ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΕΦΑΡΜΟΣΜΕΝΩΝ ΜΑΘΗΜΑΤΙΚΩΝ ΚΑΙ ΦΥΣΙΚΩΝ ΕΠΙΣΤΗΜΩΝ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

«ΜΑΘΗΜΑΤΙΚΗ ΠΡΟΤΥΠΟΠΟΙΗΣΗ σε ΣΥΓΧΡΟΝΕΣ ΤΕΧΝΟΛΟΓΙΕΣ και την ΟΙΚΟΝΟΜΙΑ»

ΤΙΤΛΟΣ ΜΕΤΑΠΤΥΧΙΑΚΗΣ ΕΡΓΑΣΙΑΣ

INTEGRAL APPROXIMATION OF MULTIDIMENSIONAL PDFS AND ITS CONNECTION WITH LARGE SAMPLE THEORY

ΑΒΕΚΛΟΥΡΗΣ ΑΓΓΕΛΟΣ

ΑΡΙΘΜΟΣ ΜΗΤΡΩΟΥ: **09313001**

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:
ΑΘΑΝΑΣΟΥΛΗΣ Α. Γ

ΑΘΗΝΑ, Ιούνιος 2015

Angelos Aveklouris

MSc Student, NTUA
angelos1990@yahoo.com

Gerassimos A. Athanassoulis

Professor, NTUA
makathan@gmail.com

INTEGRAL APPROXIMATION OF MULTIDIMENSIONAL PDFS AND ITS CONNECTION WITH LARGE SAMPLE THEORY

Athens, June 2015

Acknowledgments

After a year, I finally finished my master dissertation and I am ready for a new challenge in my life. During this year, I improved my mathematical skills and I learnt new things.

First of all I would like to thank my supervising professor, NTUA professor Gerasimos Athanassoulis, for the trust he showed me, for all the insights and the working tools he helped me to develop. He showed me new mathematical fields and the way towards research. Besides, he taught me how I should treat a problem. In addition, he gave me confidence and encouragement. A special thank you for introducing and encouraging me to work on Stochastic Mathematics and Probability.

I would like to thank NTUA professor Vassilis Papanicolaou and all my professors in my master program for offering their knowledge in mathematics and their help to improve my mathematical skills.

I thank my colleagues and my friends for the anxieties, the thoughts and the pleasant moments we shared. Especially, I thank Ph.D student Kostas Mamis who always answered my questions and my colleague and good friend, Alexis Mpotsis, for helping me with this text.

Last, I would like to thank my parents, my brother and my uncle George for their love and constant support whenever I needed it. A special thank to my friend Lena for her love and her encouragement.

Summary

In many systems (e.g. random/stochastics dynamical systems) in which researchers are interested in finding probability density functions (pdf(s)), which describe the system, we have observed that they approximate these pdfs by a superposition of some appropriate functions (which are called approximation to the identity or delta family), which is different at any approximated point, generally; see *Athanassoulis, Sapsis (2008)*. Furthermore, when researchers have data, they use estimators which are similar with the aforementioned approximation form (delta estimators); see *Brox et all. (2007), Bengio (2003)*. Although their results are very effective, there is not a proved mathematical construction in the literature which allows this approximation. Our motivation was to cover this gap in the literature, although this good has not been fully achieved, and there is one question remaining unanswered.

Initially, we want to give a rigorous mathematical construction of this approximation, and secondly to link this result with other mathematical fields (classical analysis, estimation Statistics, and Large Sample Theory) highlighting their similarities.

In Chapter 2, we show the importance of the approximation method which we shall use, and the necessity of existence of a discrete approximation of a multidimensional pdf. In particular, in the first part of this chapter, we prove completely the Fourier inversion Theorem. This is an important theorem in mathematical analysis, which is based on the convergence which we shall describe. Continuing, in the second part of Ch. 2, we cite an application of a random dynamical system where we can realize that we need a discrete approximation form.

