha}^n = \{\{\tilde{\alpha}^n_{j,s}\}\}_{(j,s)\in\{0,1,....,h\}^2}$ and $\alpha^n = \{\{\alpha^n_{j,s}\}\}_{(j,s)\in\{0,1,....,h\}^2}$ by $\alpha^n_{j,s} = \tilde{\alpha}^n_{j,s} = 0$ if $s > j$.*

Remark that $\tilde{\alpha}^n_{j,j} = 0$, i.e. $\tilde{A}^n_j = \tilde{A}_j + \sum_{s=0}^{j-1} \tilde{\alpha}^n_{j,s} A_s$.

Now the $\tilde{A}^n_j$'s are estimators of the $\tilde{A}_j$.

Theorem 10 *With the previous notations $\alpha^n \stackrel{p}{\to} 0$ and $\tilde{\alpha}^n \stackrel{p}{\to} 0$. Moreover, if $\{\Lambda_\ell\}$ is IID, $\alpha^n \stackrel{a.s.}{\to} 0$ and $\tilde{\alpha}^n \stackrel{a.s.}{\to} 0$.*

Now, in order to obtain asymptotic distributions of α^n and $\tilde{\alpha}^n$, we need to use stochastics "O(.)" and "o(.)" (cf [9] page 8, section 1.2.5).

Notations 2.1.3 *A sequence of random variable X_n is bounded in probability, if, for every $\epsilon > 0$, there exists M_ϵ and N_ϵ such that $P\{|X_n| \leq M_\epsilon\} \geq 1 - \epsilon$ for all $n \geq N_\epsilon$. Then, one writes $X_n = O_P(1)$.*

Moreover, we write $X_n = O_P(Z_n)$ for two sequences of random variable X_n and Z_n, if $X_n/Z_n = O_P(1)$ and $X_n = o_P(Z_n)$ if $X_n/Z_n \stackrel{p}{\to} 0$.

In the vector case, we define the stochastic o_p and O_p by the following way. For example, we denote $(Z_{n,0}, Z_{n,1}, ..., Z_{n,h}) = o_p(\phi(n)^{-1})$ if $Z_{n,s} = o_p(\phi(n)^{-1})$ for all s=0,1,...,h, and we do the same for O_p.

In particular, $X_n = O_P(1)$ if $X_n \stackrel{d}{\to} X$ (cf also Problem 1.P.3 of [9]). Then, the following result allow to know asymptotic distributions of A^n_j.

Theorem 11 *Let $\phi(n) > 0$ be a real sequence such that $\phi(n) \to \infty$ as $n \to \infty$. Assume $\mathbb{E}\{\Lambda^4_s\} < \infty$ for all s=0,1,..,h. We suppose that*

$$\frac{\phi(n)}{n}\big[\sum_{\ell=1}^{n}\big[\Lambda'_{\ell.}\Lambda_{\ell.}\big] - \mathbb{E}\{\Lambda'\Lambda\}\big] = O_p(1)\ .$$

Then,

$$\alpha^n = e^n + o_p(\phi(n)^{-1})$$

where $e^n = \{\{\int \mathcal{J}_{i,s}.d\mu_n\}\}_{(i,s)\in\{0,1,...,h\}^2}$ *with* $\mathcal{J}_{i,s}(\chi) = -A_i(\chi)A_s(\chi)$ *if* $s < i$, $\mathcal{J}_{i,i}(\chi) = \frac{1-A_i(\chi)^2}{2}$ *if s=i, and* $\mathcal{J}_{i,s} \equiv 0$ *if* $s > i$.

Moreover,

$$\tilde{\alpha}^n = \tilde{e}^n + o_p(\phi(n)^{-1})$$

where $\tilde{e}^n = \{\{\int \tilde{\mathcal{J}}_{i,s}.d\mu_n\}\}_{(i,s)\in\{0,1,...,h\}^2}$ *with* $\tilde{\mathcal{J}}_{i,s}(\chi) = -\tilde{A}_i(\chi)A_s(\chi)$ *if* $s < i$, *and* $\tilde{\mathcal{J}}_{i,s} \equiv 0$ *if* $s \geq i$.

This result is remarkable because by elementary properties of orthogonal functions, $\alpha^n_{i,s} = \int A^n_i A_s d\mu$ if $i < s$ and $\alpha^n_{i,i} = \int A^n_i A_i d\mu - 1$.

2.1.2 Proofs

At first, we introduce the following notations.

Notations 2.1.4 *For all* $(i,s) \in \{0,1,...,h\}^2$, *we set* $\tilde{\alpha}^n_{-1,s} = \tilde{\alpha}^n_{i,-1} = \alpha^n_{-1,s} = \alpha^n_{i,-1} = 0$. *We set* $A = (A_0, A_1, ..., A_h)$, $A^n = (A^n_0, A^n_1, ..., A^n_h)$. *For all* $i \in \{0,1,...,h\}$, *we set* $[A[_i = (A_{-1}, A_0, A_1, ..., A_{i-1})$, $[A^n[_i = (A^n_{-1}, A^n_0, A^n_1, ..., A^n_{i-1})$, $\tilde{\alpha}^n_i = (\tilde{\alpha}^n_{i,0}, \tilde{\alpha}^n_{i,1}, ..., \tilde{\alpha}^n_{i,h})$, *and* $[\alpha^n[_i = \{\{\alpha^n_{j,s}\}\}_{(j,s)\in\{0,1,....,i-1\}^2}$.

With these notations, the following result is easily proved.

Lemma 2.1.2 *Under the previous notations ,* $\tilde{\alpha}^n_i = (\tilde{\alpha}^n_{i,0}, \tilde{\alpha}^n_{i,1}, ..., \tilde{\alpha}^n_{i,i-1}, 0, ..., 0)$. *Moreover* $(A^n)' = A' + \alpha^n A'$.

On the other hand, $\tilde{A}_i = \chi_i - \big(\int \chi_i [A[_i dm\big)([A[_i)'$, $\tilde{A}^n_i = \chi_i - \big(\int \chi_i [A^n[_i dm_n\big)([A^n[_i)'$, $([A^n[_i)' = ([A[_i)' + [\alpha^n[_i([A[_i)'$ *and* $\tilde{A}^n_i = \tilde{A}_i + \tilde{\alpha}^n_i(A)' = \tilde{A}_i + A(\tilde{\alpha}^n_i)'$.

We deduce the following lemma

Lemma 2.1.3 *For all* $i \in \{0,1,...,h\}$, *the following equalities hold :*

$$a)\ \tilde{A}^n_i = \tilde{A}_i + \Big(\int \chi_i [A[_i dm - \int \chi_i [A[_i dm_n\Big)([A[_i)' - \Big(\int \chi_i [A[_i dm_n\Big)\big([\alpha^n[_i + ([\alpha^n[_i)'\big)([A[_i)'$$

$$..... - \Big(\int \chi_i [A[_i dm_n\Big)([\alpha^n[_i)'[\alpha^n[_i([A[_i)' \ ,$$

$$b)\ \int \tilde{A}^n_i \tilde{A}^n_i dm_n = \int \tilde{A}_i \tilde{A}_i dm_n + \tilde{\alpha}^n_i\Big(\int A' \tilde{A}_i dm_n\Big) + \Big(\int \tilde{A}_i A dm_n\Big)(\tilde{\alpha}^n_i)' + \tilde{\alpha}^n_i\Big(\int A' A dm_n\Big)(\tilde{\alpha}^n_i)' \ .$$

c) If $i \neq s$, $\alpha^n_{i,s} = \frac{\tilde{\alpha}^n_{i,s}}{||\tilde{A}^n_i||_n}$ *if* $||\tilde{A}^n_i||_n \neq 0$, $\alpha^n_{i,s} = 0$ *if* $||\tilde{A}^n_i||_n = 0$.

$$\alpha^n_{i,i} = \frac{||\tilde{A}_i||^2 - ||\tilde{A}^n_i||^2_n}{\big(||\tilde{A}_i|| + ||\tilde{A}^n_i||_n\big)||\tilde{A}^n_i||_n} \ \text{ if } ||\tilde{A}^n_i||_n \neq 0,\ \alpha^n_{i,i} = -1 \ \text{ if } ||\tilde{A}^n_i||_n = 0,$$

Proof of theorem 10 We prove by recurence on i that $\tilde{\alpha}^n_{i,s}$ and $\alpha^n_{i,s}$ converge in probability to 0 for every $s \in \{-1, 0, 1, ..., h\}$.

If i=-1, the result is obvious : $\alpha^n_{-1,s} = \tilde{\alpha}^n_{-1,s} = 0$.

Now, we suppose that, for every all $(s,t) \in \{-1, 0, 1, ..., i-1\} \times \{-1, 0, 1, ..., h\}$, $\alpha^n_{s,t} \xrightarrow{p} 0$.

By our assumption, $\int \chi_i [A[_i dm_n \xrightarrow{p} \int \chi_i [A[_i dm$. Then, by lemma 2.1.3-a, $\tilde{\alpha}^n_i \xrightarrow{p} 0$ and $\tilde{\alpha}^n_{i,s} \xrightarrow{p} 0$.

Now, $\int A_i A_s dm_n \xrightarrow{p} \int A_i A_s dm$. Then, by lemma 2.1.3-b, we deduce $||\tilde{A}^n_i||_n \xrightarrow{p} ||\tilde{A}_i||$.

Since $\Lambda_0, \Lambda_1, ..., \Lambda_h$ are linearly independent, $||\tilde{A}_i|| \neq 0$. Let g be the function g(a)=1/a if $a \neq 0$, and g(0)=0. Then, $g\big(||\tilde{A}^n_i||_n\big) \xrightarrow{p} ||\tilde{A}_i||^{-1}$ (cf page 24 [9]). Therefore, if $s < i$, by lemma 2.1.3-c , $\alpha^n_{i,s} = g\big(||\tilde{A}^n_i||_n\big)\tilde{\alpha}^n_{i,s} \xrightarrow{p} 0$.

We prove similarly that $\alpha^n_{i,i} \xrightarrow{p} 0$.

We prove the convergence with probability 1 by the same way. ■.

In order to prove theorem 11, we need the following lemma which one proves by means of elementary properties of sequences of random variables (cf [9] chapter 1).

Lemma 2.1.4 *Let K_n, Z_n and Z^*_n be three sequences of random variables defined on $(\Omega, \mathcal{A}, P)$ such that $\phi(n)Z_n = O_P(1)$, $\phi(n)Z^*_n = O_P(1)$ and $K_n \xrightarrow{p} K \in \mathbb{R}$.*

Then, $\phi(n)KZ_n = O_P(1)$, $\phi(n)K_nZ_n = O_P(1)$, $\phi(n)Z_n + \phi(n)Z^*_n = O_P(1)$, and $K_nZ_n = KZ_n + o_P(\phi(n)^{-1})$.

Moreover, $Z_n \xrightarrow{p} 0$ and $K_nZ_n \xrightarrow{p} 0$.

Finally, $Z_nZ^*_n = K_nZ_nZ^*_n + o_P(\phi(n)^{-1}) = KZ_nZ^*_n + o_P(\phi(n)^{-1}) = o_P(\phi(n)^{-1})$.

Now we can prove the following properties

Lemma 2.1.5 *Under the assumptions of theorem 11, $\phi(n)\alpha^n_{i,s} = O_P(1)$ for all $(i,s) \in \{-1, 0, 1, ..., h\}^2$.*

Proof We prove this lemma by recurence on i. If i=-1, the result is obvious : $\alpha^n_{-1,s} = 0$.

Let $i \in \{0, 1, ..., h\}$. We suppose that, for every all $(s,t) \in \{-1, 0, 1, ..., i-1\} \times \{-1, 0, 1, ..., h\}$, $\phi(n)\alpha^n_{s,t} = O_p(1)$.

Therefore, $\phi(n)[\alpha^n[_i = O_P(1)$. Moreover, $\phi(n)\big(\int \chi_i [A[_i dm_n - \int \chi_i [A[_i dm\big) = O_p(1)$ (lemma 2.1.4) and $\int \chi_i [A[_i dm_n \xrightarrow{p} \int \chi_i [A[_i dm$. Then, by lemma 2.1.4 and 2.1.3-a, $\phi(n)\tilde{\alpha}^n_{i,s} = O_p(1)$, i.e. $\phi(n)\tilde{\alpha}^n_i = O_p(1)$.

Therefore, if $s < i$, by by lemma 2.1.4 and 2.1.3-c, $\phi(n)\alpha^n_{i,s} = \phi(n)g\big(||\tilde{A}^n_i||_n\big)\tilde{\alpha}^n_{i,s} = O_p(1)$.

Moreover, by lemma 2.1.4, $\phi(n)\big(\int (\tilde{A}_i)^2 dm_n - ||\tilde{A}_i||^2\big) = O_p(1)$. Therefore, by lemma 2.1.4 and 2.1.3-b, $\phi(n)\big(||\tilde{A}^n_i||^2_n - ||\tilde{A}_i||^2\big) = O_p(1)$. We deduce $\phi(n)\alpha^n_{i,i} = O_P(1)$. ■

We deduce the following lemma.

Lemma 2.1.6 *Under the assumptions of theorem 11, $\phi(n)\tilde{\alpha}^n_{i,s} = O_P(1)$ for all $(i,s) \in \{0, 1, ..., h\}^2$.*

Proof By lemma 2.1.3-b, $||\tilde{A}^n_i||_n \xrightarrow{p} ||\tilde{A}_i||$. Then, if $i \neq s$, and if n large enough, by lemma 2.1.4, $\phi(n)\tilde{\alpha}^n_{i,s} = \phi(n)||\tilde{A}^n_i||_n \alpha^n_{i,s} = O_P(1)$. Moreover, if i=s, $\tilde{\alpha}^n_{i,s} = 0$. ■

We deduce the following lemma

Lemma 2.1.7 *Under the assumptions of theorem 11, $\phi(n)\int \tilde{A}^n_i \tilde{A}^n_i dm_n - \phi(n)\int \tilde{A}_i \tilde{A}_i dm_n \xrightarrow{p} 0$.*

Proof By lemma 2.1.3-b,

$$\int \tilde{A}_i^n \tilde{A}_i^n dm_n = \int \tilde{A}_i \tilde{A}_i dm_n + \tilde{\alpha}_i^n \big(\int A' \tilde{A}_i dm_n \big) + \big(\int \tilde{A}_i A dm_n \big) (\tilde{\alpha}_i^n)' + \tilde{\alpha}_i^n \big(\int A' A dm_n \big) (\tilde{\alpha}_i^n)' \ .$$

Now, $A = (A_0, A_1, ..., A_h)$ and $\tilde{\alpha}_i^n = (\tilde{\alpha}_{i,0}^n, \tilde{\alpha}_{i,1}^n, ..., \tilde{\alpha}_{i,i-1}^n, 0, 0, ..., 0)$. Then, $\phi(n)\tilde{\alpha}_i^n \big(\int A' \tilde{A}_i dm_n \big)$ $= \sum_{s=0}^{i-1} \phi(n) \tilde{\alpha}_{i,s}^n \int A_s \tilde{A}_i dm_n \xrightarrow{p} 0$ because $\phi(n)\tilde{\alpha}_{i,s}^n = O_p(1)$.

By the same way, $\phi(n) \big(\int \tilde{A}_i A dm_n \big) (\tilde{\alpha}_i^n)' \xrightarrow{p} 0$.

Moreover, $\tilde{\alpha}_i^n \xrightarrow{p} 0$. Then, $\phi(n)\tilde{\alpha}_i^n \big(\int A' A dm_n \big) (\tilde{\alpha}_i^n)' \xrightarrow{p} 0$. ■

Now we can prove the fundamental theorem.

Proof of first equations of theorem 11. Let $\alpha_i^n = (\alpha_{i,0}^n, \alpha_{i,1}^n, ..., \alpha_{i,h}^n)$, i.e. $A_i^n = A_i + \alpha_i^n A'$ and $A_s^n = A_s + A(\alpha_s^n)'$. Then,

$$< A_i^n, A_s^n >_n = \int A_i A_s dm_n + \alpha_i^n \big(\int A' A_s dm_n \big) + \big(\int A_i A dm_n \big) (\alpha_s^n)' + \alpha_i^n \big(\int A' A dm_n \big) (\alpha_s^n)'$$

$$= \int A_i A_s dm_n + \alpha_{i,s}^n \big(\int A_s A_s dm_n \big) + \alpha_{s,i}^n \big(\int A_i A_i dm_n \big) + o_p(\phi(n)^{-1}) \ (cflemma\ 2.1.4)$$

$$= \int A_i A_s dm_n + \alpha_{i,s}^n + \alpha_{s,i}^n + o_p(\phi(n)^{-1}) \ (cflemma\ 2.1.4) \ .$$

Then, $1 = \int A_i A_i dm_n + 2\alpha_{i,i}^n + o_p(\phi(n)^{-1})$ and $0 = \int A_i A_s dm_n + \alpha_{i,s}^n + o_p(\phi(n)^{-1})$ if $s < i$. ■

Proof of second equations of theorem 11. If $i > s$,
$\tilde{\alpha}_{i,s}^n = ||\tilde{A}_i^n||_n \alpha_{i,s}^n = ||\tilde{A}_i^n||_n [- \int A_i A_s dm_n + o_p(\phi(n)^{-1})]$ where $\phi(n)[\int A_i A_s dm_n - 0] = O_p(1)$. Now, by lemma 2.1.4, $||\tilde{A}_i^n||_n \int A_i A_s dm_n = ||\tilde{A}_i|| \int A_i A_s dm_n + o_p(\phi(n)^{-1}) = \int \tilde{A}_i A_s dm_n + o_p(\phi(n)^{-1})$.
■

2.1.3 Asymptotic distribution

It is easy to deduce from theorem 11 the asymptotic distributions of $\tilde{A}^n$ and A^n. For example if Λ_ℓ is IID and if $\mathbb{E}\{\Lambda_\ell^4\}\big) < \infty$, on can apply the CLT. For example, if $s < i$, $\sqrt{n} \int A_i A_s dm_n$ has asymptotically a normal distribution. Then, $\sqrt{n}\{\{\alpha_{i,s}^n\}\}$ has asymptotically a normal distribution. Then, it is enough to apply classical theorems in order to deduce the asymptotic covariance of $\sqrt{n}\{\{\alpha_{i,s}^n\}\}$.

2.2 Estimation of higher order variances

We will deduct from the previous results estimators of Gauss Jacobi Quadrature and of Higher order variances.

At first, we need empirical orthogonal polynomials.

Notations 2.2.1 *Let $j \in \mathbb{N}$ such that P_j exists. Let $\{X_\ell\}_{\ell \in \mathbb{N}}$, $X_\ell \in \mathbb{R}$, be a sequence of random variables such that $(1/n) \sum_{\ell=1}^n X_\ell^s \xrightarrow{p} \mathbb{E}\{X^s\} < \infty$ for all $s \in \mathbb{N}$.*

For all $n \in \mathbb{N}^*$, we denote by m_n the empirical measure associated at the sample $\{X_\ell\}_{\ell=1,2,..,n}$.

Under these conditions the assumptions of the proposition 2.1.1 hold.

Proposition 2.2.1 *Let $h \in \mathbb{N}*$. We set $\{\Lambda_\ell\}_{\ell\in\mathbb{N}}$, $\Lambda_\ell = \big(X_\ell^0, X_\ell^1, ..., X_\ell^h\big)$. Then,*

$$(1/n)\sum_{\ell=1}^{n} \Lambda'_\ell \Lambda_\ell \xrightarrow{p} \mathbb{E}\{\Lambda'\Lambda\} \ .$$

Proof We know that a sequence of random vectors $Y_n \xrightarrow{p} Y$ if $P\big\{\big|Y_n - Y\big| < a\big\} \to 1$ as $n \to \infty$ for every $a > 0$. Donc, $P\big\{\big|Y_n - Y\big| \geq a\big\} = P\big\{\complement\{\big|Y_n - Y\big| < a\big\}\ \} = 1 - P\big\{\big|Y_n - Y\big| < a\big\} \to 1 - 1 = 0$.

One can choose in $\mathbb{R}^{h+1}$ the maximum norm : $||(x_0, x_1, ..., x_h)|| = Max_{s=0,1,...,h}(|x_s|)$. Then, convergence in probability of Λ_ℓ means that

$$P\big\{\big|\big|(1/n)\sum_{\ell=1}^{n} \Lambda_\ell - \mathbb{E}\{\Lambda\}\big|\big| \geq a\big\} = P\big\{Max_{s=0,1,...,h}\big|(1/n)\sum_{\ell=1}^{n} X_\ell^s - \mathbb{E}\{X^s\}\big| \geq a\big\} \to 0.$$

Let $a > 0$. Then, for all $\epsilon > 0$, for all $s \in \{0, 1, ..., h\}$, there exists N_s such that if $n \geq N_s$, $P\big\{\big|(1/n)\sum_{\ell=1}^{n} X_\ell^s - \mathbb{E}\{X^s\}\big| \geq a\big\} \leq \epsilon/(h+1)$.
Now, if $n \geq Ma = Max_{s=0,1,...,h}(N_s)$,

$$P\big\{Max_s\big|(1/n)\sum_{\ell=1}^{n} X_\ell^s - \mathbb{E}\{X^s\}\big| \geq a\big\} \leq P\big\{\ \cup_s \ \big|(1/n)\sum_{\ell=1}^{n} X_\ell^s - \mathbb{E}\{X^s\}\big| \geq a\big\}$$

$$\leq \sum_s P\big\{\big|(1/n)\sum_{\ell=1}^{n} X_\ell^s - \mathbb{E}\{X^s\}\big| \geq a\big\} \leq (h+1)(\epsilon/(h+1)) = \epsilon \ .$$

That prove that $\sum_{\ell=1}^{n} \Lambda_\ell \xrightarrow{p} \mathbb{E}\{\Lambda_\ell\}$. We deduce that $(1/n)\sum_{\ell=1}^{n} \Lambda'_\ell \Lambda_\ell \xrightarrow{p} \mathbb{E}\{\Lambda'\Lambda\}$ by changing h in 2h. ■

Then, we can define empirical orthogonal polynomials.

Notations 2.2.2 *Let $\{\tilde{P}_j^n\}_{j=0,1,..,h}$ and $\{P_j^n\}_{j=0,1,..,h}$ be the family of orthogonal polynomials such that, for all $i \leq h$, $\tilde{P}_i^n(x) = \tilde{A}_i^n(1, x, x^2, ..., x^h)$, $P_i^n(x) = A_i^n(1, x, x^2, ..., x^h)$.*

Notations 2.2.3 *We set $P_j(x) = \sum_{t=0}^{j} a_{j,t}x^t$, $\tilde{P}_j(x) = \sum_{t=0}^{j} \tilde{a}_{j,t}x^t$ and $P_j^n(x) = \sum_{t=0}^{j} a_{j,t}^n x^t$, $\tilde{P}_j^n(x) = \sum_{t=0}^{j} \tilde{a}_{j,t}^n x^t$.*

Then, the $\tilde{P}_i^n$'s are the orthogonal polynomials associated to the empirical measure m_n.

Proposition 2.2.2 *Under the previous notations, $\tilde{P}_i(x) = \tilde{A}_i(1, x, x^2, ..., x^h)$, $P_i(x) = A_i(1, x, x^2, ..., x^h)$.*
Moreover, there exists N such if $n \geq N$, $\{P_i^n\}_{i=0,1,..,h}$ is the family of orhonormal polynomials of $L^2(\mathbb{R}, m_n)$ and $\{\tilde{P}_i^n\}_{i=0,1,..,h}$ is the family of orthogonal polynomials of $L^2(\mathbb{R}, m_n)$ such that $\tilde{a}_{j,j}^n = 1$.

We deduce from theorem 10 that $\tilde{P}_i^n$ is an estimate of $\tilde{P}_i$.

Proposition 2.2.3 *For all $i \leq h$, $\tilde{P}_i^n = P_i + \sum_{s=0}^{i-1} \tilde{\alpha}_{i,s}^n P_s$ and $P_i^n = P_i + \sum_{s=0}^{i} \alpha_{i,s}^n P_s$, where, for all (j,s), $\tilde{\alpha}_{j,s}^n \xrightarrow{p} 0$ and $\alpha_{j,s}^n \xrightarrow{p} 0$. Moreover, for all (i,s), $\tilde{\alpha}_{i,s}^n$ and $\alpha_{i,s}^n$ converges almost surely to 0 if X_ℓ is IID.*

Now, one can define empirical Gauss Jacobi Quadrature and empirical variances of order j.

Notations 2.2.4 *Let $j \in \mathbb{N}$. We denote by $(\sigma_j^n)^2$ the variance of order j of m_n, by $x_{j,1}^n, x_{j,2}^n, ..., x_{j,j}^n$ the zeros of $\tilde{P}_j^n$, by $\lambda_{j,1}^n, \lambda_{j,2}^n, ..., \lambda_{j,j}^n$ the weights of Gauss Jacobi quadrature, if these numbers exists. If not, one defines theses variables by 0.*

Remark that $(\sigma_1^n)^2$ is the classical empirical variances. Moreover $(\sigma_2^n)^2 = M_4^n - \frac{(M_3^n - M_1^n M_2^n)^2}{M_2^n - (M_1^n)^2} - (M_2^n)^2$, where M_j^n is the empirical moment of order j.

Clearly, these estimators converges almost surely if X_ℓ is IID. More generally, the following proposition holds.

Proposition 2.2.4 *Under the previous assumptions, $\sigma_j^n \xrightarrow{p} \sigma_j$. Moreover, for all j=1,2,..,h, for all s=1,2,..,j, $x_{j,s}^n \xrightarrow{p} x_{j,s}$ and $\lambda_{j,s}^n \xrightarrow{p} \lambda_{j,s}$, respectively. These estimators converges almost surely if X_ℓ is IID.*

Proof Clearly $\tilde{a}_{j,s}^n \xrightarrow{p} \tilde{a}_{j,s}$ and $a_{j,s}^n \xrightarrow{p} a_{j,s}$. Then, by theorem page 24 of [9], $g(\tilde{a}_{j,0}^n, \tilde{a}_{j,1}^n, ..., \tilde{a}_{j,j}^n) \xrightarrow{p} g(\tilde{a}_{j,0}, \tilde{a}_{j,1}, ..., \tilde{a}_{j,j})$ if g is continous with P-probability 1. Then, for example, $x_{j,s}^n \xrightarrow{p} x_{j,s}$.
One prove the almost sure convergence by the same way ■.

In order to study asymptotic distributions, we will suppose assumptions as those of theorem 11 (cf also section 2.1.3). But they can be complicated. Also in order to simplify we assume from now that X_ℓ is IID.

Notations 2.2.5 *We suppose now that $\{X_\ell\}_{\ell\in\mathbb{N}}$, $X_\ell \in \mathbb{R}$, is an IID sequence of random vectors such that $X_0 = X$ and $\mathbb{E}\{X^{4h}\} < \infty$.*
For all $n \in \mathbb{N}^*$, we denote by $<,>_n$ and $||\,||_n$ the scalar product and the norm of $L^2(\mathbb{R}^{h+1}, m_n)$.

Indeed, for example we want to know the asymptotic distributions of empirical variances. But if we keep the hypothesis of Theorem 11, we assume then that $\frac{\phi(n)}{n}\left[\sum_{\ell=1}^n \left[\Lambda_{\ell.}'\Lambda_{\ell.}\right] - \mathbb{E}\{\Lambda'\Lambda\}\right] \xrightarrow{p} G$ where G is a random vector. This can give assumptions and notations too complicated (e.g. $\Lambda_{\ell.} = \left(X_\ell^0, X_\ell^1, ..., X_\ell^h\right)$). So we just give the results in the IID case. But the reader can easily obtain the same type of results under weakest hypotheses, for example under the strong mixing assumption.

Now, by the CLT, one can apply the result of section 2.1 with $\phi(n) = \sqrt{n}$

Lemma 2.2.1 *Under the previous assumptions , $\tilde{\alpha}^n = \tilde{e}^n + o_p\left(n^{-1/2}\right)$ and $\alpha^n = e^n + o_p\left(n^{-1/2}\right)$.*

Indeed, clearly, by the CLT, $\frac{\sqrt{n}}{n}\left[\sum_{\ell=1}^n \left[\Lambda_{\ell.}'\Lambda_{\ell.}\right] - \mathbb{E}\{\Lambda'\Lambda\}\right] \xrightarrow{p} G$ where G is a random vector such that $G \sim N(0, C)$. Then, assumptions of theorem 11 hold.

At first, one generalizes the result about the asymptotical distribution of empirical classical variance.

Theorem 12 *Under the previous assumptions, $\sqrt(n)\left[(\sigma_j^n)^2 - \sigma_j^2\right]$ has asymptotically a normal distribution with mean 0 and variance $\mathbb{E}\left\{\tilde{P}_j(X)^4\right\} - \sigma_j^4$.*

Proof : By lemma 2.1.7, $\sqrt{n}\Big(\int \tilde{P}_j^n \tilde{P}_j^n.dm_n - \int \tilde{P}_j\tilde{P}_j.dm\Big)$ has asymptotically the same distribution as $\sqrt{n}\Big(\int \tilde{P}_j\tilde{P}_j.dm_n - \sigma_j^2\Big) = \sum_{\ell=1}^n \frac{\tilde{P}_j(X_\ell)\tilde{P}_j(X_\ell) - \sigma_j^2}{\sqrt{n}}$. Then, it is enough to use the Central Limit Theorem. ■

We obtain now the asymptotical distribution of the estimators of Gauss Jacobi Quadrature.

Theorem 13 *For all s=1,2,...,j, we set $x_{j,s}^n = x_{j,s} + \eta_s$. Then, $\sqrt{n}\{\eta_s\}_{s=1,..,j}$ has asymptotically a normal distribution with mean 0 and covariance matrix $\{G_{s,t}\}_{(s,t)\in\{1,2,...,j\}^2}$ where*

$$G_{s,t} = \mathbb{E}\left\{\frac{\left[\sum_{v=0}^{j-1} P_v(x_{j,s})P_v(X)P_j(X)\right]\left[\sum_{v=0}^{j-1} P_v(x_{j,t})P_v(X)P_j(X)\right]}{P_j'(x_{j,s})P_j'(x_{j,t})}\right\} .$$

This theorem is deduced from the following lemma by using the CLT.

Lemma 2.2.2 *For all s = 1,2,...,j,*

$$\eta_s = \frac{1}{\tilde{P}'_j(x_{j,s})} \int \tilde{P}_j(t)\Big(\sum_{v=0}^{j-1} P_v(x_{j,s})P_v(t)\Big).m_n(dt) + o_p(n^{-1/2})\,.$$

Proof We prove this lemma for s=1. First, one proves that $\sqrt{n}\ \eta_1$ is asymptotically normal. By notations 2.1.2, we know that

$$\tilde{P}_j^n \ = \ \tilde{P}_j + \sum_{v=0}^{j-1} \tilde{\alpha}_{j,v}^n P_v\,.$$

Therefore, because $\tilde{P}_j^n(x_{j,1}^n) = 0$,

$$\tilde{P}_j(x_{j,1}^n) \ = \ -\sum_{v=0}^{j-1} \tilde{\alpha}_{j,v}^n P_v(x_{j,1}^n)\,.$$

Therefore, $\sqrt{n}\tilde{P}_j(x_{j,1}^n) \ = \ \sqrt{n}(x_{j,1}^n - x_{j,1})(x_{j,1}^n - x_{j,2})....(x_{j,1}^n - x_{j,j})$ is asymptotically normal. Moreover $(x_{j,1}^n - x_{j,2})^{-1}....(x_{j,1}^n - x_{j,j})^{-1}$ converges almost surely to $(x_{j,1} - x_{j,2})^{-1}....(x_{j,1} - x_{j,j})^{-1}$. Therefore,by the theorem of page 19 of [9] $\sqrt{n}(x_{j,1}^n - x_{j,1})$ is asymptotically normal, i.e. $\sqrt{n}\ \eta_1$ is asymptotically normal.

By the same way, one proves that $\sqrt{n}\ \eta_s$ is asymptotically normal for s=2,3,...,j.

Now one can prove the lemma. Indeed, one can write the following equalities :

$$\tilde{P}_j^n(x) \ = (x - x_{j,1} - \eta_1)(x - x_{j,2} - \eta_2).....(x - x_{j,j} - \eta_j)$$

$$\begin{array}{ll}
= & (x - x_{j,1})(x - x_{j,2}).....(x - x_{j,j}) \\
- & \eta_1(x - x_{j,2})(x - x_{j,3}).....(x - x_{j,j}) \\
- & \eta_2(x - x_{j,1})(x - x_{j,3}).....(x - x_{j,j}) \\
 & ..... \\
- & \eta_j(x - x_{j,1})(x - x_{j,2}).....(x - x_{j,j-1}) \ + o_p(n^{-1/2}) \ ,
\end{array}$$

by lemma 2.1.4. Therefore,

$$\tilde{P}_j^n(x_{j,1}) \ = -\eta_1(x_{j,1} - x_{j,2})(x_{j,1} - x_{j,3}).....(x_{j,1} - x_{j,j}) + \ o_p(n^{-1/2}) = -\eta_1\tilde{P}'_j(x_{j,1}) + o_p(n^{-1/2})\,.$$

Now, by theorem 11,

$$\begin{aligned}
\tilde{P}_j^n(x_{j,1}) \quad &= \quad \tilde{P}_j(x_{j,1}) + \sum_{v=0}^{j-1} \tilde{\alpha}_{j,v}^n P_v(x_{j,1}) \\
&= \quad \sum_{v=0}^{j-1} \tilde{\alpha}_{j,v}^n P_v(x_{j,1}) \\
&= \quad \sum_{v=0}^{j-1} \Big(- \int \tilde{P}_j P_v dm_n \Big) P_v(x_{j,1}) + o_p(n^{-1/2})\,.
\end{aligned}$$

We deduce the lemma. ■

Theorem 14 *Let $j \in \mathbb{N}$ such that $\mathbb{E}\{X^{4j}\} < \infty$. For all s=1,2,...,j, and for all u=1,2,..,j, we set $L_s(x) = \frac{\tilde{P}_j(x)}{x-x_{j,s}}$ and $h_u^s(x) = \frac{\tilde{P}_j(x)}{(x-x_{j,s})(x-x_{j,u})}$. We define D_u^s by*

$$\begin{aligned} D_u^s &= \sum_{r=1,r\neq s}^{j} \left(\frac{h_r^s(x_{j,s})\mathbb{E}\{L_s(X)\}}{L_s(x_{j,s})} - \mathbb{E}\{h_r^s(X)\} \right) \left(\frac{P_u(x_{j,r})}{L_s(x_{j,s})\tilde{P}_j'(x_{j,r})} \right) \\ &\quad - \frac{P_u(x_{j,s})\mathbb{E}\{L_s(X)\}}{L_s(x_{j,s})^3} \sum_{r=1,s\neq r}^{j} h_r^s(x_{j,s}) \, . \end{aligned}$$

Then, $\sqrt{n}\{\lambda_{j,s}^n - \lambda_{j,s}\}_{s=1,..,j}$ has asymptotically the normal distribution with mean 0 and covariance matrix $\{O_{s,t}\}_{1\leq s,t\leq j}$ where

$$O_{s,t} = \mathbb{E}\left\{ \left[\ell_s(X) + \sum_{u=0}^{j-1} D_u^s P_u(X)\tilde{P}_j(X) \right] \left[\ell_t(X) + \sum_{u=0}^{j-1} D_u^t P_u(X)\tilde{P}_j(X) \right] \right\} - \lambda_{j,s}\lambda_{j,t} \, .$$

This theorem is deduced from the following lemma by using CLT.

Lemma 2.2.3 *For all $s \in \{1, 2, .., j\}$,*

$$\lambda_{j,s}^n = \int \ell_s(t).m_n(dt) + \int \Big(\sum_{u=0}^{j-1} D_u^s P_u(t)\tilde{P}_j(t) \Big).m_n(dt) + o_p(n^{-1/2}) \, .$$

Proof We prove this lemma for s=1. We simplify $L^1(t)$ in L(t) and h_r^1 in h_r : $L(t) = (t-x_{j,2})(t-x_{j,3})....(t-x_{j,j})$, $h_2(t) = (t-x_{j,3})(t-x_{j,4})....(t-x_{j,j})$, $h_3(t) = (t-x_{j,2})(t-x_{j,4})....(t-x_{j,j})$ etc. Moreover, we set $L^n(t) = (t-x_{j,2}^n)(t-x_{j,3}^n)....(t-x_{j,j}^n)$.