In the following chapter (Chapter 3) we present the main results of our work. In particular, we present a series of theorems establishing that any multidimensional pdf can be approximated by a superposition of Gaussian pdfs (or more general functions/kernels). In the simplest case, we can assume spherically symmetric kernels, and easily can be generalized to ellipsoidal ones in orientation of the usual basis of $\mathbb{R}^N$. In the most interesting case, which we study here, we can assume general ellipsoidal kernels in any orientation. Firstly, we prove that any pdf can “generate” a family of approximation to the identity (or delta family). Subsequently, we show the central theorem of this chapter, in which we prove an integral approximation of any pdf h, namely $\lim_{\mathbf{G}_{\boldsymbol{\lambda}}(x),\prec} \int_{\mathbb{R}^N} K_{\mathbf{G}_{\boldsymbol{\lambda}}(x)}(\boldsymbol{x}-\boldsymbol{u})h(\boldsymbol{u})d\boldsymbol{u} = h(\boldsymbol{x})$,

where $\mathbf{G}_{\boldsymbol{\lambda}}(\boldsymbol{x})$ is a family of matrices in an appropriate directed set. In the third part of this chapter, using the Riemann Sum, we give a discrete approximation form by a superposition of Gaussian pdfs, and we prove that we can change this superposition at any approximated point of support of pdf. It means, we can achieve better approximation with less points of the partition, i.e., we can decrease the computational cost. In other words, we prove that $h(\boldsymbol{x}) \approx \sum_{\ell \in \mathcal{L}} p_\ell \, K(\boldsymbol{x}-\boldsymbol{u}_\ell \,; \tilde{\mathbf{C}}(\boldsymbol{x}))$, where $\tilde{\mathbf{C}}(\boldsymbol{x})$ is the localized covariance matrix,

and p_ℓ are unknown constants. As we shall see, there are many common points with the Fourier inversion Theorem. The basic deference is that we require (it is a nature condition) positivity of family of approximation to the identity. This additional condition makes it different, and it is the reason we study the approximation of pdfs separately.

Last, in Chapter 4, we deal with estimation of pdfs. We define the delta sequences estimators, $\hat{f}_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} K_m(\boldsymbol{x},\boldsymbol{x}_i)$, and we present the most basic of the kernel estimators (which are delta ones). Then, we focus on theorems of asymptotic unbiasedness, ;i.e., $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x})\right)-f(\boldsymbol{x}) \xrightarrow{n\to\infty} 0$, and consistency of variation term, i.e., $\hat{f}_n(\boldsymbol{x})-\mathrm{E}\left(\hat{f}(\boldsymbol{x})\right) \xrightarrow{P} 0$ in $L_1\left(\mathbb{R}^N,\mathfrak{B}^N,\mu\right)$. In last years, functional data have been developed in statistics. For this reason, we generalized the above definitions in infinite dimensional spaces, and we present some recent published results in Banach spaces.

At first glance, integral approximation, and estimation of a pdf are different methods. However, the fact that the expected value of a delta estimator can be written in the following form, $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x})\right) = \int_{\mathbb{R}^N} K_{m_n}(\boldsymbol{x},\mathbf{z}) f(\mathbf{z})\mu(dz)$, allow us to link these two methods. Studying the bias problem, we shall show that $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x})\right) \xrightarrow{n\to\infty} f(\boldsymbol{x})$, which is the same problem as the integral approximation of a pdf. On the other hand, if we assume all p_ℓ are equal, then the discrete form which we derive in Chapter 3, looks like with a delta estimator. Finally, as we shall notice, either we refer to the classical analysis, either to the approximation-estimation of pdfs, the properties of families of approximation to the identity play a key role on proving the theorems.