We know that

$$\begin{aligned} \lambda_{j,1} &= \int \ell_1(x).m(dx) \\ &= \frac{\int (t-x_{j,2})(t-x_{j,3})...(t-x_{j,j}).m(dt)}{(x_{j,1}-x_{j,2})(x_{j,1}-x_{j,3})...(x_{j,1}-x_{j,j})} \\ &= \frac{\int L(t).m(dt)}{L(x_{j,1})} \end{aligned}$$

and that

$$\begin{aligned} \lambda_{j,1}^n &= \int \ell_1^n(x).m_n(dx) \\ &= \frac{\int (t-x_{j,2}^n)(t-x_{j,3}^n)...(t-x_{j,j}^n).m_n(dt)}{(x_{j,1}^n-x_{j,2}^n)(x_{j,1}^n-x_{j,3}^n)...(x_{j,1}^n-x_{j,j}^n)} \\ &= \frac{\int L^n(t).m_n(dt)}{L^n(x_{j,1}^n)} \quad . \end{aligned}$$

By proposition 2.2.4, we deduce easily that $\int L^n(t).m_n(dt)$ and $L^n(x_{j,1}^n)$ converge almost surely to $\int L(t).m(dt)$ and $L(x_{j,1})$, respectively. Therefore,

$$\lambda_{j,1}^n \; - \lambda_{j,1} \;\; = \;\; \frac{\int L^n(t).m_n(dt)}{L^n(x_{j,1}^n)} \;\; - \;\; \frac{\int L(t).m(dt)}{L(x_{j,1})}$$

$$= \;\; \frac{\int L^n(t).m_n(dt) \; - \; \int L(t).m(dt)}{L^n(x_{j,1}^n)} \;\; + \;\; \frac{\int L(t).m(dt)}{L^n(x_{j,1}^n)} - \;\; \frac{\int L(t).m(dt)}{L(x_{j,1})}$$

$$= \;\; \frac{\int L^n(t).m_n(dt) \; - \; \int L(t).m(dt)}{L(x_{j,1})} \; - \; \frac{\Big[\int L(t).m(dt)\Big]\Big[L^n(x_{j,1}^n) \; - \; L(x_{j,1})\Big]}{L(x_{j,1})^2} \; + \; o_p(n^{-1/2}) \; ,$$

if $\sqrt(n)\big[\int L^n(t).m_n(dt) \; - \; \int L(t).m(dt)\big]$ and $\sqrt{n}\big[L^n(x_{j,1}^n) \; - \; L(x_{j,1})\big]$ are asymptotically normal. Then, we prove this result now. Indeed,

$$\int L^n(t).m_n(dt) \; - \; \int L(t).m(dt)$$

$$= \int (t - x_{j,2}^n)(t - x_{j,3}^n)....(t - x_{j,j}^n).m_n(dt) \; - \; \int (t - x_{j,2})(t - x_{j,3})....(t - x_{j,j}).m(dt)$$

$$\begin{aligned}
&= \int (t - x_{j,2})(t - x_{j,3})....(t - x_{j,j}).[m_n - m](dt) \\
&- \; \eta_2 \int (t - x_{j,3})(t - x_{j,4})....(t - x_{j,j}).m_n(dt) \\
&- \; \eta_3 \int (t - x_{j,2})(t - x_{j,4})....(t - x_{j,j}).m_n(dt) \\
&- \; ..... \\
&- \; \eta_j \int (t - x_{j,2})(t - x_{j,3})....(t - x_{j,j-1}).m_n(dt) \; + \; o_p(n^{-1/2})
\end{aligned}$$

$$= \int L(t).[m_n - m](dt) \; - \; \eta_2 \mathbb{E}\{h_2(X)\} - \; \eta_3 \mathbb{E}\{h_3(X)\} \; - \; ..... - \; \eta_j \mathbb{E}\{h_j(X)\} + \; o_p(n^{-1/2}) \quad .$$

By the same way,

$$\begin{aligned}
& L^n(x_1^n) \; - \; L(x_{j,1}) \\
= \; & (x_{j,1}^n - x_{j,2}^n)(x_{j,1}^n - x_{j,3}^n)...(x_{j,1}^n - x_{j,j}^n) \; - \; (x_{j,1} - x_{j,2})(x_{j,1} - x_{j,3})...(x_{j,1} - x_{j,j}) \\
= \; & (x_{j,1}^n - x_{j,2})(x_{j,1}^n - x_{j,3})...(x_{j,1}^n - x_{j,j}) \; - \; (x_{j,1} - x_{j,2})(x_{j,1} - x_{j,3})...(x_{j,1} - x_{j,j}) \\
- \; & \eta_2(x_{j,1}^n - x_{j,3})(x_{j,1}^n - x_{j,4})...(x_{j,1}^n - x_{j,j}) \\
- \; & \eta_3(x_{j,1}^n - x_{j,2})(x_{j,1}^n - x_{j,4})...(x_{j,1}^n - x_{j,j}) \\
- \; & ..... \\
- \; & \eta_j(x_{j,1}^n - x_{j,2})(x_{j,1}^n - x_{j,3})...(x_{j,1}^n - x_{j,j-1}) + \; o_p(n^{-1/2})
\end{aligned}$$

$$\begin{aligned}
= \; & (x_{j,1} - x_{j,2})(x_{j,1} - x_{j,3})...(x_{j,1} - x_{j,j}) \; - \; (x_{j,1} - x_{j,2})(x_{j,1} - x_{j,3})...(x_{j,1} - x_{j,j}) \\
+ \; & \eta_1(x_{j,1} - x_{j,3})(x_{j,1} - x_{j,4})...(x_{j,1} - x_{j,j})
\end{aligned}$$

$$
\begin{aligned}
&+ \quad \eta_1(x_{j,1}-x_{j,2})(x_{j,1}-x_{j,4})...(x_{j,1}-x_{j,j}) \\
&+ \quad ..... \\
&+ \quad \eta_1(x_{j,1}-x_{j,2})(x_{j,1}-x_{j,3})...(x_{j,1}-x_{j,j-1}) \\
&- \quad \eta_2(x_{j,1}-x_{j,3})(x_{j,1}-x_{j,4})...(x_{j,1}-x_{j,j}) \\
&- \quad \eta_3(x_{j,1}-x_{j,2})(x_{j,1}-x_{j,4})...(x_{j,1}-x_{j,j}) \\
&- \quad ..... \\
&- \quad \eta_j(x_{j,1}-x_{j,2})(x_{j,1}-x_{j,3})...(x_{j,1}-x_{j,j-1}) + \ o_p(n^{-1/2}) \\
&= \quad \eta_1\big[h_2(x_{j,1})+...+h_j(x_{j,1})\big] \ - \ \eta_2 h_2(x_{j,1}) \ - .... - \ \eta_j h_j(x_{j,1}) + \ o_p(n^{-1/2}) \ .
\end{aligned}
$$

Therefore,

$$
\lambda_{j,1}^n - \lambda_{j,1}
$$

$$
\begin{aligned}
&= \quad \frac{\int L(t).[m_n-m](dt)}{L(x_{j,1})} \\
&- \quad \eta_2 \frac{\mathbb{E}\{h_2(X)\}}{L(x_{j,1})} \ - \eta_3 \frac{\mathbb{E}\{h_3(X)\}}{L(x_{j,1})} \ - ..... \ - \eta_j \frac{\mathbb{E}\{h_j(X)\}}{L(x_{j,1})} \\
&- \quad \eta_1 \frac{\mathbb{E}\{L(X)\}}{L(x_{j,1})^2}\big[h_2(x_{j,1})+h_3(x_{j,1})+....+h_j(x_{j,1})\big] \\
&+ \quad \eta_2 h_2(x_{j,1})\frac{\mathbb{E}\{L(X)\}}{L(x_{j,1})^2} + ... + \eta_j h_j(x_{j,1})\frac{\mathbb{E}\{L(X)\}}{L(x_{j,1})^2} + o_p(n^{-1/2})
\end{aligned}
$$

$$
\begin{aligned}
&= \quad \frac{\int (t-x_{j,2})(t-x_{j,3})...(t-x_{j,j}).[m_n-m](dt)}{(x_{j,1}-x_{j,2})(x_{j,1}-x_{j,3})...(x_{j,1}-x_{j,j})} \\
&- \quad \eta_1 \frac{\mathbb{E}\{L(X)\}}{L(x_{j,1})^2}\big[h_2(x_{j,1})+h_3(x_{j,1})+....+h_j(x_{j,1})\big] \\
&+ \quad \eta_2 \frac{1}{L(x_{j,1})}\left[\frac{h_2(x_{j,1})\mathbb{E}\{L(X)\}}{L(x_{j,1})} - \mathbb{E}\{h_2(X)\}\right] \\
&+ \quad \eta_3 \frac{1}{L(x_{j,1})}\left[\frac{h_3(x_{j,1})\mathbb{E}\{L(X)\}}{L(x_{j,1})} - \mathbb{E}\{h_3(X)\}\right] \\
&+ \quad ..... \\
&+ \quad \eta_j \frac{1}{L(x_{j,1})}\left[\frac{h_j(x_{j,1})\mathbb{E}\{L(X)\}}{L(x_{j,1})} - \mathbb{E}\{h_j(X)\}\right] + o_p(n^{-1/2})
\end{aligned}
$$

$$
= \quad \int \ell_1(x).[m_n-m](dx)
$$

$$
\begin{aligned}
&- \quad \left(\int \tilde{P}_j(t)\Big[\sum_{u=0}^{j-1} P_u(x_{j,1})P_u(t)\Big].m_n(dt)\right)\frac{\mathbb{E}\{L(X)\}}{\tilde{P}_j'(x_{j,1})L(x_{j,1})^2}\big[h_2(x_{j,1})+....+h_j(x_{j,1})\big]\\
&+ \quad \left(\int \tilde{P}_j(t)\Big[\sum_{u=0}^{j-1} P_u(x_{j,2})P_u(t)\Big].m_n(dt)\right)\frac{1}{L(x_{j,1})\tilde{P}_j'(x_{j,2})}\left[\frac{h_2(x_{j,1})\mathbb{E}\{L(X)\}}{L(x_{j,1})}-\mathbb{E}\{h_2(X)\}\right]\\
&+ \quad \left(\int \tilde{P}_j(t)\Big[\sum_{u=0}^{j-1} P_u(x_{j,3})P_u(t)\Big].m_n(dt)\right)\frac{1}{L(x_{j,1})\tilde{P}_j'(x_{j,3})}\left[\frac{h_3(x_{j,1})\mathbb{E}\{L(X)\}}{L(x_{j,1})}-\mathbb{E}\{h_3(X)\}\right]\\
&+ \quad .....\\
&+ \quad \left(\int \tilde{P}_j(t)\Big[\sum_{u=0}^{j-1} P_u(x_{j,j})P_u(t)\Big].m_n(dt)\right)\frac{1}{L(x_{j,1})\tilde{P}_j'(x_{j,j})}\left[\frac{h_j(x_{j,1})\mathbb{E}\{L(X)\}}{L(x_{j,1})}-\mathbb{E}\{h_j(X)\}\right]\\
&+ \quad o_p(n^{-1/2})
\end{aligned}
$$

$$
\begin{aligned}
&= \quad \frac{\int L(t).[m_n-m](dt)}{L(x_{j,1})}\\
&- \quad \left(\int \tilde{P}_j(t)\Big[\sum_{u=0}^{j-1} P_u(x_{j,1})P_u(t)\Big].m_n(dt)\right)\frac{\mathbb{E}\{L(X)\}}{L(x_{j,1})^3}\big[h_2(x_{j,1})+....+h_j(x_{j,1})\big]\\
&+ \quad \int\Big\{\sum_{u=0}^{j-1}\Big[\sum_{r=2}^{j}\Big(\frac{h_r(x_{j,1})\mathbb{E}\{L(X)\}}{L(x_{j,1})}-\mathbb{E}\{h_r(X)\}\Big)\frac{P_u(x_{j,r})}{L(x_{j,1})\tilde{P}_j'(x_{j,r})}\Big]P_u(t)\tilde{P}_j(t)\Big\}.m_n(dt)\\
&+ \quad o_p(n^{-1/2})\ .
\end{aligned}
$$

We deduce the lemma. ■

Chapter 3

Detection of points of concentration

3.1 Introduction

The Gauss Jacobi quadrature can allow to detect the concentration points of a probability. For example let X be a random variable with a uniform density on [0,1] except at one point x_0 where it has a nonzero measure : $P\{X = x_0\} = b > 0$. Then the Gauss Jacobi quadrature will help to detect this point and its weight. This will be true for several points of nonzero measure. This is also the case for probabilities with concentrations in certain points, i.e. near certain points, the probability is very close to a Dirac measure.

3.1.1 Complement of the results of section 1.1.2

The results of section 1.1.2 show that in the case of sufficiently smooth density, the distance of successive roots $x_{j,s}$ converge to 0 and the same is true for weight $\lambda_{j,s}$. We can ask ourselves what happens when X has a density of this type except at one point x_0 where it has a nonzero measure : $P\{X = x_0\} = b > 0$. We can easily have a first answer thanks to Stieltjes Markov's inequality. But for further details, it seems difficult to get them quickly by mathematical theorems. Then the simplest is to make simulations.

That's what we did: in all simulations which we have made, we found that the properties of Section 1.1.2 remain true except near points of nonzero measure. That is to say that, generally, $x_{j,s+1} - x_{j,s} \to 0$ and $\lambda_{j,s} \to 0$ as $j \to \infty$.

But it is quite clear that it will be different for the points closest to x_0 such that $P\{X = x_0\} = b > 0$. Indeed, suppose that two consecutive roots check $x_{j,k} < x_0 < x_{j,k+1}$ with $P\{X = x_0\} = b > 0$. Then, by Stieljes Markov inequality,

$$\sum_{x_{j,s} < x_{j,k}} \lambda_{j,s} \leq F(x_{j,k} - 0) \;\; and \;\; F(x_{j,k+1} + 0) \leq \sum_{x_{j,s} \leq x_{j,k+1}} \lambda_{j,s} \; .$$

Therefore,

$$\begin{aligned} F(x_0 + 0) - F(x_0 - 0) &\leq F(x_{j,k+1} + 0) - F(x_{j,k} - 0) \\ &\leq \sum_{x_{j,s} \leq x_{j,k+1}} \lambda_{j,s} - \sum_{x_{j,s} < x_{j,k}} \lambda_{j,s} = \lambda_{j,k} + \lambda_{j,k+1} \; . \end{aligned}$$

Therefore, $\lambda_{j,k} + \lambda_{j,k+1} \geq b$. This means that there will always weights which are not too small even if j is large.

In fact, in some simulations which we have made, we found enough frequently some weights $\lambda_{j,k_j} \to b$ as $j \to \infty$. Moreover, $x_{j,k_j+1} \approx x_0$ ou $x_{j,k_j} \approx x_0$ and $x_{j,k+1} - x_{j,k} \to 0$ as $j \to \infty$.

So it seems that in many cases, in order to find the non-zero measure in a point and its weight, it suffices to find the roots x_{j,k_j} close to a x_0 and weights $\lambda_{j,k_j} \approx b$.

This is what will confirm the following examples.

3.1.2 Example 1

In this example we chose a probability mixing a continuous density and a point x_0 such that $P\{X = x_0\} > 0$. The continuous density is that of a Gaussian mixture centered in -3, -1, with standard deviation 1, 2. The Dirac measure is located in $x_0 = 4$. The weights of the two Gaussian density and of the Dirac measure (in -3, -1, 4) are the weights 2/9, 4/9, 1/3.

We have calculated the $x^n_{j,t}$'s and the $\lambda^n_{j,t}$'s empirically, i.e. we used a sample of this mixture Gaussian. The histogram of the empirical density is represented in Figure 3.1. Note that in this one, it is simultaneously represented the curve of the density of mixture Gaussian and the non-zero measure in $x_0 = 4$. In order that graphs are consistent, it would be required that the Dirac measure is infinite. It's impossible. Also, one has been obliged to adopt the following convention.

Convention 3.1.1 *In Figure 3.1, the abscissae and ordinates do not represent the actual values of the empirical density nor the probability of the value in $x_0 = 4$: it is the only to give an idea of the distribution of probability.*

In this report, we'll keep the same convention for representing any figure which has both a density and a measure $m(\{x_0\}) > 0$.

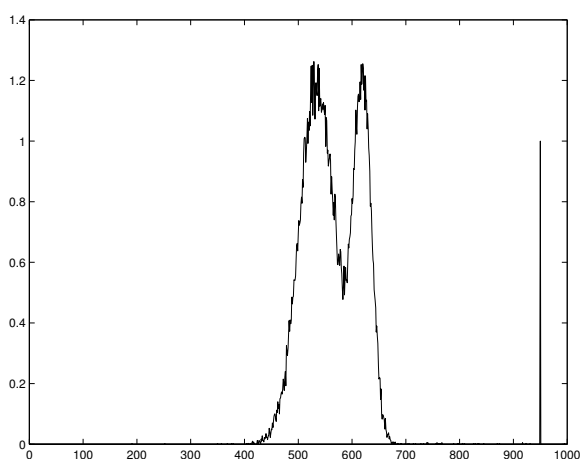

Figure 3.1: Gaussian mixture in -4,-3 and Dirac measure in 4

Variances We study first the variances of higher order. We find the following table.

j=	1	2	3	4
σ_j^2	11.0370	63.0308	453.1219	2354.33

As a matter of fact, one sees that these variances tell us nothing about the concentration in 4. Also, in order to compare, we also calculated the variances of the Gaussian mixture alone (without the point of concentration 4). We find the following table.

j=	1	2	3	4
σ_j^2	3.2234	24.5524	290.0836	1208.25

This shows that the study of variances does not tell us much about point of concentration. Even the variance of order 2 of the Gaussian mixture does not clearly indicate the concentration around two tops, which should be the case for a Gaussian mixture with two components.

Note that the variances are even smaller when we removed the point of concentration. In fact, this is normal because the probability will stop to be close to 0 in 4 and moments will be bigger.

So generally, we can not use the variances of higher order to teach us the existence of a measure $m(\{x_0\}) > 0$ mixed with a probability which has a continuous density. But we will see that this is not the case with the parameters of the Gauss Jacobi quadrature.

Gauss Jacobi Quadrature Indeed, the points and weights of Gauss Jacobi quadrature allows to detect the concentration in the point $x_0 = 4$. Indeed, the parameters of the Gauss Jacobi quadrature are given by the following tables.

	1	2	3	4	5	6	7	8
x_{1s}	-0.2274	0	0	0	0	0	0	0
x_{2s}	3.4109	-3.2730	0	0	0	0	0	0
x_{3s}	3.9011	-4.8453	-1.7292	0	0	0	0	0
x_{4s}	-6.3918	3.9776	-3.3357	-0.4282	0	0	0	0
x_{5s}	-7.7252	3.9960	-4.5646	-2.2141	0.6129	0	0	0
x_{6s}	-8.7553	-5.6417	4.0051	-3.2739	1.7758	-0.9448	0	0
x_{7s}	-9.5629	-6.5870	4.0981	3.8168	-4.1424	-2.0990	0.3619	0
x_{8s}	-10.2119	-7.4893	5.7010	3.9945	-5.0952	-3.1244	1.1844	-1.1042
x_{9s}	-10.6159	-8.2123	6.2676	-5.9674	3.9995	-3.9839	1.8163	-2.2345
x_{10s}	-10.8087	-8.6956	6.4217	-6.6345	-4.7074	4.0022	-3.0406	2.5305
x_{11s}	-10.8938	-9.0235	6.4882	-7.1783	-5.3689	4.0106	3.6130	-3.7158
x_{12s}	-10.9303	-9.2567	6.5310	-7.6459	-5.9684	4.6708	3.9965	-4.3445

	9	10	11	12
x_{9s}	-0.2326	0	0	0
x_{10s}	-1.3096	0.6104	0	0
x_{11s}	-2.1612	1.4062	-0.4045	0
x_{12s}	1.9751	-2.8888	-1.3550	0.3182

	1	2	3	4	5	6	7	8
$\lambda_{1,s}$	1.0000	0	0	0	0	0	0	0
$\lambda_{2,s}$	0.4557	0.5443	0	0	0	0	0	0
$\lambda_{3,s}$	0.3658	0.1791	0.4551	0	0	0	0	0
$\lambda_{4,s}$	0.0317	0.3454	0.3893	0.2336	0	0	0	0
$\lambda_{6,s}$	0.0052	0.3383	0.1696	0.3761	0.1108	0	0	0
$\lambda_{6,s}$	0.0011	0.0557	0.3333	0.3348	0.0431	0.2320	0	0
$\lambda_{7,s}$	0.0003	0.0176	0.1909	0.1529	0.2029	0.3218	0.1137	0
$\lambda_{8,s}$	0.0001	0.0050	0.0002	0.3375	0.0821	0.3057	0.0579	0.2116
$\lambda_{9,s}$	0.0000	0.0016	0.0000	0.0296	0.3355	0.1943	0.0298	0.2817
$\lambda_{10,s}$	0.0000	0.0007	0.0000	0.0123	0.1010	0.3337	0.2708	0.0138
$\lambda_{11,s}$	0.0000	0.0004	0.0000	0.0055	0.0484	0.3210	0.0175	0.2019
$\lambda_{12,s}$	0.0000	0.0002	0.0000	0.0026	0.0232	0.0008	0.3351	0.1223

	9	10	11	12
$\lambda_{9,s}$	0.1273	0	0	0
$\lambda_{10,s}$	0.1954	0.0723	0	0
$\lambda_{11,s}$	0.2453	0.0383	0.1217	0
$\lambda_{12,s}$	0.0208	0.2452	0.1740	0.0756

This shows that the measure nonzero in $x_0 = 4$ with a weight 1/3 implies roots close to 4 and weights close to 0,333 while the other roots and weights are much more varied.

Note that some weights are close to 0. This is due to properties of the pure Gaussian (without Dirac measure) or to that of Gaussian mixtures: compare with the examples in the following sections 3.1.3 , 3.1.4, 3.1.5 , 3.1.6 (N(0,1) distribution) and 3.1.7, 3.1.8, 3.1.9 (mixture of uniform distributions).

In order to better appreciate these results, in the following table, we shall group together the roots close to 4 and the associated weight by using the following convention.

Convention 3.1.2 *In the tables giving the two roots around points x_s with measures $m(\{x_s\}) > 0$, we are only interested by the roots with significant weight. When there is an alone significant weight we put a 0 for the following weight and we did the same thing for root.*

There are then for the roots close to x=4 and the associated weights the following table.

j	$x_{j,k}$	$x_{j,k+1}$	$\lambda_{j,k}$	$\lambda_{j,k+1}$
3	3.9011	0	0.3658	0
4	3.9776	0	0.3454	0
5	3.9960	0	0.3383	0
6	4.0051	0	0.3333	0
7	4.0981	3.8168	0.1909	0.1529
8	3.9945	0	0.3375	0
9	3.9995	0	0.3355	0
10	4.0022	0	0.3337	0
11	4.0106	0	0.3210	0
12	3.9965	0	0.3351	0

By Stieljes Markov's inequality, we know that the weight at x = 4, checks $\lambda_{j,k} + \lambda_{j,k+1} \geq P\{X = 4\}$ when $x_{j,k} < 4 < x_{j,k+1}$.

So we see that there is a weight of about 0,333 concentrated around 4: indeed, $P\{X \in [4 - a, 4 + a]\} \leq \lambda_{j,k} + \lambda_{j,k+1}$ where a is small enough .

Remark 3.1.3 *One can find practical problems for calculating the elements of the Gauss Jacobi quadrature. Indeed, it is known that the computation of orthogonal polynomials P_j can be difficult when j increases a little. This is because the moments become very large or very small depending on the case. The accuracy of calculations becomes delicate and one can sometimes find values enough distant of real values. A fortiori, it is even more true for the Gauss Jacobi quadrature.*

3.1.3 Example 2 : Gaussian standard case

We study now the standard Gaussian case : without measure concentrated at a point. We will see that for the first roots, it appears significant weight at x = 0. In order to make it to vanish, we have to choose root and weight of large enough order.

Therefore around x = 0, we find the roots and weight following.

j	$x_{j,k}$	$x_{j,k+1}$	$\lambda_{j,k}$	$\lambda_{j,k+1}$
3	0.0075	0	0.6709	0
4	0.7816	-0.6968	0.4410	0.4671
5	0.1057	0	0.5309	0
6	0.7879	-0.4427	0.3657	0.4449
7	0.1651	0	0.4348	0
8	0.6528	-0.3623	0.3280	0.3803
9	0.1216	0	0.3645	0
10	0.5370	-0.3207	0.2945	0.3265
11	0.0905	0	0.3148	0
12	0.4472	-0.2924	0.2643	0.2844

This shows that, at the beginning, the weight around 0 decreases slowly although there have no points of nonzero measure.

However the roots are well distributed, but associated weights may be almost zero. For example for j = 12, we have the following table.

$x_{12,s}$	4.3503	3.1757	-3.5296	-3.0934	2.6400	-2.4636	1.9207	-1.7602	1.1829	-1.0183	0.4472	-0.2924
$\lambda_{12,s}$	0.0001	0.0013	0.0003	0.0022	0.0095	0.0132	0.0451	0.0596	0.1479	0.1722	0.2643	0.2844

3.1.4 Example 3

One takes a standard normal curve with a point of concentration 0 and weight 1/4 for N(0.1) and 3/4 for the point of concentration with the nonzero measure at x = 0 : $m(\{0\}) = 3/4$.

Because the results of the standard Gaussian case, it is not surprising that the weight close to 0 does not tend fast enough to the weight of the point of concentration.

Note that the roots $x_{j,k}$ where j is odd have a single point with a strong concentration and that the roots of even order have two points around 0. This result is normal: it is due to the symmetry with respect to 0.

j	$x_{j,k}$	$x_{j,k+1}$	$\lambda_{j,k}$	$\lambda_{j,k+1}$
3	0.0060	0	0.9172	0
4	0.3908	-0.2998	0.4297	0.5438
5	0.0003	0	0.8873	0
6	-0.2835	0.2572	0.4546	0.4960
7	0.0005	0	0.8649	0
8	0.2334	-0.1870	0.4169	0.5079
9	0.0033	0	0.8472	0
10	0.2094	-0.1355	0.3655	0.5388
11	0.0039	0	0.8335	0
12	0.2048	-0.0921	0.2920	0.5946

Regarding the other roots and weights, we find similar results to those of the Gaussian case: the roots are well distributed and the other weights are lower (which is normal because you have to remove the weights near x = 0) .

Finally, we see that the weight around x = 0 decreases slowly and we have to take j bigger than 12 in order that the weights associated approaches 0.75.

3.1.5 Example 4

Now, we take a weight for the Dirac measure smaller than in the previous example. One takes a mixture of N (0,1) and of a point of concentration in x = 0 with weight 1/2 for N (0,1) and 1/2 the point of concentration : $m(\{0\}) = 1/2$.

j	$x_{j,k}$	$x_{j,k+1}$	$\lambda_{j,k}$	$\lambda_{j,k+1}$
3	-0.0000	0	0.8294	0
4	0.4661	-0.4574	0.4665	0.4732
5	0.0018	0	0.7574	0
6	0.3627	-0.3392	0.4297	0.4524
7	0.0014	0	0.7126	0
8	0.2766	-0.2730	0.4126	0.41789
9	-0.0037	0	0.6783	0
10	-0.2351	0.2049	0.3739	0.4124
11	-0.0028	0	0.6523	0
12	-0.1813	0.1788	0.3752	0.3757

As previously, we see that the weight around 0 decreases slowly.

In order to find weight approaching the normal weight (0.5), we have to use orthogonal polynomials of bigger order. For example if j = 23, there is a root and a weight $\lambda_{j,k+1} = 0.5827$.

3.1.6 Example 5

In this example, we take two points of concentration in -4 and 4 and a Gaussian mixture centered -1,0 with standard deviation 1, 2. The respective weights in x=-4,4,-1,0 are 0.2, 0.4, 0.2, 0.2 : $m(\{-4\}) = 0.2$ and $m(\{4\}) = 0.4$.

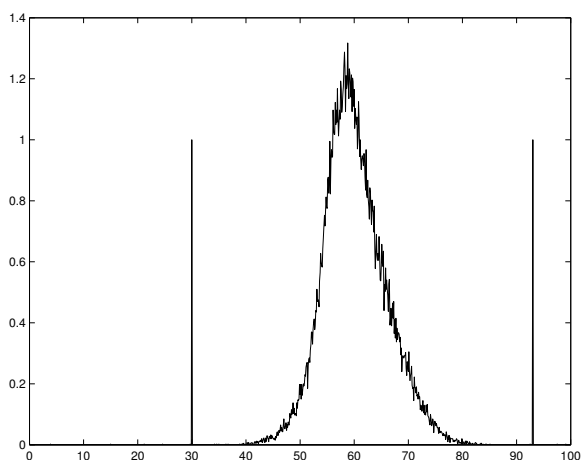

Figure 3.2: Two points of nonzero measure in 4 and -4

Then, one has the following table.

j	$x_{j,k}$	$x_{j,k'}$	$\lambda_{j,k}$	$\lambda_{j,k'}$
5	4.2142	-3.9737	0.2936	0.2150
6	-3.9986	3.5103	0.2059	0.3375
7	4.0160	-4.0258	0.3095	0.1917
8	-3.9430	4.4904	0.2079	0.1772
9	-3.9913	3.9020	0.2066	0.3032
10	-4.0001	4.1900	0.2032	0.2394
11	3.8007	-4.0082	0.2796	0.1983
12	4.0488	-3.9490	0.2668	0.1756
12	4.0051	-4.0087	0.2108	0.2087

In this table, we do not give two roots around each points of concentration, but the roots close to 4 and -4. Of course, the weight $\lambda_{j,k}$ and $\lambda_{j,k'}$ are the associated weights.

This shows that the roots approach slowly 4 and -4. Moreover the weights associated converge slowly.

3.1.7 Example 6

We choose mixtures of two uniform distributions and of two points of nonzero measures. We choose uniform mixture on the intervals [-2, -1], [3/4, 3/2] and the points -3, 2 , as having a non-zero measure. The weights associated with these intervals and in points -3, 2, of nonzero measure are respectively 0.2, 0.4, 0.2, 0.2.

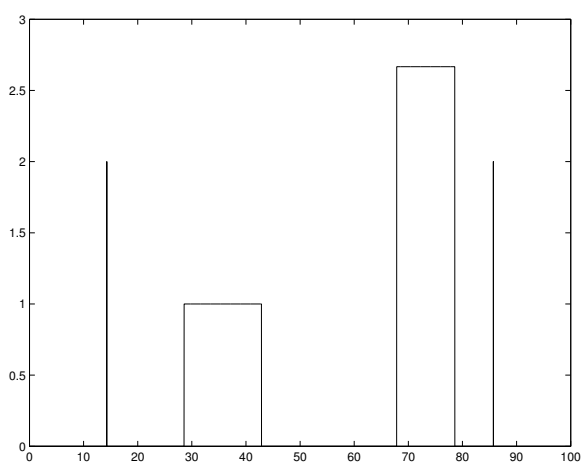

Figure 3.3: Two points with a nonzero measure

j	$x_{j,k}$	$x_{j,k'}$	$\lambda_{j,k}$	$\lambda_{j,k'}$
3	-2.8337	1.6785	0.2724	1.6785
4	-2.9880	1.9522	0.2072	1.9522
5	-2.9973	1.9812	0.2021	1.9812
6	-2.9998	-1.8015	0.2002	1.9980
7	-3.0000	1.9993	0.2000	1.9993
8	-3.0000	1.9999	0.2000	1.9999
9	-3.0000	2.0000	0.2000	0.2000
10	-3.0000	2.0000	0.2000	0.2000
11	-3.0000	2.0000	0.2000	0.2000
12	-3.0000	2.0000	0.2000	0.2000

We note that the concentration points (-3,2) appear much more distinctly than in the previous example.

We note also that the roots are almost all concentrated on the intervals of uniform mixing, i.e. on the interval [-2, -1] and [3/4, 3/2] : for example the roots of P_{12} are : -3.0000, 2.0000 and -1.9621, -1.8077, -1.5662, -1.291, -1.0662 and 1.4613, 1.3114, 1.0914, 0.8704, 0.7396,

This specifies proposition 1.1.9 according to which there may not be three successive roots $x_{j,k} < x_{j,k+1} < x_{j,k+2}$ such that $P\{X \in [x_{j,k}, x_{j,k+2}]\} = 0$ if $\lambda_{j,k+1} > 0$.

This shows that the points of concentration and weight appear clearly : they are the roots and weights associated close to points with a nonzero measure.

3.1.8 Example 7

One chooses two uniform mixtures on [-2, -1] and [3/4, 3/2] and a non-zero measure in x=0 with respective weights 0.4, 0.4, 0.2,

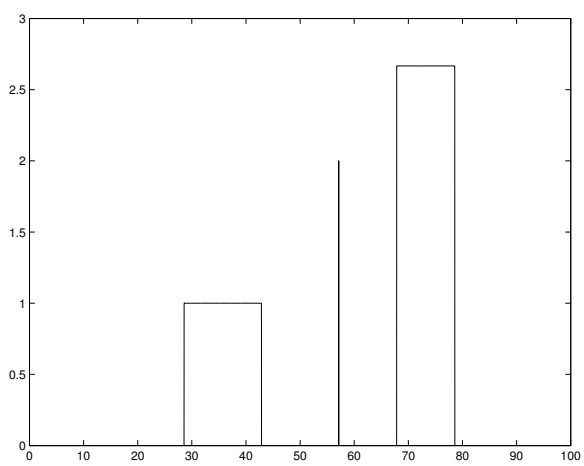

Figure 3.4: One point with a nonzero measure

j	$x_{j,k}$	$\lambda_{j,k}$
5	-0.0520	0.2055
6	0.0560	0.2048
7	-0.0103	0.1991
8	0.0050	0.2009
9	-0.0018	0.1996
10	0.0004	0.2001
11	-0.0003	0.1999
12	0.0000	2.0000

In this example, the points of concentration and weight appear clearly: they are the roots near x=0 and the weights associated from a certain j.

Moreover as in the previous example, the roots are almost all concentrated on the intervals of the uniform mixture, i.e. on the intervals [-2, -1] and [3/4, 3/2]: for example the roots of P_{12} are : 0.0000 and -1.9742, -1.8658, -1.6833, -1.4546, -1.2165, -1.0445 and 1.4754, 1.3722, 1.2014, 0.9927, 0.8071 .

3.1.9 Example 8

We choose uniform distributions on connected intervals and a point of nonzero measure located inside of their supports. Then we take an uniform mixture of the intervals [-3.0], [0,0.750] and a non-zero measure in -1 with weights, respectively 0.3, 0.5, 0.2.

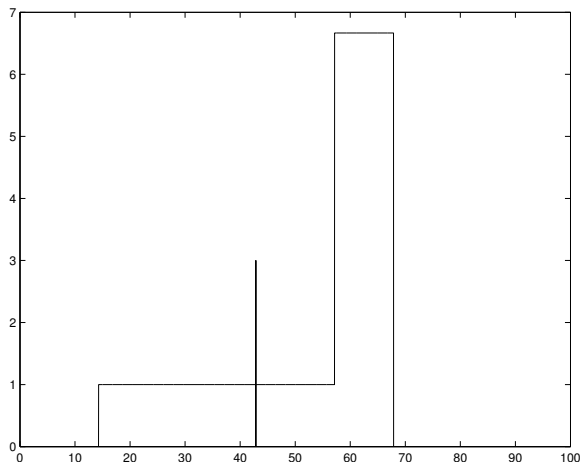

Figure 3.5: One point with a nonzero measure

Roots close to -1 and the associated weights are given in the following table.

j	$x_{j,k}$	$x_{j,k+1}$	$\lambda_{j,k}$	$\lambda_{j,k+1}$
5	-1.0055	0	0.3157	0
6	-1.3163	-0.7185	0.1797	0.2035
7	-1.0081	0	0.2852	0
8	-1.2715	-0.8540	0.1320	0.2036
9	-1.0145	0	0.2650	0
10	-1.2385	-0.9095	0.2067	0.0755
11	-1.0156	0	0.2521	0
12	-1.2521	-0.9544	0.0695	0.2228
13	-1.0389	0	0.2234	0
14	-1.0308	0	0.2310	0
15	-0.9705	0	0.2273	0
16	-1.0026	-0	0.2422	0

The points of concentration and weights appear clearly: they are the roots close to x=0 and the weights associated from a certain order.

3.1.10 Conclusion

The Gauss Jacobi quadrature can effectively detect the points with a non-zero measure. It suffices in many cases to see near which points, the weights $\lambda_{j,s}$ are greater than a certain limit. But because we can calculate the orthogonal polynomials only until a finite order, it may also mean that the probability is concentrated near a given point, but in a measure weaker than the Dirac measure.

Chapter 4

Application : mixtures

We will apply the previous results to find the different components of a Gaussian mixture. In order to do this, at first we need some properties.

4.1 Some properties

At first we introduce the type of mixtures that we will study.

Notations 4.1.1 *Let Y be a real random variable. We assume that there exists $p \in \mathbb{N}$ such that $Y = \sum_{r=1}^{p} \delta_r Z_r$, where $Z_r - z_r = \sigma_r \epsilon_r \sim N(0, \sigma_r^2)$, $z_r \in \mathbb{R}$, where the δ_r are random variables such that $\delta_r = 0$ or 1 , $\sum_{r=1}^{p} \delta_r = 1$, and where $(\delta_1, \delta_2, ..., \delta_p), \epsilon_1, \epsilon_2, ..., \epsilon_p$ are independent.*

We assume that the z_r's , $r = 1, 2, .., p$, are distincts. We set $P(\delta_r = 1) = q_r$ for $r = 1, 2, ..., p$.

Because $Z_r = z_r + \sigma_r \epsilon_r$, then $Y = \sum_{r=1}^{p} \delta_r (z_r + \sigma_r \epsilon_r) = \sum_{r=1}^{p} \delta_r z_r + \sum_{r=1}^{p} \delta_r \sigma_r \epsilon_r$.

One want to estimate the z_r 's and q_r's by using Gauss Jacobi Quadrature. Unfortunately, the Gauss-Jacobi quadrature of Y does not give the z_r 's and q_r's directly. Indeed, let us denote by $y_{p,s}$ and $\lambda'_{p,s}$, the zeros and the weights of orthogonal polynomials associated to Y. Then, generally, $y_{p,s} \neq z_s$ and $\lambda'_{p,s} \neq q_s$.

It is true only when $\sigma_p(Y) = 0$.

Proposition 4.1.1 *We denote by $\{\tilde{Q}_j\}$ the familly of orthogonal polynomials associated with $\sum_{r=1}^{p} \delta_r z_r$: $\tilde{Q}_j(z) = \sum_{t=0}^{j} b_{j,t} z^t$ where $b_{j,j} = 1$. Then, the z_r's are the zeros $z_{p,r}$ of $\tilde{Q}_p$. Moreover, let $\ell_r^S(z) = \frac{\tilde{Q}_p(z)}{(z - z_r)\tilde{Q}'_p(z_r)}$. Then, $q_r = \int \ell_r^S(z).m_S(dz)$ for $r = 1, 2...., p$ where m_S is the distribution of $S = \sum_{r=1}^{p} \delta_r z_r$.*

Therefore, we have to use the Gauss Jacobi Quadrature not of Y but of $S = \sum_{r=1}^{p} \delta_r z_r$. Therefore, we need to know moments of $\sum_{r=1}^{p} \delta_r z_r$. In this aim, one has the idea to use "negative variance" of Gaussian (cf Section V-2, page IV-28, [5]) .

Indeed, recall that this naming is based on the following property(cf lemma 1.1, (viii) [8])

Proposition 4.1.2 *Soit $G \sim N(0, \sigma^2)$, $U \sim N(0, 1)$ and $\epsilon \sim N(0, 1)$ three independent random variable. Then, for all $q \in \mathbb{N}$,*

$$\mathbb{E}\Big\{ \big[G + iU\big]^q \Big\} = \mathbb{E}\Big\{ \big[\sqrt{\sigma^2 - 1}\ \epsilon\big]^q \Big\}$$

Proof Let $h_j(x) = x^j + \sum_{s=0}^{j-1} g_{j,s} x^j$ be the j-eme Hermite orthogonal polynomial. Then, one know that $h_j(x) = \mathbb{E}\big\{(iU + x)^j\big\}$ (cf Théorème 1-4, page IV-7, [5], lemma 1.1, (ii) [8], [13], [12]).