Σύνοψη

Σε αρκετά συστήματα (π.χ. τυχαία/στοχαστικά δυναμικά συστήματα) όπου οι ερευνητές ενδιαφέρονται να βρούνε άγνωστες συναρτήσεις πυκνότητας πιθανότητας (σπп), οι οποίες περιγράφουν το σύστημα, παρατηρήσαμε πως προσεγγίζουν αυτές τις άγνωστες σπп με μια υπέρθεση κατάλληλων συναρτήσεων (οικογένειες προσέγγισης την μονάδας ή δέλτα οικογένειες), διαφορετική γενικά σε κάθε σημείο προς προσέγγιση; Βλέπε *Athanassoulis, Sapsis (2008.* Επιπλέον, όταν οι ερευνητές έχουν δεδομένα, χρησιμοποιούν εκτιμητές που μοιάζουν με την προαναφερθείσα προσέγγιση (δέλτα εκτιμητές); Βλέπε see *Brox et all. (2007), Bengio (2003).* Παρόλο που τα αποτελέσματα στα οποία καταλήγουν είναι αρκετά ικανοποιητικά στην βιβλιογραφία δεν έχουμε, όσο και αν ψάξαμε σε όποια πηγή μπορούσαμε, βρει μια αποδεδειγμένη μαθηματική κατασκευή που να επιτρέπει την παραπάνω προσέγγιση.Στόχος αυτής της εργασίας ήταν να καλύψει ακριβώς αυτό το κενό, παρόλο που δεν επιτευχθεί από πλήρη επιτυχία κι ένα ερώτημα παραμένει αναπάντητο.

Επιθυμούμε αρχικά να δώσουμε μια αυστηρή μαθηματική κατασκευή αυτής της προσέγγισης, και δεύτερον, να συνδέσουμε το αποτέλεσμα μας με άλλους τομείς των μαθηματικών (κλασσική ανάλυση, εκτιμητική στατιστική και θεωρία μεγάλων δειγμάτων), αναδεικνύοντας τα κοινά τους στοιχεία.

Στο Κεφάλαιο 2 δείχνουμε την σημαντικότητα της προσέγγισης που θα χρησιμοποιήσουμε και την αναγκαιότητα να υπάρχει μια διακριτή μορφή προσέγγισης μιας σπп. Συγκεκριμένα στο πρώτο μέρος του κεφαλαίου αποδεικνύουμε πλήρως το θεώρημα αντιστροφής του μετασχηματισμού Fourier. Μια πολύ σημαντική εφαρμογή στη μαθηματική ανάλυση όπου βασίζεται στην σύγκλιση που θα περιγράψουμε. Στη συνέχεια, στο δεύτερο μέρος του ίδιου κεφαλαίου παραθέτουμε μια εφαρμογή ενός στοχαστικού δυναμικού συστήματος κατά το οποίο γίνεται αντιληπτό πως χρειαζόμαστε μία διακριτή μορφή προσέγγισης.

Στο επόμενο κεφάλαιο (Κεφ. 3) παρουσιάζουμε τα πιο σημαντικά αποτελέσματα αυτής της δουλείας. Συγκεκριμένα παρουσιάζουμε μια σειρά από θεωρήματα δείχνοντας ότι κάθε πολυδιάστατη σπп μπορεί να προσεγγιστεί όσο καλά θέλουμε από μια υπέρθεση Gaussian σπп (ή και γενικότερων συναρτήσεων). Στην απλούστερη περίπτωση μπορούμε να θεωρήσουμε σφαιρικούς πυρήνες, και εύκολα να γενικεύσουμε σε ελλειψοειδής στην κατεύθυνση της συνήθης βάσης του $\mathbb{R}^N$. Στην πιο ενδιαφέρουσα περίπτωση, την οποία μελετάμε εδώ, θεωρούμε γενικούς ελλειψοειδής πυρήνες σε οποιοδήποτε προσανατολισμό. Αρχικά αποδεικνύουμε ότι κάθε σπп ΄΄γεννά΄΄ μια τέτοια οικογένεια. Στη συνέχεια, δείχνουμε το κεντρικό θεώρημα του κεφαλαίου στο οποίο αποδεικνύουμε μια συνεχή προσέγγιση για κάθε σπп., h, δηλ. $\lim\limits_{\mathbf{G}_{\lambda}(\boldsymbol{x}),\prec} \int\limits_{\mathbb{R}^N} K_{\mathbf{G}_{\lambda}(\boldsymbol{x})}(\boldsymbol{x}-\boldsymbol{u})h(\boldsymbol{u})d\boldsymbol{u} = h(\boldsymbol{x})$,