Then, by using this property, one prove this proposition (cf [5], [8]). ■.

Then, iU+G behaves like a Gaussian distribution with variance negative if $\sigma^2 < 1$. Now, if $\sigma^2 > 1$, we find a usual Gaussian distribution. If $\sigma^2 = 1$, the probability is concentrated in 0.

To use the negative variances in order to solve our problem, we will first prove the following property which is a consequence of proposition 4.1.2.

Proposition 4.1.3 *Let $U \sim N(0,1)$ be a random variable independent with $(\delta_1, \delta_2, ..., \delta_p)$ and $(\epsilon_1, \epsilon_2, ..., \epsilon_p)$. Then, for all $q \in \mathbb{N}$,*

$$\mathbb{E}\Big\{\Big(\sum_{t=1}^{p}(\delta_t z_t + \sigma_t \epsilon_t) + isU\Big)^q\Big\} = \mathbb{E}\big\{(T_s)^q\big\} \ ,$$

where $T_s = \sum_{t=1}^{p} \delta_t (z_t + \sqrt{\sigma_t^2 - s^2}\epsilon_t)$.

Proof We have

$$\mathbb{E}\Big\{\Big(\sum_{s=1}^{p}\delta_s(z_s + \sigma_s\epsilon_s)\Big)^q\Big\}$$

$$= \mathbb{E}\Big\{\sum_{j_1+....+j_p=q}\frac{q!}{j_1!.....j_p!}\Big(\delta_1^{j_1}(z_1+\sigma_1\epsilon_1)^{j_1}\delta_2^{j_2}(z_2+\sigma_2\epsilon_2)^{j_2}.....\delta_p^{j_p}(z_p+\sigma_p\epsilon_p)^{j_p}\Big)\Big\}$$

$$= \mathbb{E}\Big\{\frac{q!}{q!}\Big(\delta_1^q(z_1+\sigma_1\epsilon_1)^q + \delta_2^q(z_2+\sigma_2\epsilon_2)^q + ... + \delta_p^q(z_p+\sigma_p\epsilon_p)^q\Big)\Big\}$$

$$= \mathbb{E}\Big\{\Big(\delta_1^q(z_1+\sigma_1\epsilon_1)^q + \delta_2^q(z_2+\sigma_2\epsilon_2)^q + ... + \delta_p^q(z_p+\sigma_p\epsilon_p)^q\Big)\Big\}$$

$$= \mathbb{E}\Big\{\delta_1^q(z_1+\sigma_1\epsilon_1)^q\Big\} + \mathbb{E}\Big\{\delta_2^q(z_2+\sigma_2\epsilon_2)^q\Big\} + ... + \mathbb{E}\Big\{\delta_p^q(z_p+\sigma_p\epsilon_p)^q\Big\}$$

$$= \mathbb{E}\Big\{\delta_1(z_1+\sigma_1\epsilon_1)^q\Big\} + \mathbb{E}\Big\{\delta_2(z_2+\sigma_2\epsilon_2)^q\Big\} + ... + \mathbb{E}\Big\{\delta_p(z_p+\sigma_p\epsilon_p)^q\Big\}$$

$$= q_1\mathbb{E}\Big\{(z_1+\sigma_1\epsilon_1)^q\Big\} + q_2\mathbb{E}\Big\{(z_2+\sigma_2\epsilon_2)^q\Big\} + ... + q_p\mathbb{E}\Big\{(z_p+\sigma_p\epsilon_p)^q\Big\}$$

where $q_s = P\{\delta_s = 1\}$.

Therefore,

$$\mathbb{E}\Big\{\Big(\big(\sum_{s=1}^{p}\delta_s(z_s+\sigma_s\epsilon_s)\big) + isU\Big)^q\Big\}$$

$$= \mathbb{E}\Big\{\sum_{s=0}^{q}C_q^s\big(\sum_{s=1}^{p}\delta_s(z_s+\sigma_s\epsilon_s)\big)^s(isU)^{q-s}\Big\}$$

$$= \sum_{s=0}^{q}C_q^s\mathbb{E}\Big\{\big(\sum_{s=1}^{p}\delta_s(z_s+\sigma_s\epsilon_s)\big)^s\Big\}\mathbb{E}\Big\{(isU)^{q-s}\Big\}$$

$$= \sum_{s=0}^{q}C_q^s\Big[q_1\mathbb{E}\Big\{(z_1+\sigma_1\epsilon_1)^s\Big\} + q_2\mathbb{E}\Big\{(z_2+\sigma_2\epsilon_2)^s\Big\} + ... + q_p\mathbb{E}\Big\{(z_p+\sigma_p\epsilon_p)^s\Big\}\Big]\mathbb{E}\Big\{(isU)^{q-s}\Big\}$$

$$= \sum_{s=0}^{q}C_q^s q_1\mathbb{E}\Big\{(z_1+\sigma_1\epsilon_1)^s\Big\}\mathbb{E}\Big\{(isU)^{q-s}\Big\} + ... + \sum_{s=0}^{q}C_q^s q_p\mathbb{E}\Big\{(z_p+\sigma_p\epsilon_p)^s\Big\}\mathbb{E}\Big\{(isU)^{q-s}\Big\}$$

$$= q_1\mathbb{E}\Big\{\big[(z_1+\sigma_1\epsilon_1)+(isU)\big]^q\Big\} + ... + q_p\mathbb{E}\Big\{\big[(z_p+\sigma_p\epsilon_p)+(isU)\big]^q\Big\}$$

$$= q_1\mathbb{E}\Big\{\big[(z_1+\sqrt{\sigma_1^2-s^2}\ \epsilon)\big]^q\Big\} + ... + q_p\mathbb{E}\Big\{\big[(z_p+\sqrt{\sigma_p^2-s^2}\ \epsilon)\big]^q\Big\}$$

where $\epsilon \sim N(0,1)$.

Indeed,

$$\mathbb{E}\Big\{\big[\sigma_2\epsilon_2+(isU)\big]^q\Big\} = \mathbb{E}\Big\{s^q\big[(\sigma_2/s)\epsilon_2+iU\big]^q\Big\} = s^q\mathbb{E}\Big\{\big[(\sigma_2/s)\epsilon_2+iU\big]^q\Big\}$$

$$= s^q\mathbb{E}\Big\{\big[\sqrt{(\sigma_2/s)^2-1}\ \epsilon\big]^q\Big\}$$

(by proposition 4.1.2)

$$= \mathbb{E}\Big\{s^q\big[\sqrt{(\sigma_2/s)^2-1}\ \epsilon\big]^q\Big\} = \mathbb{E}\Big\{\big[s\sqrt{(\sigma_2/s)^2-1}\ \epsilon\big]^q\Big\} = \mathbb{E}\Big\{\big[\sqrt{\sigma_2^2-s^2}\ \epsilon\big]^q\Big\}\ .$$

Therefore,

$$\mathbb{E}\Big\{\big[z_1+\sigma_1\epsilon_1+(isU)\big]^q\Big\} = \mathbb{E}\Big\{\sum_{s=0}^{q} C_q^s z_1^s\big[\sigma_1\epsilon_1+(isU)\big]^{q-s}\Big\}$$

$$= \sum_{s=0}^{q} C_q^s z_1^s\mathbb{E}\Big\{\big[\sigma_1\epsilon_1+(isU)\big]^{q-s}\Big\} = \sum_{s=0}^{q} C_q^s z_1^s\ \mathbb{E}\Big\{\big[\sqrt{\sigma_1^2-s^2}\ \epsilon)\big]^{q-s}\Big\}$$

$$= \mathbb{E}\Big\{\sum_{s=0}^{q} C_q^s z_1^s\ \big[\sqrt{\sigma_1^2-s^2}\ \epsilon)\big]^{q-s}\Big\} = \mathbb{E}\Big\{\big[(z_1+\sqrt{\sigma_1^2-s^2}\ \epsilon)\big]^q\Big\}\ .$$

Now we have seen above that $\mathbb{E}\Big\{\Big(\sum_{s=1}^{p}\delta_s(z_s+\sigma_s\epsilon_s)\Big)^q\Big\} = \sum_{t=1}^{p} q_t\mathbb{E}\Big\{(z_t+\sigma_t\epsilon_t)^q\Big\}$.
Then, replace σ_t by $\sqrt{\sigma_t^2-s^2}$. Then,

$$\mathbb{E}\Big\{\Big(\sum_{t=1}^{p}\delta_t(z_t+\sqrt{\sigma_t^2-s^2}\epsilon_s)\Big)^q\Big\} = \sum_{t=1}^{p} q_t\mathbb{E}\Big\{(z_t+\sqrt{\sigma_t^2-s^2}\epsilon_t)^q\Big\}\ . \blacksquare$$

Then, we can use the previous property in order to find the various parameters σ, z_r and q_r. For example, if $\sigma_r = \sigma$ for r=1,2,....,p, $\mathbb{E}\{[Y+i\sigma U]^q\} = \mathbb{E}\{S^q\}$ (if not, if $s = min_t\sigma_t$, T_s have the same moments as a degenerate Gaussian mixture, i.e. with a standard deviation of one component which is equal to zero). Therefore if we choose $s = \sigma$, $Y + isU$ has the same moments as $S = \sum_{r=1}^{p}\delta_r z_r$. Then its variance of order p will be null. Therefore, we can find $s = \sigma$ by studying the variances of order p of isU + Y. When σ is obtained, it suffices to calculate the parameters of the Gauss Jacobi quadrature which will be the associated the z_r's and q_r's.

If all the σ_t's are not identical, we will apply the results of Chapter 3 to Y+isU.

Now, in order to make our calculations, we need also the following properties.

Proposition 4.1.4 *Let Z_s, s=1,2,...,p, be some p random variables independent of $(\delta_1, ..., \delta_p)$. Then, the random variable $\delta_1 Z_1 + ... + \delta_p Z_p$ has the density $g = q_1 f_{Z_1} + ... + q_p f_{Z_p}$ where the f_{Z_t}'s are the densities of the Z_t's.*

Proof We know that

$$\mathbb{E}\big\{e^{it[\delta_1 Z_1 + ... + \delta_p Z_p]}\big\} = \mathbb{E}\big\{e^{it\delta_1 Z_1} .....e^{it\delta_p Z_p}\big\}$$

$$= \mathbb{E}_{\delta_1=1}\big\{e^{it\delta_1 Z_1} .....e^{it\delta_p Z_p}\big\} + \mathbb{E}_{\delta_2=1}\big\{e^{it\delta_1 Z_1} .....e^{it\delta_p Z_p}\big\} + ... + \mathbb{E}_{\delta_p=1}\big\{e^{it\delta_1 Z_1} .....e^{it\delta_p Z_p}\big\}$$

$$= \mathbb{E}_{\delta_1=1}\big\{e^{it\delta_1 Z_1}\big\} + \mathbb{E}_{\delta_2=1}\big\{e^{it\delta_2 Z_2}\big\} + ... + \mathbb{E}_{\delta_p=1}\big\{e^{it\delta_p Z_p}\big\}$$

$$= \mathbb{E}_{\delta_1=1}\big\{e^{itZ_1}\big\} + \mathbb{E}_{\delta_2=1}\big\{e^{itZ_2}\big\} + ... + \mathbb{E}_{\delta_p=1}\big\{e^{itZ_p}\big\}$$

$$= \mathbb{E}\big\{\mathbb{1}_1(\delta_1)e^{itZ_1}\big\} + \mathbb{E}\big\{\mathbb{1}_1(\delta_2)e^{itZ_2}\big\} + ... + \mathbb{E}\big\{\mathbb{1}_1(\delta_p)e^{itZ_p}\big\}$$

$$= \mathbb{E}\big\{\mathbb{1}_1(\delta_1)\big\}\mathbb{E}\big\{e^{itZ_1}\big\} + \mathbb{E}\big\{\mathbb{1}_1(\delta_2)\big\}\mathbb{E}\big\{e^{itZ_2}\big\} + ... + \mathbb{E}\big\{\mathbb{1}_1(\delta_p)\big\}\mathbb{E}\big\{e^{itZ_p}\big\}$$

$$= P\{\delta_1 = 1\}\mathbb{E}\big\{e^{itZ_1}\big\} + P\{\delta_2 = 1\}\mathbb{E}\big\{e^{itZ_2}\big\} + ... + P\{\delta_p = 1\}\mathbb{E}\big\{e^{itZ_p}\big\}$$

$$= q_1\mathbb{E}\big\{e^{itZ_1}\big\} + q_2\mathbb{E}\big\{e^{itZ_2}\big\} + ... + q_p\mathbb{E}\big\{e^{itZ_p}\big\}$$

$$= \mathbb{E}\big\{e^{itx}[q_1 f_{Z_1}(x) + ... + q_p f_{Z_p}(x)]dx\big\} \ . \ \blacksquare$$

Proposition 4.1.5 *If the Z_i's are Gaussian, the random variable $\delta_1 Z_1 + ... + \delta_p Z_p$ is completely determined by its moments.*

Proof Indeed, $\mathbb{E}\big\{e^{it[\delta_1 Z_1 + ... + \delta_p Z_p]}\big\} = q_1\mathbb{E}\big\{e^{itZ_1}\big\} + q_2\mathbb{E}\big\{e^{itZ_2}\big\} + ... + q_p\mathbb{E}\big\{e^{itZ_p}\big\}$ where the Z_i's are Gaussian. Therefore we can write each $\mathbb{E}\big\{e^{itZ_p}\big\}$ as a convergent serie which is function of Gaussian moments. Therefore the characteristic function is completely determined by the moments. $\blacksquare$

4.2 First application to mixtures

4.2.1 Method

At first, we show how one can find the various parameters of a Gaussian mixture when $\sigma_r = \sigma$.

Notations 4.2.1 *In this section 4.2, we assume that $\sigma_r^2 = \sigma^2$ for all $r \in \{1, 2, ..., p\}$.*

Then, we can use the negative variances of Gaussian distributions in order to obtain the parameters of the mixture. Indeed, we have the following proposition.

Proposition 4.2.1 *Under the previous assumptions, $Y = \sum_{r=1}^p \delta_r z_r + \sigma\epsilon_0$ where $\epsilon_0 \sim N(0, 1)$.*

Proof We know that $Y = \sum_{r=1}^p \delta_r z_r + \big(\sum_{r=1}^p \delta_r \epsilon_r\big)\sigma$.
By proposition 4.1.4, $\sum_{r=1}^p \delta_r \epsilon_r$ has the density $q_1 f_{Z_1}(x) + ... + q_p f_{Z_p}(x) = [q_1 + ... + q_p]\frac{e^{-x^2/2}}{\sqrt{2\pi}} = \frac{e^{-x^2/2}}{\sqrt{2\pi}}$.

Then, $Y = \sum_{r=1}^p \delta_r z_r + \big(\sum_{r=1}^p \delta_r \epsilon_r\big)\sigma = \sum_{r=1}^p \delta_r z_r + \sigma\epsilon_0$ where $\epsilon_0 \sim N(0, 1)$. $\blacksquare$

One is therefore in case the variable $S = \sum_{r=1}^p \delta_r z_r$ is disturbed by a Gaussian noise $\sigma\epsilon$. We will see that it is easy to find the parameters of this mixture by using the variances of higher order.

Indeed, assume that we want to estimate the z_t's and q_t's. We can estimate the moments of S by using negative variances because

$$\mathbb{E}\Big\{\Big(\sum_{t=1}^{p}(\delta_t z_t + \sigma\epsilon_t) + isU\Big)^q\Big\} = \mathbb{E}\Big\{\Big(\sum_{t=1}^{p}\delta_t\big[z_t + \sqrt{\sigma^2 - s^2}\ \epsilon_t\big]\Big)^q\Big\}\ .$$

It is thus enough find s such that $\sigma^2 = s^2$: in this case, the moments of $\sum_{t=1}^{p}(\delta_t z_t + \sigma\epsilon_t) + isU$ have a variance of order p which is equal to zero. Indeed, it is the variance of order p of S which is concentrated in p distinct points.

Then, to find s such that $\sigma^2 = s^2$ is easy. Indeed, if we take s increasing from 0, $\sum_{t=1}^{p}(\delta_t z_t + \sigma\epsilon_t) + isU$ has the same moments as the real random variable $T_s = \sum_{t=1}^{p}(z_t + \sqrt{\sigma - s^2}\ \epsilon_t)$. We can then calculate the variance of order p of T_s. When it vanishes, we know the s is reached. Then we know the parameters searched $s = \sigma$ and also the z_r's and q_r's by using Gauss Jacobi quadrature.

Concretely, if we take s more and more large, variance of order p decreases to 0 and then becomes negative. Then, it is not difficult to find $s = \sigma$ by calculating the variance of order p for any value of s.

4.2.2 Examples

First example Suppose we have a sample y_ℓ , $\ell = 1, 2, ..., 10000$, of Y. Suppose that Y is a mixture of Gaussian components centered at [-2,0,3], with standard deviation 0.722 and with weights 1/6, 2/6, 3/6. Then, one calculate the variance of order 3 of T_s for some values of s. In order to calculate the moments of Y + isU we need a sample $y_\ell + isu_\ell$ of Y+isU. In order to do this, it is enough to add any pseudo-random sample u_ℓ of N(0,1).

In the first figure 4.1 we have the graph of the variance of order 3 when s varies from 0 to 1: we take σ_3^2 defined in SectionA.1.4 by replacing the moments M_s of X by the empirical moments of Y+isU.

The red curve represents the imaginary part. Indeed, for a sample we must take into account the imaginary part because it can be not exactly zero: we have a sample of X +isU and not the exact distributions.

Anyway, this figure shows that the variance vanishes near s = 0.72.

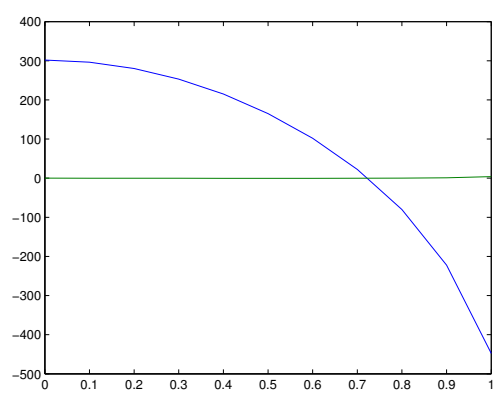

Figure 4.1: Variance of order 3 of T_s

Second Example In figure 4.2, the blue curve represents the real part of the variance of order 3 of a sample of X + isU when X is a Gaussian mixture of components centered in -2,0,3, of standard deviation 1,1,1 of weight 1/6, 2/6, 3/6 when s varies from 0 to 1. The red curve represents the imaginary part. We see that the variance vanishes near s = 1.

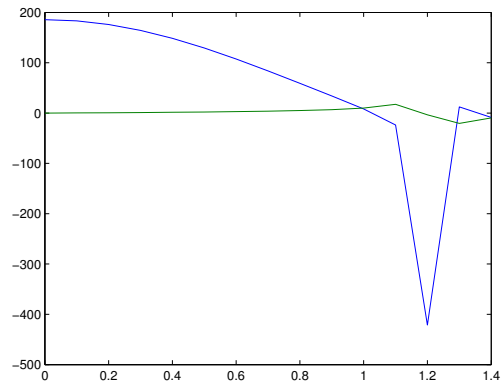

Figure 4.2: Variance of order 3 of T_s

To calculate σ, we'll take a curve defined over a smaller interval near the point where $s = \sigma$. In the figure 4.3, then we represent the same curves for intervals much smaller. This allows to conclude that $s_0 = \sigma = 0.996$.

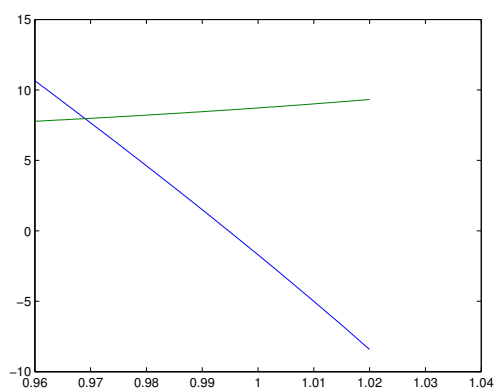

Figure 4.3: Variance of order 3 of T_s

With the moments, we can then estimate the parameters of the Gauss Jacobi quadrature.

We estimate the roots by -1.9735, 0.0653, 3.0232 (instead of -2,0,3) with weights 0.1776, 0.3274, 0.4950 (instead of 0.1667, 0.3333, 0.5000).

Because these parameters were obtained for a sample of the given law, we will compare this result with a new simulated sample. In the figure 4.4 it has been plotted in red the original curve obtained by histogram estimation (function "hist" of Matlab). In blue, we represent the curve

obtained from the simulation of a distribution of a mixture which has the estimated parameters.

Thus we understand that we can assume that the parameters obtained by this estimation are quite correct

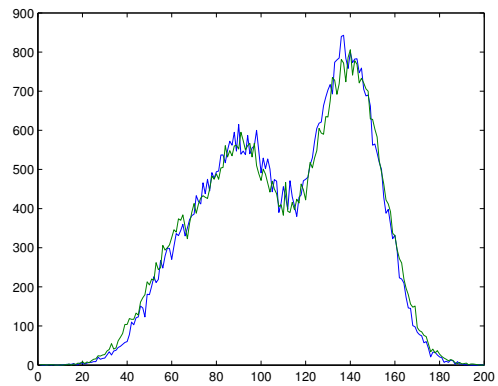

Figure 4.4: Histograms of theoretical and estimated mixtures

4.3 Second application to mixtures

4.3.1 Presentation

We have seen how to apply Proposition 4.1.3 when $Y = \sum_{r=1}^{p} \delta_r [z_r + \sigma \epsilon_r]$. Now we will apply this result when the σ_r 's are different.

Then, by proposition 4.1.3, we know that

$$\mathbb{E}\Big\{\Big(\sum_{t=1}^{p}(\delta_t z_t + \sigma\epsilon_t) + isU\Big)^q\Big\} = \mathbb{E}\Big\{\Big(\sum_{t=1}^{p}\delta_t\big[z_t + \sqrt{\sigma_t^2 - s^2}\ \epsilon_t\big]\Big)^q\Big\}\ .$$

Therefore, if $s^2 = Min_{r=1,\dots,p}(\sigma_r^2)$, $T_s = \sum_{t=1}^{p}(\delta_t z_t + \sigma_t \epsilon_t) + isU$ has the same moments as a Gaussian mixture with p components, including one degenerate, i.e. with a point of nonzero measure in z_{t_0} where $\sigma_{t_0}^2 = Min_{r=1,\dots,p}(\sigma_r^2) : P\{T_s = z_{t_0}\} = q_t > 0$. So we find s by applying a method of the same type as that described in chapter 3. Then, in order to find this point z_t, we use the properties of the Gauss Jacobi quadrature instead of those of the variances of order p.

It then suffices to apply the method described in Chapter 3 in order to find the point of concentration z_t, its weight q_t and its variance σ_t^2.

After we removed the points corresponding to the Gaussian component of the mixture and we start again.

In order to show how we had to proceed, we will study an example.

4.3.2 Example

One chooze a sample with size 100000 of a Gaussian mixture with components centered in -3,0,2, with standard deviation 1, and with weights 0.4, 0.3, 0.3. (cf figure 4.5).

4.3.3 Calculation of the first standard deviation

The first standard deviation which we can know by using the Gauss Jacobi quadrature is the smallest standard deviation. Indeed, adding $i\sigma_t U$, we find that the moments of $Y + i\sigma_t U$ are the same as those of a real random variable having a point of concentration.

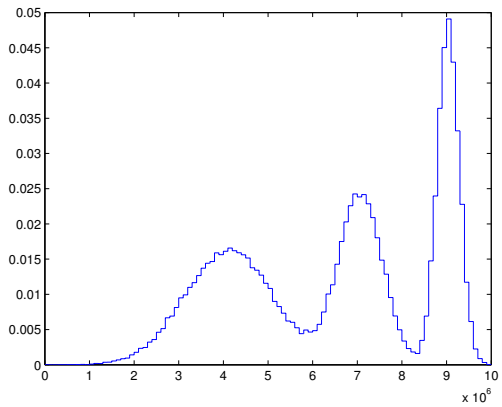

Figure 4.5: Gaussian mixture with components centered in -3,0,2

Now, for a sample, this is only estimation obtained thanks to a complex random variable. Then, it is not quite sure that all the moments are real. We must take the real part since the imaginary part of the moments tend to 0 as the sample size tends to ∞.

Therefore one estimates the moments and then the elements of the Gauss Jacobi quadrature with different s. In this report, we'll do it for s = 1/5 and s = 1/4.

Study of s=1/5 For simplicity, in this section we denote always by $\lambda_{j,s}$ and $x_{j,s}$ the weights and the roots of the Gauss Jacobi quadrature associated with the sample of the random variable Y + isU, that is to say that the theoretical probabilities have the same moments as $\sum_{r=1}^{p} \delta_r \big(z_r + \sqrt{\sigma_r^2 - s^2}\ \epsilon_r\big)$. Then, for s=1/5, we have the following tables.

	1	2	3	4	5	6	7	8
$x_{1,s}$	-0.6021	0	0	0	0	0	0	0
$x_{2,s}$	-3.1159	1.3471	0	0	0	0	0	0
$x_{3,s}$	-3.9540	1.8420	-1.1128	0	0	0	0	0
$x_{4,s}$	-4.6964	-2.9009	2.0039	-0.1215	0	0	0	0
$x_{5,s}$	-5.1886	-3.5613	2.0427	-1.6815	0.2911	0	0	0
$x_{6,s}$	-5.6402	-4.1587	-2.6969	2.0721	0.9964	-0.3988	0	0
$x_{7,s}$	-6.0472	-4.6755	-3.3621	-1.9065	2.1145	1.5975	-0.0231	0
$x_{8,s}$	-6.4284	-5.1461	-3.9369	-2.6990	2.1680	1.8255	-0.9921	0.1971
$x_{9,s}$	-6.7533	-5.5361	-4.4093	-3.2956	2.2072	1.9006	-2.1171	0.4523
$x_{10,s}$	-6.9618	-5.8040	-4.7501	-3.7108	-2.6261	2.2346	1.9395	-1.2471
$x_{11,s}$	-7.0993	-6.0243	-5.0521	-4.0767	-3.0758	-2.0081	2.3790	2.0854
$x_{12,s}$	-7.1747	-6.1922	-5.2991	-4.3857	-3.4654	-2.5049	2.9729	2.1919
$x_{13,s}$	-7.1992	-6.2764	-5.4498	-4.6123	3.1542	-3.7849	-2.9287	2.2038
$x_{14,s}$	-7.1980	-6.2709	-5.4370	-4.5881	3.1361	-3.7446	-2.8661	2.2019
$x_{15,s}$	-13.2195	7.4003	-7.1993	-6.2765	-5.4493	-4.6105	3.1619	-3.7812
$x_{16,s}$	-7.2320	-7.2320	-6.3839	-5.5899	-4.8075	-4.0369	3.1403	3.1403

	9	10	11	12	13	14	15	16
$x_{9,s}$	-0.4043	0	0	0	0	0	0	0
$x_{10,s}$	0.8678	-0.0280	0	0	0	0	0	0
$x_{11,s}$	1.7481	-0.4754	0.3394	0	0	0	0	0
$x_{12,s}$	1.8889	-1.3269	0.5672	-0.1748	0	0	0	0
$x_{13,s}$	1.9050	-2.0216	0.7431	-0.7594	0.0334	0	0	0
$x_{14,s}$	1.9019	-1.9129	0.6214	-0.0497	-0.0497	-0.4318	0	0
$x_{15,s}$	-2.9226	-2.0126	2.2042	1.9056	0.7514	-0.7462	0.0413	0
$x_{16,s}$	-3.2426	-2.4225	2.2181	1.9230	-1.4819	1.0409	-0.3944	0.2726

	1	2	3	4	5	6	7	8
$\lambda_{1,s}$	1.0000	0	0	0	0	0	0	0
$\lambda_{2,s}$	0.4367	0.5633	0	0	0	0	0	0
$\lambda_{3,s}$	0.2217	0.3860	0.3922	0	0	0	0	0
$\lambda_{4,s}$	0.0681	0.2982	0.3104	0.3233	0	0	0	0
$\lambda_{5,s}$	0.0229	0.2239	0.2874	0.2073	0.2585	0	0	0
$\lambda_{6,s}$	0.0065	0.1172	0.2389	0.2632	0.1220	0.2522	0	0
$\lambda_{7,s}$	0.0017	0.0502	0.2032	0.1563	0.2128	0.1126	0.2632	0
$\lambda_{8,s}$	0.0004	0.0181	0.1253	0.2016	0.1415	0.1678	0.1262	0.2191
$\lambda_{9,s}$	0.0001	0.0064	0.0650	0.1757	0.0976	0.2040	0.1388	0.1358
$\lambda_{10,s}$	0.0000	0.0027	0.0346	0.1322	0.1724	0.0729	0.2217	0.0904
$\lambda_{11,s}$	0.0000	0.0012	0.0176	0.0886	0.1658	0.1132	0.0108	0.1950
$\lambda_{12,s}$	0.0000	0.0006	0.0093	0.0555	0.1356	0.1448	0.0000	0.1099
$\lambda_{13,s}$	0.0000	0.0004	0.0059	0.0352	0.0000	0.0994	0.1425	0.0986
$\lambda_{14,s}$	0.0000	0.0004	0.0061	0.0373	0.0000	0.1047	0.1463	0.1005
$\lambda_{15,s}$	-0.0000	-0.0000	0.0000	0.0004	0.0059	0.0354	0.0000	0.0999
$\lambda_{16,s}$	-0.0000	0.0000	0.0003	0.0038	0.0229	0.0735	-0.0000	0.0000

j	9	10	11	12	13	14	15	16
$\lambda_{9,s}$	0.1767	0	0	0	0	0	0	0
$\lambda_{10,s}$	0.0545	0.2187	0	0	0	0	0	0
$\lambda_{11,s}$	0.1016	0.1490	0.1572	0	0	0	0	0
$\lambda_{12,s}$	0.1913	0.0718	0.0929	0.1882	0	0	0	0
$\lambda_{13,s}$	0.2006	0.0955	0.0548	0.0858	0.1814	0	0	0
$\lambda_{14,s}$	0.1994	0.0918	0.0911	0.0099	0.0120	0.2003	0	0
$\lambda_{15,s}$	0.1428	0.0949	0.0982	0.2009	0.0535	0.0872	0.1808	0
$\lambda_{16,s}$	0.1267	0.1172	0.0861	0.2100	0.0572	0.0251	0.1334	0.1437

Now we recapitulate the results for roots close to 2 and the weights associated.

j	x_{j,j_0}	x_{j,j_0+1}	λ_{j,j_0}	λ_{j,j_0+1}
3	1.8420	0	0.3860	0
4	2.0039	0	0.3104	0
5	2.0427	0	0.2874	0
6	2.0721	0	0.2632	0
7	2.1145	1.5975	0.2128	0.1126
8	2.1680	1.8255	0.1415	0.1678
9	2.2072	1.9006	0.0976	0.2040
10	2.2346	1.9395	0.0729	0.2217
11	2.0854	1.7481	0.1950	0.1016
12	2.1919	1.8889	0.1099	0.1913
13	2.2038	1.9050	0.0986	0.2006
14	2.2019	1.9019	0.1005	0.1994
15	2.2042	1.9056	0.0982	0.2009
16	2.2181	1.9230	0.0861	0.2100

By Stieljes Markov's inequality, we see that there is a weight of about 0.3 concentrated around 2. Indeed, let λ_{j,j_0} and λ_{j,j_0+1} and x_{j,j_0}, x_{j,j_0+1}, $x_{j,j_0} < 2 < x_{j,j_0+1}$, be the weights and the roots associated around 2. Then, $P\{G_3 \in [2-a, 2+a]\} \leq \lambda_{j,j_0} + \lambda_{j,j_0+1}$ where G_3 represents the third Gaussian component when a is small. Now there is no perfectly correct point in order to identify it. We see only that it seems that there is a concentration close to 2.

This is normal because s is not well chosen. Indeed, at this point 2, we have a Gaussian distribution with a variance of about $1/16 - 1/25 = 0.225$. This means that the probability is mainly concentrated on an interval of length 0.2 around 2. In order to be sure there are no points of nonzero measure, we should therefore consider the P_j where j is greater than 16. But the fact that there is no concentration points x_{js} seems enough accurate in order to indicate that for s = 1/5, there is no point of nonzero measure.

Study of s=1/4 For s=1/4, we have the following tables.

j	1	2	3	4	5	6	7	8
$x_{1,s}$	-0.6021	0	0	0	0	0	0	0
$x_{2,s}$	-3.1124	1.3408	0	0	0	0	0	0
$x_{3,s}$	-3.9413	1.8229	-1.1275	0	0	0	0	0
$x_{4,s}$	-4.6980	-2.9385	1.9695	-0.1865	0	0	0	0
$x_{5,s}$	-5.2363	1.9932	-3.6727	-2.0133	0.0862	0	0	0
$x_{6,s}$	-5.6896	-4.2653	-2.8746	1.9991	-0.9214	0.2928	0	0
$x_{7,s}$	-6.1204	-4.8151	-3.5750	2.0010	-2.2863	0.5220	-0.3507	0
$x_{8,s}$	-6.6569	-5.4122	-4.2570	2.0014	-3.1082	-1.8523	0.6346	-0.1903
$x_{9,s}$	44.9122	-6.6532	-5.4092	-4.2543	2.0014	-3.1058	-1.8497	0.6351
$x_{10,s}$	-7.0150	-5.7824	-4.7133	-3.6697	-2.5791	2.0658	2.0143	-1.1356
$x_{11,s}$	-7.1355	-5.9331	1.9053	1.9053	-4.9239	-3.9265	2.0012	-2.8875
$x_{12,s}$	-7.2054	3.1253	3.1253	-6.0623	-5.1399	-4.2139	-3.2681	2.0019
$x_{13,s}$	-7.2129	-6.0833	-5.1832	-4.2808	3.5464	-3.3681	2.2172	2.2172
$x_{14,s}$	-9.4594	-7.2232	-6.1087	-5.2331	-4.3565	2.9473	2.9473	-3.4813
$x_{15,s}$	-7.1460	-7.1460	-6.2943	-5.4456	-4.6159	-3.8111	-2.9820	2.8109
$x_{16,s}$	-7.1096	-6.6096	-6.6096	-5.6843	-4.8616	-4.0829	-3.2883	2.7362

	9	10	11	12	13	14	15	16
$x_{9,s}$	-0.1896	0	0	0	0	0	0	0
$x_{10,s}$	0.9556	0.0315	0	0	0	0	0	0
$x_{11,s}$	-1.7116	0.6040	-0.1993	0	0	0	0	0
$x_{12,s}$	-2.2752	0.8774	-0.8637	0.0661	0	0	0	0
$x_{13,s}$	1.9999	-2.4191	-1.2108	0.5951	-0.1226	0	0	0
$x_{14,s}$	-2.5807	2.0062	1.5700	-1.5540	0.3870	-0.3131	0	0
$x_{15,s}$	2.8109	2.0691	1.9616	-2.1131	-0.9498	0.5362	-0.1068	0
$x_{16,s}$	2.7362	2.3039	1.9923	-2.4793	-1.5876	0.7108	-0.4781	0.1401

	1	2	3	4	5	6	7	8
$\lambda_{1,s}$	1.0000	0	0	0	0	0	0	0
$\lambda_{2,s}$	0.4363	0.5637	0	0	0	0	0	0
$\lambda_{3,s}$	0.2222	0.3899	0.3879	0	0	0	0	0
$\lambda_{4,s}$	0.0651	0.2953	0.3203	0.3193	0	0	0	0
$\lambda_{5,s}$	0.0187	0.3062	0.1997	0.2010	0.2745	0	0	0
$\lambda_{6,s}$	0.0050	0.0976	0.2359	0.3017	0.1549	0.2050	0	0
$\lambda_{7,s}$	0.0011	0.0359	0.1723	0.2999	0.1735	0.1162	0.2011	0
$\lambda_{8,s}$	0.0001	0.0085	0.0818	0.2994	0.1918	0.1192	0.0820	0.2171
$\lambda_{9,s}$	-0.0000	0.0001	0.0086	0.0821	0.2994	0.1918	0.1189	0.0819
$\lambda_{10,s}$	0.0000	0.0027	0.0363	0.1374	0.1724	-0.0565	0.3533	0.0912
$\lambda_{11,s}$	0.0000	0.0016	-0.0000	0.0000	0.0228	0.1063	0.2996	0.1762
$\lambda_{12,s}$	0.0000	0.0000	-0.0000	0.0009	0.0130	0.0711	0.1521	0.2987
$\lambda_{13,s}$	0.0000	0.0008	0.0115	0.0632	0.0000	0.1412	0.0000	-0.0001
$\lambda_{14,s}$	-0.0000	0.0000	0.0007	0.0099	0.0549	-0.0000	0.0000	0.1281
$\lambda_{15,s}$	-0.0000	0.0000	0.0004	0.0055	0.0336	0.0946	0.1395	-0.0000
$\lambda_{16,s}$	0.0000	0.0001	0.0000	0.0030	0.0204	0.0702	0.1256	0.0000

	9	10	11	12	13	14	15	16
$\lambda_{9,s}$	0.2171	0	0	0	0	0	0	0
$\lambda_{10,s}$	0.0366	0.2265	0	0	0	0	0	0
$\lambda_{11,s}$	0.0968	0.0879	0.2088	0	0	0	0	0
$\lambda_{12,s}$	0.1317	0.0364	0.0890	0.2071	0	0	0	0
$\lambda_{13,s}$	0.3007	0.1379	0.0680	0.0797	0.1970	0	0	0
$\lambda_{14,s}$	0.1409	0.2922	0.0137	0.0675	0.1309	0.1613	0	0
$\lambda_{15,s}$	0.0000	0.0942	0.2084	0.0992	0.0619	0.0848	0.1779	0
$\lambda_{16,s}$	0.0001	0.0048	0.2967	0.1195	0.0570	0.0414	0.1092	0.1521

Then, for the roots close to 2 and the associated weights we have the following table.

j	x_{j,j_0}	x_{j,j_0+1}	λ_{j,j_0}	λ_{j,j_0+1}
3	1.8229	0	0.3899	0
4	1.9695	0	0.3203	0
5	1.9932	0	0.3062	0
6	1.9991	0	0.3017	0
7	2.0010	0	0.2999	0
8	2.0014	0	0.2994	0
9	2.0014	0	0.2994	0
10	2.0658	2.0143	0.3533	0.0912
11	1.9053	2.0012	0.0000	0.2996
12	2.0019	0	0.2987	0
13	2.2172	1.9999	0.0000	0.3007
14	2.0062	0	0.2922	0
15	2.0691	1.9616	0.0942	0.2084
16	2.3039	1.9923	0.0048	0.2967

Conclusion Finally, we see little difference between the results s = 1/5 and s = 1/4 if we confine ourselves to use the Markov Stieljes inequality. In order that the result appears more clearly, we must in fact take orthogonal polynomials of degree greater than 16.