όπου $\mathbf{G}_{\lambda}(\boldsymbol{x})$ είναι μια οικογένεια πινάκων που ανήκει σε κάποιο κατάλληλο κατευθυνόμενο σύνολο. Στο τρίτο μέρος, χρησιμοποιώντας το Riemann άθροισμα, δίνουμε μια διακριτή προσέγγιση μιας σπп από υπέρθεση Gaussian συναρτήσεων. Επιπλέον, το σημαντικότερο είναι ότι σε κάθε σημείο στο στήριγμα της σπп μπορούμε να αλλάζουμε αυτή την υπέρθεση. Αυτό είναι πολύ σημαντικό αποτέλεσμα διότι μπορούμε να πετύχουμε καλύτερη προσέγγιση

με λιγότερα σημεία της διαμέρισης που χρησιμοποιούμε για την προσέγγιση, το οποίο σημαίνει μείωση του υπολογιστικού κόστους. Με άλλα λόγια θα δείξουμε ότι $h(\boldsymbol{x}) \approx \sum_{\ell \in \mathcal{L}} p_\ell \, K\left(\boldsymbol{x} - \boldsymbol{u}_\ell \, ; \tilde{\mathbf{C}}(\boldsymbol{x}) \right)$, όπου $\tilde{\mathbf{C}}(\boldsymbol{x})$ είναι ο τοπικός πίνακας συνδιακύμανσης. Όπως θα δούμε υπάρχουν αρκετά κοινά στοιχεία με τα θεωρήματα της κλασσικής ανάλυσης. Η ουσιαστική διαφορά είναι ότι ζητάμε οι οικογένειες προσέγγισης της μονάδας, που θα χρησιμοποιήσουμε, να είναι θετικές (φυσική συνθήκη). Αυτή η επιπλέον ιδιότητα διαφοροποιεί τα πράγματα και είναι ο λόγος για τον οποίο πρέπει να μελετήσουμε ξεχωριστά την προσέγγιση των σππ.

Τέλος, στο Κεφάλαιο 4, ασχολούμαστε με την εκτίμηση σππ. Ορίζουμε γενικά τους δέλτα εκτιμητές $\hat{f}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_m\left(\boldsymbol{x}, \boldsymbol{x}_i \right)$, και γίνεται η παρουσίαση των πιο σημαντικών δέλτα εκτιμητών. Στη συνέχεια, επικεντρωνόμαστε θεωρήματα ασυμπτωτικής αμεροληψίας, δηλ. $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) - f(\boldsymbol{x}) \xrightarrow{n \to \infty} 0$, και συνέπειας, δηλ. $\hat{f}_n(\boldsymbol{x}) - \mathrm{E}\left(\hat{f}(\boldsymbol{x}) \right) \xrightarrow{P} 0$ των δέλτα εκτιμητών στον $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$. Τα τελευταία χρόνια στην στατιστική έχουν αναπτυχθεί τα λεγόμενα συναρτησιακά δεδομένα. Για αυτό τον λόγο, γενικεύουμε τους παραπάνω ορισμούς σε Banach χώρους και παρουσιάζουμε τα σημαντικά αποτελέσματα που πρόσφατα έχουν δημοσιευτεί σε τέτοιους χώρους.

Με μια πρώτη ματιά, η προσέγγιση και η εκτίμηση μιας σππ φαίνονται διαφορετικές μέθοδοι. Παρόλο αυτά, το γεγονός πως η μέση τιμή ενός δέλτα εκτιμητή γράφεται στην ακόλουθη μορφή, $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) = \int_{\mathbb{R}^N} K_{m_n}(\boldsymbol{x}, \mathbf{z}) \, f(\mathbf{z}) \mu(dz)$, μας επιτρέπει να συνδέσουμε αυτές τις δύο μεθόδους. Συγκεκριμένα μελετώντας το πρόβλημα της αμεροληψίας θα δείξουμε ότι $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) \xrightarrow{n \to \infty} f(\boldsymbol{x})$, όπου είναι το ίδιο πρόβλημα που αντιμετωπίζουμε στη προσέγγιση μιας σππ. Από την άλλη μεριά, αν στην διακριτή μορφή προσέγγισης που δίνουμε στο Κεφ. 3 θεωρήσουμε όλα τα p_ℓ ίσα, τότε μοιάζει αρκετά με έναν δέλτα εκτιμιτή. Επίσης, όπως θα παρατηρήσουμε είτε αναφερόμαστε στην κλασσική ανάλυση, είτε στη προσέγγιση-εκτίμηση σππ καθοριστικό ρόλο στις αποδείξεις των θεωρημάτων παίζουν οι ιδιότητες των οικογενειών προσέγγισης της μονάδας.