On the other hand, the number of weights close to 0.3 concentrated in a single root near 2 is much more important. In the simulations made otherwise, this is what is significant.

Estimation

In the previous tables, we have give the results only for two values of s. But in fact, in order to find the first $s = \sigma_3$, the point of concentration and the associated weight, we had to repeat this operation a larger number of times. Finally, by repeating many times this type of calculation we find that the best estimation for the concentration point is $z_3 = 2.02 \pm 0.01$. It is indeed at this point that the roots are as close to a break as in the case above.

The calculation of s is a little more imprecise: we find $\sigma_3 = 0.25 \pm 0.02$. About the weight, it is estimated to be close to $q_3 = 0.3 \pm 0.03$.

As a matter of fact, we should study more in detail the behavior of the Gauss Jacobi quadrature close to the points of nonzero measure for more certainty. One can also take orthogonal polynomials of degree higher than 16.

The weight can also be refined during the removal of elements of 3-th Gaussian component (cf after). This too should be clarified.

Note however that, for this example, these estimates are rather correct.

4.3.4 Suppression of the first Gaussian component

We therefore estimated the parameters of the first Gaussian component. To find the second using the same technique, one must first eliminate points coming from the first component. There are various techniques possible. One chose to use a simple technique by eliminating $N_3 = nq_3$ points where n = 100000 is the size of sample. Normally N_3 is a good estimation of the number of points coming really from the first Gaussian component.

In order to eliminate the points of the first Gaussian component, we choose the following method. We know there are about a point of the sample coming from that component in a partition of $\mathbb{R}$ in N_3 intervals whose the Gaussian probability of each is $1/N_3$. One can then choose to remove the point nearest point of the middle of each interval.

This is what we do in the following Matlab program.

```
function y = RepartGauss(N,Esp,sigma)
L=1/N;
if N/2 − fix(N/2) > 0
for k=1:fix(N/2)
Mm(k) = sqrt(2)*sigma*erfinv(2*k*L);
end
Im = fliplr(Mm);
M=[Im,Mm];
M=[-Im,0,Mm];
end

if N/2-fix(N/2)==0
for k=1:N/2
Mm(k) = sqrt(2)*sigma*erfinv(2*(L/2+(k-1)*L));
end
Im = fliplr(Mm);
M=[Im,0,Mm];
M=[-Im,Mm];
end

y=M+Esp;


function y=VarSuppresDeUnEchGauss(X,N,Esp,sigma)
DD=size(X);
n=DD(2);
M=RepartGauss(N,Esp,sigma);
E=1:0;
for k=1:N
A=1;
j=0;
Mini=min(abs(X-M(k)));
while A==1
j=j+1;
if abs(X(j)-M(k)) ==Mini
Re=X(j);
X=[X(1:j-1),X(j+1:n)];
n=n-1;
A=0;
end
end
end
y=X;
```

The result obtained is provided in Figure 4.6. In this figure 4.6, in a sample of n = 100000 we have delete $nq_3 = 29500$ points supposed to belong to the first Gaussian component. The graph is the function histogram of Matlab (hist).

Note that these graphs can be used to better determine the weight, that is to say the size of the sample which has to be removed. Indeed, the curve has peaks on his right incompatible with

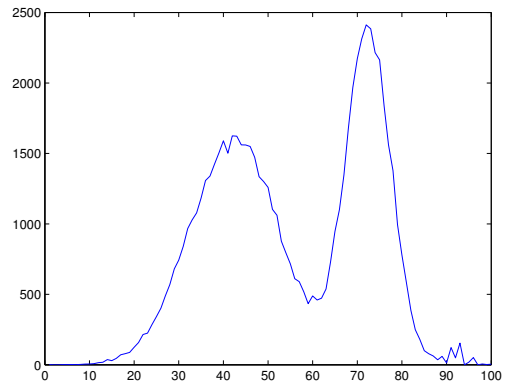

Figure 4.6: Suppression of the normal component centered in 2

a Gaussian mixture with two components.

Then, one can try another weight close to the estimation which we got. For example, if one removes only 30150 point we have the graph of Figure 4.7.

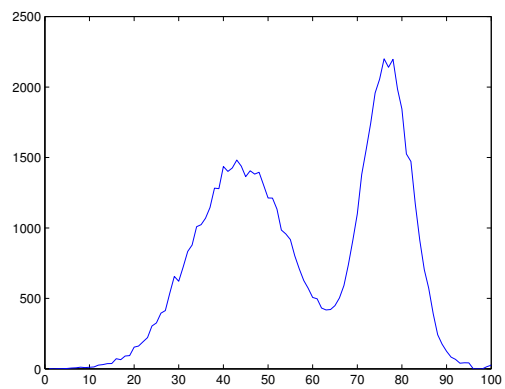

Figure 4.7: Suppression of 30150 elements of the normal component centered in 2

If we remove 30500 points we have the graph of Figure 4.8.
This kind of results can eventually allow us to refine the choice of s.

4.3.5 Estimation of the second Gaussian component

Then, we have eliminated the first Gaussian component. Thus we can apply the same method as before to find the second one.

The theoretical points of concentration which we must find are -3 and 0 and the weights $(0.4, 0.3) * (1000/705) = (0.5674, 0.4255)$: indeed, if you remove a component, the weight change.

Then we proceed as before. For example, for s = 1/2, we find the following tables.

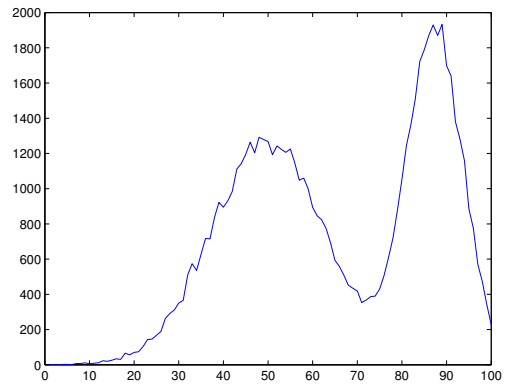

Figure 4.8: Suppression of 30500 elements of the normal component centered in 2

	1	2	3	4	5	6	7	8
$x_{1,s}$	-1.6913	0	0	0	0	0	0	0
$x_{2,s}$	-3.4429	-0.1445	0	0	0	0	0	0
$x_{3,s}$	-4.0317	-2.1841	0.1523	0	0	0	0	0
$x_{4,s}$	-4.5255	-2.9674	0.9399	-0.2487	0	0	0	0
$x_{5,s}$	-5.0765	-3.7960	-2.5014	1.4345	-0.0680	0	0	0
$x_{6,s}$	11.2814	-5.0631	-3.7812	-2.4895	1.4424	-0.0658	0	0
$x_{7,s}$	0.9211	0.9211	-5.1013	-3.8404	-2.5575	1.3459	-0.0921	0
$x_{8,s}$	-9.6775	-5.2213	-4.0607	1.4925	1.4925	-2.9102	-1.6563	0.0201
$x_{9,s}$	-8.0995	-5.5455	-4.4788	-3.3913	-2.2586	1.6054	1.6054	1.3315
$x_{10,s}$	-7.3226	-7.3226	-5.1064	2.1558	2.1558	1.6360	-4.0071	-2.9664
$x_{11,s}$	28.6401	-7.3263	-7.3263	-5.1077	2.1546	2.1546	1.6344	-4.0087
$x_{12,s}$	-7.2111	-6.9217	-6.9217	-5.2096	-4.0926	-3.0460	-1.9959	3.1344
$x_{13,s}$	18.8979	-7.2508	-6.9393	-6.9393	-5.2043	-4.0896	-3.0438	-1.9941
$x_{14,s}$	-7.0392	-7.0392	-5.9610	-5.9610	-5.0132	-3.9360	-2.9025	-1.8360
$x_{15,s}$	-322.2582	-7.0391	-7.0391	-5.9609	-5.9609	-5.0132	-3.9360	-2.9025
$x_{16,s}$	-7.1775	-7.1775	-6.3276	-6.3276	-5.3557	-4.3760	-3.4468	-2.4124

	9	10	11	12	13	14	15	16
$x_{9,s}$	-0.0528	0	0	0	0	0	0	0
$x_{10,s}$	-1.9245	-0.0187	0	0	0	0	0	0
$x_{11,s}$	-2.9682	-1.9263	-0.0188	0	0	0	0	0
$x_{12,s}$	1.9661	1.9661	1.5180	-0.0240	0	0	0	0
$x_{13,s}$	3.1862	1.9736	1.9736	1.5239	-0.0238	0	0	0
$x_{14,s}$	2.4317	1.7740	1.7740	0.7133	0.7133	0.0070	0	0
$x_{15,s}$	-1.8359	2.4318	1.7740	1.7740	0.7131	0.7131	0.0070	0
$x_{16,s}$	-1.6254	-1.6254	3.1937	2.8374	1.7779	1.7779	1.2243	-0.0470

	1	2	3	4	5	6	7	8
$\lambda_{1,s}$	1.0000	0	0	0	0	0	0	0
$\lambda_{2,s}$	0.4690	0.5310	0	0	0	0	0	0
$\lambda_{3,s}$	0.2653	0.3141	0.4207	0	0	0	0	0
$\lambda_{4,s}$	0.1093	0.3954	0.0839	0.4114	0	0	0	0
$\lambda_{5,s}$	0.0426	0.1789	0.3295	0.0391	0.4098	0	0	0
$\lambda_{6,s}$	0.0000	0.0431	0.1833	0.3246	0.0383	0.4107	0	0
$\lambda_{7,s}$	-0.0002	-0.0002	0.0388	0.1741	0.3300	0.0448	0.4127	0
$\lambda_{8,s}$	0.0000	0.0275	0.1339	0.0091	0.0091	0.2852	0.1349	0.4002
$\lambda_{9,s}$	0.0001	0.0073	0.0961	0.1930	0.2671	-0.0008	-0.0008	0.0457
$\lambda_{10,s}$	0.0001	0.0001	0.0295	-0.0003	-0.0003	0.0361	0.1501	0.2109
$\lambda_{11,s}$	0.0000	0.0001	0.0001	0.0295	-0.0003	-0.0003	0.0360	0.1497
$\lambda_{12,s}$	0.0005	0.0002	0.0002	0.0174	0.1558	0.1897	0.2136	0.0001
$\lambda_{13,s}$	-0.0000	0.0004	0.0002	0.0002	0.0187	0.1535	0.1929	0.2111
$\lambda_{14,s}$	0.0001	0.0001	-0.0001	-0.0001	0.0361	0.1551	0.2204	0.1651
$\lambda_{15,s}$	0.0000	0.0001	0.0001	-0.0001	-0.0001	0.0361	0.1551	0.2203
$\lambda_{16,s}$	0.0000	0.0000	-0.0000	-0.0000	0.0139	0.0912	0.1677	0.2891

	9	10	11	12	13	14	15	16
$\lambda_{9,s}$	0.3924	0	0	0	0	0	0	0
$\lambda_{10,s}$	0.1867	0.3871	0	0	0	0	0	0
$\lambda_{11,s}$	0.2112	0.1868	0.3873	0	0	0	0	0
$\lambda_{12,s}$	0.0000	0.0000	0.0389	0.3835	0	0	0	0
$\lambda_{13,s}$	0.0001	0.0002	0.0002	0.0378	0.3847	0	0	0
$\lambda_{14,s}$	0.0023	0.0012	0.0012	-0.0030	-0.0030	0.4247	0	0
$\lambda_{15,s}$	0.1652	0.0023	0.0012	0.0012	-0.0030	-0.0030	0.4246	0
$\lambda_{16,s}$	0.0076	0.0076	-0.0001	0.0007	-0.0009	-0.0009	0.0566	0.3675

Then, we find for the roots close to 0 and the associated weights the following table.

j	x_{j,j_0}	x_{j,j_0+1}	λ_{j,j_0}	λ_{j,j_0+1}
3	1.8420	0	0.3860	0
4	2.0039	0	0.3104	0
5	2.0427	0	0.2874	0
6	2.0721	0	0.2632	0
7	5 2.1145	1.5975	0.2128	0.1126
8	2.1680	1.8255	0.1415	0.1678
9	2.2072	1.9006	0.0976	0.2040
10	2.2346	1.9395	0.0729	0.2217
11	2.0854	1.7481	0.1950	0.1016
12	2.1919	1.8889	0.1099	0.1913
13	2.2038	1.9050	0.0986	0.2006
14	2.2019	1.9019	0.1005	0.1994
15	2.2042	1.9056	0.0982	0.2009
16	2.2181	1.9230	0.0861	0.2100

Calculation of the last Gaussian component

We must find the 3rd Gaussian component which is centered in -3 by our assumption and close to -3 by our estimates. In order to do this, it suffices to suppress the second Gaussian compo-

nent. For this, we apply the same method as in order to suppress the second Gaussian component.

In Figure 4.9, we have removed 28905 terms of the second Gaussian component among the 70500 terms remaining.

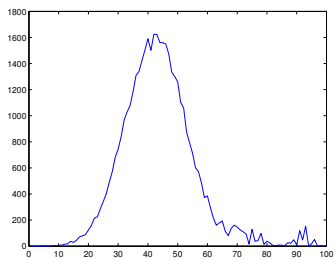

Figure 4.9: Suppression of 28905 terms of the second Gaussian component

In Figure 4.10, we have removed 30000 terms of the second Gaussian component among the 70500 terms remaining.

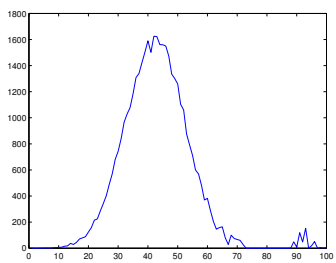

Figure 4.10: Suppression of 30000 terms of the second Gaussian component

It remains only to estimate the parameters of a single Gaussian component, which is very easy.

Final Estimation

By applying the method described here, we therefore found for the different Gaussian components.
$z_3 = 2.02 \pm 0.01$, $z_2 = 0.03 \pm 0.01$, $z_1 = -3.02 \pm 0.01$.
$\sigma_3 = 0.25 \pm 0.01$, $\sigma_2 = 0.50 \pm 0.01$, $\sigma_1 = 1.01 \pm 0.02$.
$q_3 = 0.3 \pm 0.03$, $q_2 = 0.3 \pm 0.03$, $q_1 = 0.4 \pm 0.02$.

Conclusion

So we have relatively good results especially when you consider that these results can be refined by a more detailed study of negative variances and of properties of the Gauss Jacobi quadrature and also of the method of removal of a Gaussian component.

Therefore in some cases this method can give good results for the calculation of Gaussian mixtures.

Of course, one may want to compare this method with other methods to find the Gaussian components, that is to say, essentially, the EM algorithm : by applying it to the sample which we have just estimate, it was found.

$z_3 = 2.1$, $z_2 = -0.2$ et $z_1 = -2.99$.
$\sigma_3 = 0.26$, $\sigma_2 = 0.49$, $\sigma_1 = 0.98$.
$q_3 = 0.28$, $q_2 = 0.31$, $q_1 = 0.39$.

So, in this case, our method gives result best than those of the EM algorithm.

Of course, all this should be clarified. But that is not purpose of this report which is to show the potentialities of the higher order variances and of Gauss Jacobi quadrature.

Chapter 5

Higher Order Regression

One application of the higher order variances is the introduction of higher order regression.

5.1 Notations and theorems

5.1.1 Notations

In order to introduce these, the easiest way is to cite a case we have actually a higher order regression.

Notations 5.1.1 *Let $(X,Y) \in \mathbb{R}^2$ be a random vector such that $\mathbb{E}\{|X^pY^{p'}|\} < \infty$ for all, p, $p' \in \mathbb{N}$. One supposes that there exists $p \in \mathbb{R}$ tel que Y = S(X) + $\sigma\epsilon$ where $S(X) = \sum_{s=1}^{p} \delta_s f_s(X)$ where δ_s = 0 or 1 , $\sum_{s=1}^{p} \delta_s = 1$, where $\epsilon \sim N(0,1)$ and where (δ_1, δ_2,...., δ_p) , ϵ and X are independent. We set P(δ_s=1) = q_s pour s=1,2,...,p. Moreover, we denote by m the distribution of X.*

Assume that $f_s \in L^2(\mathbb{R}, m)$ for s=1,2,...,p. Assume that $f_s(x) \neq f_{s'}(x)$ for m-almost all x and s≠s'.

Clearly, one can call curves $x \to f_s(x)$ regression curves of order p. Indeed, the knowledge of the curves $x \to f_s(x)$ and of the weights q_s allow to predict better the behavior of Y than a classical regression would make.

In this section, we shall show quickly how these curves can be found by applying the properties of the Gauss Jacobi quadrature and of variances of higher order. We will develop these points in details later.

If $\sigma = 0$, generally the variance of order p of conditional probability is equal at zero. Then, we introduce the following notations.

Notations 5.1.2 *Let μ^x a regular version of the conditional probability of Y given x. We suppose that $\mathbb{E}\{Y^{2p}\} < \infty$. We denote by $\{\tilde{Q}_j^x\}$, j=0,1,...,p, the family of orthonormal polynomials of μ^x : $\tilde{Q}_j^x(y) = y^j + d_{j-1}^x y^{j-1} + ... + d_1^x y + d_0^x$, and by $y_{p,1}^x$, $y_{p,2}^x$,...., $y_{p,p}^x$ the zeros of $\tilde{Q}_p^x$. We denote by $(\sigma_p^x)^2$ the associated variance of order p and by $\lambda_{p,1}^x$, $\lambda_{p,2}^x$,...., $\lambda_{p,p}^x$ the associated weights of Gauss-Jacobi quadrature. Moreover, we denote by $\{Q_j^x\}$, j=0,1,...,p, the family of orthonormal polynomials of μ^x.*

Then, the following property is a consequence of the properties of orthogonal polynomials.