TABLE OF CONTENTS

NOTATION

$\left(K_\delta\right)_{\delta>0}$	an approximation to the identity
λ	Lebesgue measure
$\mathrm{Leb}(f)$	Lebesgue set of a function $f \in L^1\left(\mathbb{R}^N\right)$
$\hat{f}$	Fourier transform of a function $f \in L^1\left(\mathbb{R}^N\right)$ (Chapter 2)
$x(t;\omega)$	random (or stochastic) response (or output)
$y(t;\omega)$	random (or stochastic) excitation o(or input
$\mathrm{P}_{\mathscr{X}}$	infinite dimensional measure of the Borel sets of the sample (functional) Banach $\mathscr{X}$ space
$\mathscr{F}_{xy}\ u,v$	characteristic functional of joint response-excitation
$\varphi_{x(t)y(s)}(u,v)$	characteristic function of joint response-excitation
$f_{x(t)y(s)}(a,\mathrm{b})$	joint (two-time) response excitation probability density function
$f_{y(t)}(\mathrm{b})$	excitation pdf
$\delta\ \cdot - t$	Dirac delta generalized function at time t
$\dfrac{\delta \mathscr{F}_{xy}\ u,v}{\delta u(t)}$	Volterra $u-$partial derivative of $\mathscr{F}_{xy}$
$C_b\ \mathbb{R}^N \to \mathbb{R}$	the space of bounded continuous functions defined on $\mathbb{R}^N$
f_{Gauss}	Gaussian pdf
$\mathbf{Q}$	proper rotation matrix
$\mathcal{E}_{\mathbf{Q}}\left(\mathbf{0};\delta/\lambda\right)$	ellipsoidal by orientation $\mathbf{Q}$
$\mathbf{C}$	covariance matrix of Gaussian pdf
$\tilde{\mathbf{C}}$	$\lambda-$dependent localized covariance matrix
ρ_{ij}	correlation coefficients of covariance matrix $\mathbf{C}$
$\tilde{\rho}_{ij}$	correlation coefficients of the $\lambda-$dependent localized covariance matrix
$\left(\Omega,\mathcal{F},P\right)$	probability space
$\operatorname{ess\,sup} f$	essential supremum of f
$\hat{f}$	estimator of f (Chapter 4)
$\hat{f}_B$	balloon estimator
$\hat{f}_S$	sample point estimator
$\hat{f}_{BS}$	binned sample point estimator
K_{m_n}	delta sequence
E	Banach space

ABBREVIATIONS

Pdf	Probability Density Function
ARK	Approximate Reproducing Kernels
MISE	Mean Integrated Square Error
AMISE	Asymptotic MISE

CHAPTER 1

Introduction

Table of Contents

1.1. Motivation and scope of the present work

In many systems that involve random functions we are interested in finding their probability density functions (pdf(s)). Examples of such systems are the random/stochastic dynamical systems described by random/stochastic differential equations. Athanassoulis, Sapsis (2008), *New partial differential equations governing the joint, response-excitation,probability distributions of nonlinear systems, under general stochastic excitation*, and Venturi et al. (2012), *A computable evolution equation for the joint response-excitation probability density function of stochastic dynamical systems*, study such models, in which they derive first order differential equations for the joint response-excitation pdf of a particular system. In addition, we often encounter such systems in Statistics. In Statistics, researchers use various statistical estimators to estimate the unknown pdf.