Proposition 5.1.1 *Let $Q_{X,Y}$ be the probability of (X,Y). Let $\pi(y^p) = \sum_{s=0}^{p} y^s g_s(x)$ be the orthogonal projection in $L^2(\mathbb{R}^2, Q_{X,Y})$ of $(x,y) \longmapsto y^p$ onto the subspace generated by $(x,y) \longmapsto \sum_{s=0}^{p-1} y^s r_s(x)$ where $y^s r_s(x) \in L^2(\mathbb{R}^2, Q_{X,Y})$.*

We suppose that, for all t=1,2,...,p, $(y^x_{p,t})^p \in L^2(\mathbb{R}, m)$. Then, for m-almost all x,

$$\tilde{Q}^x_p(y) = y^p - \sum_{s=0}^{p-1} y^s g_s(x) = [y - y^x_{p,1}][y - y^x_{p,2}].....[y - y^x_{p,p}] \ .$$

Proof By definition of the orthogonal projection and of the orthogonalization process of Gram-schmidt, for m-almost all x, $\int \left([y - y^x_{p,1}][y - y^x_{p,2}].....[y - y^x_{p,p}]\right)^2.\mu^x(dy)$ is the minimum of $\int \left([y - k_1(x)][y - k_2(x)].....[y - k_p(x)]\right)^2.\mu^x(dy)$ when $\left(k_1(x), k_2(x), ..., k_p(x)\right) \in \mathbb{R}^p$.

Therefore, $\int\int \left([y-y^x_{p,1}][y-y^x_{p,2}].....[y-y^x_{p,p}]\right)^2.\mu^x(dy).m(dx)$ is smaller than $\int\int \left([y-k_1(x)][y-k_2(x)].....[y - k_p(x)]\right)^2.\mu^x(dy)m(dx)$ when $[y - k_1(x)][y - k_2(x)].....[y - k_p(x)] \in L^2(\mathbb{R}^2, Q_{X,Y})$.

Now, the minimum of $\int\int \left(y^p - \sum_{s=0}^{p-1} y^s r_s(x)\right)^2.\mu^x(dy)m(dx)$ when $\sum_{s=0}^{p-1} y^s r_s(x) \in L^2(\mathbb{R}^2, Q_{X,Y})$ is reached for the orthogonal projection in $L^2(\mathbb{R}^2, Q_{X,Y})$ of $(x, y) \longmapsto y^p$ onto the subspace spanned by $(x, y) \longmapsto \sum_{s=0}^{p-1} y^s r_s(x)$ and it is reached also for $(k_1(x), ..., k_p(x)) = (y^x_{p,1}, ..., y^x_{p,p})$: $[y - k_1(x)][y - k_2(x)].....[y - k_p(x)] = y^p - \sum_{s=0}^{p-1} y^s g_s(x)$. ■

5.1.2 Properties

If $\sigma = 0$, generally, $(\sigma^x_p)^2 = 0$.

Proposition 5.1.2 *We suppose $\sigma = 0$. Then, $(\sigma^x_p)^2 = 0$ for m-almost all $x \in \mathbb{R}$, i.e. $\mathbb{E}\{\tilde{Q}^x_p(Y)^2\} = 0$. Moreover, $y^x_{p,s} = f_s(x)$ for s=1,2,...,p and $q_s = \lambda^x_{p,s}$ for s=1,2,...,p .*

Remark that if $\sigma = 0$ and if $r_s(x_0) = r_t(x_0)$ for a (t,s) such that $t \neq s$, $(\sigma^{x_0}_p)^2$ does not exists. But, in this case, $(\sigma^{x_0}_{p-1})^2 = 0$.

Now recall how to compute the $\lambda^x_{p,t}$'s.

Proposition 5.1.3 *Let $\ell^x_s(y) = \frac{\tilde{Q}^x_p(y)}{(y-y^x_{p,s})\tilde{Q}^{x\,\prime}_p(y^x_{p,s})}$. Then, $\lambda^x_{p,s} = \int \ell^x_s(y).\mu^x(dy)$ for s=1,2,...,p.*
Moreover, if $\sigma = 0$, $q_s = \lambda^x_{p,s} = \mathbb{E}\{\ell^X_s(Y)\}$ for s=1,2,...,p.

For example, if $\sigma = 0$, $\ell^x_1(y) = \frac{(y-f_2(x))(y-f_3(x)).....(y-f_p(x))}{(f_1(x)-f_2(x))(f_1(x)-f_3(x)).....(f_1(x)-f_p(x))}$.

For example, suppose that p=2 and that $\sigma_2 = 0$: $q_1 = \lambda^x_{2,1}$ and $q_2 = \lambda^x_{2,2}$. Then, $(Y - f_1(X))(Y - f_2(X)) = 0$ and $\mathbb{E}\{(Y - f_1(X))^2(Y - f_2(X))^2\} = 0$.
More generally $\mathbb{E}\{(Y-f_1(X))^2(Y-f_2(X))^2\}$ is the minimum of $\mathbb{E}\{(Y-k_1(X))^2(Y-k_2(X))^2\}$ when $k_s \in L^2(\mathbb{R}, m)$ for s=1,2. It is also the minimum of $\mathbb{E}\{(Y^2 - r(X)Y - s(X))^2\}$ when r and s are functions of $L^2(\mathbb{R}, m)$: the minimum $S_2 = \mathbb{E}\{(Y^2 - r(X)Y - s(X))^2\}$ holds for $r(x) = r^*(x) = f_1(x) + f_2(x)$ and $s(x) = s^*(x) = -f_1(x)f_2(x)$.

Now, this minimum measures the distance L^2 between $(x, y) \longmapsto y^2$ and the subspace spanned by $(x, y) \longmapsto r(x)y + s(x)$ when $r, s \in L^2(\mathbb{R}, m)$. Therefore, $r^*(x)y + s^*(x)$ is the orthogonal projection of $(x, y) \longmapsto y^2$ onto the subspace spanned by r(x)y+ s(x).

5.1.3 Method of computation

At first, we prove the following result in the same way that the proposition 4.1.3.

Proposition 5.1.4 *Let $U \sim N(0,1)$ be a random variable independent with $(\delta_1, \delta_2, ..., \delta_p)$, X, and $(\epsilon_1, \epsilon_2, ..., \epsilon_p)$. Then, for all $q \in \mathbb{N}$,*

$$\mathbb{E}\Big\{\Big(\sum_{t=1}^{p}(\delta_t f_t(X) + \sigma_t \epsilon_t) + isU\Big)^q\Big\} = \mathbb{E}\Big\{T_s(X)^q\Big\}$$

where

$$T_s(X) = \sum_{t=1}^{p} \delta_t \big[(f_t(X) + \sqrt{\sigma_t^2 - s^2}\ \epsilon_t)\big]\ .$$

Then, it is enough easy to calculate the parameters of higher order regression by using the same method as in section 4.2.

For example suppose that we know the conditional density for all x. Then, in the theoretical case, it suffices to calculate s for some x and the associated measure μ^x.

Then, if for example we can calculate the conditional moments, then we shall calculate the polynomials Q_j^x and then the roots $y_{p,s}^x$ and weight $\lambda_{p,s}^x$. Otherwise, we can calculate the roots and weights in some points x and they are continued by continuity if this hypothesis holds. One will do the same in case of estimation.

5.2 Examples : regression of order 2

Now we study some examples of regression of order 2. In the following figures we have the graphs of the regression of order 2, empirical or not, when $X \in [0,1]$, $f_1(X) = cos(\pi X) - 3$ and $f_2(X) = 2sin(10(X + \sqrt{2})) + 3$, $Y = Z(X) + \sigma\epsilon$, with for weights $q_1 = 1/4$, $q_2 = 3/4$, and with $\sigma = 1$ (figures 5.1, 5.4) , $\sigma = 2$ (figures 5.2, 5.5), $\sigma = 4$ (figures 5.3, 5.6).

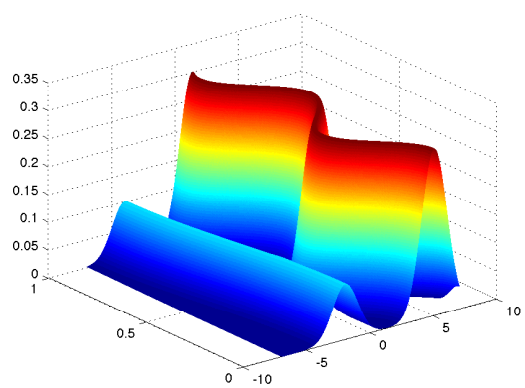

Figure 5.1: $\sigma = 1$

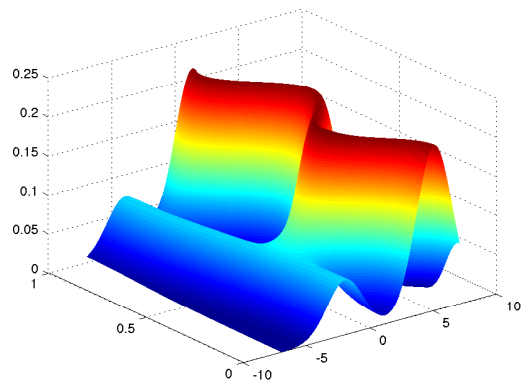

Figure 5.2: $\sigma = 2$

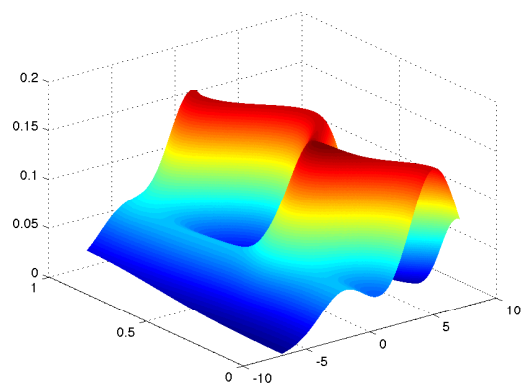

Figure 5.3: $\sigma = 4$

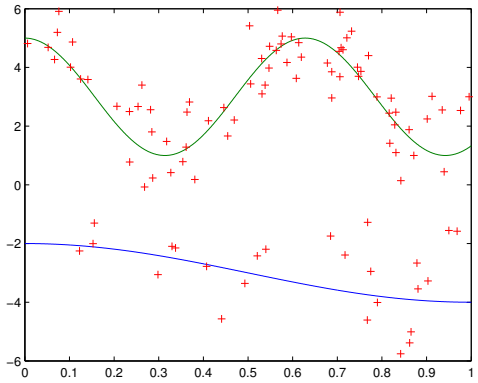

Figure 5.4: $\sigma = 1$

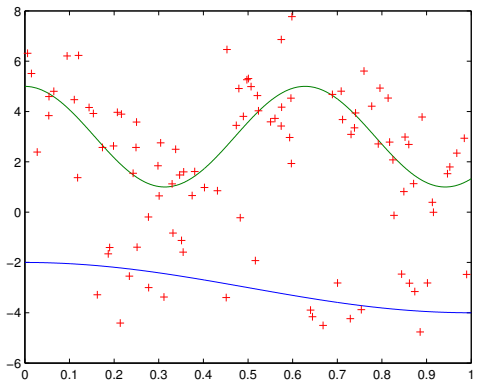

Figure 5.5: $\sigma = 2$

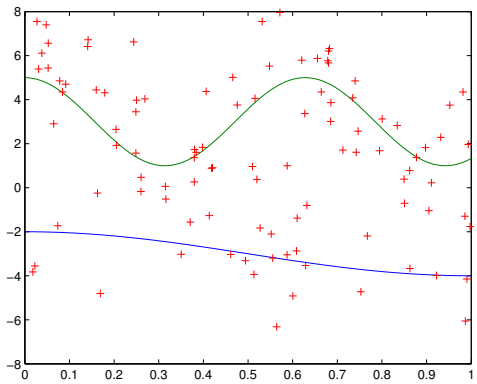

Figure 5.6: $\sigma = 4$

Appendix A

Variance of order 3

In order to better study the variances and the regression of order j, it is useful to have the exact formulas for the parameters for the first j. Here we give these parameters in case j = 3.

A.1 Elementary calculations

A.1.1 Some formulas

We know that

$$\bar{P}_2^X(x) = x^2 - \frac{M_3 - M_1 M_2}{M_2 - M_1^2}(x - M_1) - M_2 \ ,$$

$$\sigma_2^2 = M_4 - \frac{(M_3 - M_1 M_2)^2}{M_2 - M_1^2} \ - \ M_2^2$$

$$\ell_1^j(x) = \frac{(x - x_{j,2})(x - x_{j,3})...(x - x_{j,j})}{(x_{j,1} - x_{j,2})(x_{j,1} - x_{j,3})...(x_{j,1} - x_{j,j})} \ .$$

A.1.2 Polynomials

By the Gram Schmidt Process,

$$\tilde{P}_j(x) = x^j - \sum_{s=0}^{j-1} \mathbb{E}\{X^j P_s(X)\} P_s(x) \ .$$

Therefore,

$$\tilde{P}_3 = x^3 - \mathbb{E}\{X^3 P_2(X)\} P_2(x) - \mathbb{E}\{X^3 P_1(X)\} P_1(x) - \mathbb{E}\{X^3 P_0(X)\} P_0(x)$$

$$= x^3 - \frac{\left[M_5 - \frac{M_3 - M_1 M_2}{M_2 - M_1^2}(M_4 - M_3 M_1) - M_3 M_2\right]}{\sigma_2} P_2(x) - \frac{M_4 - M_3 M_1}{\sigma_1} P_1(x) - M_3$$

$$= x^3 - \frac{\left[M_5 - \frac{M_3 - M_1 M_2}{M_2 - M_1^2}(M_4 - M_3 M_1) - M_3 M_2\right]}{M_4 - \frac{(M_3 - M_1 M_2)^2}{M_2 - M_1^2} \ - \ M_2^2} \left[x^2 - \frac{M_3 - M_1 M_2}{M_2 - M_1^2}(x - M_1) - M_2\right]$$

$$-\frac{M_4 - M_3M_1}{M_2 - M_1^2}(x - M_1) - M_3$$

$$= x^3 - \frac{\left[M_5 - \frac{M_3 - M_1M_2}{M_2 - M_1^2}(M_4 - M_3M_1) - M_3M_2\right]}{M_4 - \frac{(M_3 - M_1M_2)^2}{M_2 - M_1^2} - M_2^2}x^2$$

$$+\frac{\left[M_5 - \frac{M_3 - M_1M_2}{M_2 - M_1^2}(M_4 - M_3M_1) - M_3M_2\right]}{M_4 - \frac{(M_3 - M_1M_2)^2}{M_2 - M_1^2} - M_2^2}\frac{M_3 - M_1M_2}{M_2 - M_1^2}(x - M_1) - \frac{M_4 - M_3M_1}{M_2 - M_1^2}(x - M_1)$$

$$+\frac{\left[M_5 - \frac{M_3 - M_1M_2}{M_2 - M_1^2}(M_4 - M_3M_1) - M_3M_2\right]}{M_4 - \frac{(M_3 - M_1M_2)^2}{M_2 - M_1^2} - M_2^2}M_2 - M_3 \ .$$

A.1.3 Weights

We know that

$$\ell_1^3(x) = \frac{(x - x_{3,2})(x - x_{3,3})}{(x_{3,1} - x_{3,2})(x_{3,1} - x_{3,3})} \ , \ \ell_2^3(x) = \frac{(x - x_{3,1})(x - x_{3,3})}{(x_{3,2} - x_{3,1})(x_{3,2} - x_{3,3})} \ ,$$

$$\ell_3^3(x) = \frac{(x - x_{3,1})(x - x_{3,2})}{(x_{3,3} - x_{3,1})(x_{3,3} - x_{3,2})} \ .$$

Therefore,

$$\lambda_{3,1} = \frac{M_2 - [x_{3,2} + x_{3,3}]M_1 + x_{3,2}x_{3,3}}{(x_{3,1} - x_{3,2})(x_{3,1} - x_{3,3})} \ ,$$

$$\lambda_{3,2} = \frac{M_2 - [x_{3,1} + x_{3,3}]M_1 + x_{3,1}x_{3,3}}{(x_{3,2} - x_{3,1})(x_{3,2} - x_{3,3})} \ ,$$

$$\lambda_{3,3} = \frac{M_2 - [x_{3,1} + x_{3,2}]M_1 - x_{3,1}x_{3,2}}{(x_{3,3} - x_{3,1})(x_{3,3} - x_{3,2})} \ .$$

A.1.4 Variance of order 3

By proposition 1.1.7, we know that

$$\sigma_j^2 = M_{2j} - \sum_{s=0}^{j-1}\beta_{j,s}^2 \ where \ \beta_{j,s} = \int x^j P_s(x).mdx \, .$$

Therefore, for j=3,

$$\beta_{3,2} = \int x^3 P_2(x).mdx = \frac{1}{\sigma_2}\int x^3\Big[x^2 - \frac{M_3 - M_1M_2}{M_2 - M_1^2}(x - M_1) - M_2\Big].m(dx)$$

$$= \frac{\left[M_5 - \frac{M_3 - M_1M_2}{M_2 - M_1^2}(M_4 - M_3M_1) - M_3M_2\right]}{\sigma_2} \ .$$

Moreover,

$$\beta_{3,1} = \int x^3 P_1(x).mdx = \int x^3 \big[x - M_1\big]/\sigma_1.mdx = \frac{\big[M_4 - M_3 M_1\big]}{\sigma_1} \ .$$

and

$$\beta_{3,0} = \int x^3 P_0(x).mdx = M_3 \ .$$

Therefore,

$$\sigma_3^2 = M_6 - \frac{\big[M_5 - \frac{M_3 - M_1 M_2}{M_2 - M_1^2}(M_4 - M_3 M_1) - M_3 M_2\big]^2}{M_4 - \frac{(M_3 - M_1 M_2)^2}{M_2 - M_1^2} \ - \ M_2^2} - \frac{\big[M_4 - M_3 M_1\big]^2}{M_2 - M_1^2} - M_3^2 \ .$$

Bibliography

[1] LANCASTER H. O. (1960) Orthogonal models for contingency tables. Developments in statistics. Academic Press, New York.

[2] BLACHER R. (1993) Higher Order Correlation Coefficients. Statistics 25, 1-15.

[3] BLACHER R. (1995) Central limit theorem by polynomial dependence coefficients. Journal of computational and applied mathematics 57, 45-56.

[4] BLACHER R. (1983) Indicateurs de dépendance fournis par le développement en série de la densité de probabilité. Thèse 3 Cycle, Université Joseph Fourier de Grenoble. Hal archives ouvertes : http://tel.archives-ouvertes.fr

[5] BLACHER R. (1990) Quelques applications des fonctions orthogonales en probabilités et statistiques, Thèse de l'université Joseph Fourier, Grenoble 1. Hal archives ouvertes : http://tel.archives-ouvertes.fr

[6] BLACHER R. (1984) Coefficient de correlation d'ordre (i,j) et variances d'ordre i. Rapport de recherche 463, Laboratoire TIM3 IMAG, Université de Grenoble.

[7] BLACHER R. (1984) Coefficient de correlation d'ordre (i,j) et variances d'ordre i. Symposium International de Bar Le Duc, Polynomes orthogonaux et applications, Springer Verlag, 1171, 475-486.

[8] BLACHER R. (2003) Multivariate quadratic forms of random vectors. Journal of multivariate analysis 87, 2-23.

[9] SERFLING R.J. (1980) Approximation theorems of mathematical statistics. Wiley, New York

[10] CHIHARA. (1978) An introduction to orthogonal polynomials Gordon and Breach, New York

[11] FREUD Geza. (1971) Orthogonal polynomials. Pergamon Press

[12] HIDA T. (1980) Brownian motion. Springer, Berlin

[13] SZEGO G. (1959) Orthogonal polynomials. Americal Mathematical Society, Providence.