As we have observed in these systems, researchers use a superposition of families of approximations to the identity to estimate the unknown pdf, and especially Gaussian pdfs (Gaussian kernels). In addition, they can change this family at any approximated-estimated point; see Athanassoulis, Sapsis (2008), Sec. 8, 9. This becomes clearer as we consider some articles, such as Susarla, Walter (1981), *Estimation of a Multivariate Density Function Using Delta Sequences*, who deal with positive type delta sequences in section 2. Also, Nolan, Marron (1989), *Uniform consistency of automatic and location-adaptive delta-sequence estimators*, who were the first researchers to present locally adaptive estimators, and to prove some important results. During the last few years, scientists who deal with Large Sample Theory have developed these methods, and have published many results and insights. Specifically, Sain, Scott (1996), *On locally adaptive density estimation*, Sain (2002), *Multivariate locally adaptive density estimation*, Vidal-Sanz (2005), *Pointwise universal consistency of nonparametric density estimators*, present locally adaptive delta estimators, and show asymptotic theorems for unbiasedness and consistency. Also, there is applied research, such as, Brox et al. (2007), *Nonparametric Density Estimation with Adaptive, Anisotropic Kernels for Human Motion Tracking*, Eqs. 5 and 6 of Sec. 3 who deal with the human motion tracking using exactly those estimators, and Bengio et. al. (2006) *Non-Local Manifold Parzen Windows*

After having examined the results of aforementioned research, see Athanassoulis, Sapsis (2008), Sec. 11, and Brox et al. (2007), Figure 1 (pp. 156), and Sec. 4, we have concluded that the above approximation is effective. However, we did not find in the literature (as long as we have searched, and at any source we could) a rigorous mathematical construction which allows this particular approximation.

Our initial aim was exactly to cover this gap, although this good has not been fully achieved, and there is one question remaining unanswered. We want to construct a theory, and within its frame to show that any pdf can be approximated by a superposition of families of approximation to the identity (or delta families). The most important part for the application is that we can approximate any pdf by a superposition of Gaussian pdfs, and thus we can change this superposition at any approximated point. It is obvious that our aim is not to prove a general theorem, but we want to construct the mathematical tools, such that, we could solve/cover this particular problem/gap.

Our aim is not restricted in only the aforementioned one, but we, also, have a more general one. We would like to link our technique of approximation, with other fields of mathematics.

In particular, we try to highlight the similarities between the fields of mathematical analysis, estimation Statistics and Large Sample Theory. With the proofs, which we cite and our comments where we deem necessary, we, largely, succeed in it. The reader will understand that the particular convergence and the approximation (continuous and discrete), which we use, play a significant role in mathematics, and exactly with this, we can prove strong theorems, such as the inversion of Fourier transform.

Despite the essential objectives of this work there is also a personal objective, but an equally important one. Through this thesis, I intend to broaden my knowledge and to incorporate with new techniques and ideas. This means, I intend to learn how to handle a problem and to realize that mathematics are not divided into sectors, but they are a unified whole where each time we just use the tools that we need (or construct them) in order to solve a problem. We believe that this objective has been achieved to a great extent.

Finally, we tried our text to be as complete as we can. We cite the proofs and the definitions completely, and the preliminaries that we deem necessary for better understanding of concepts. In addition, we have tried to present/analyze recent papers, so readers could be informed with recent results.

1.2. Preview of chapters

In Chapter 2, we show the importance of the approximation method which we shall use, and the necessity of existence of a discrete approximation of a multidimensional pdf. In particular, in the first part of this chapter, we prove completely the Fourier inversion Theorem. This is an important theorem in mathematical analysis, which is based on the convergence which we shall describe. Continuing, in the second part of Ch. 2, we cite an application of a random dynamical system where we can realize that we need a discrete approximation form.

In the following chapter (Chapter 3) we present the main results of our work. In particular, we present a series of theorems establishing that any multidimensional pdf can be approximated by a superposition of Gaussian pdfs (or more general functions/kernels). In the simplest case, we can assume spherically symmetric kernels, and easily can be generalized to ellipsoidal ones in orientation of the usual basis of $\mathbb{R}^N$. In the most interesting case, which we study here, we can assume general ellipsoidal kernels in any orientation. Firstly, we prove that any pdf can "generate" a family of approximation to the identity (or delta family). Subsequently, we show the central theorem of this chapter, in which we prove an integral approximation of any pdf h, namely $\lim\limits_{\mathbf{G}_{\boldsymbol{\lambda}}(\boldsymbol{x}),\prec} \int_{\mathbb{R}^N} K_{\mathbf{G}_{\boldsymbol{\lambda}}(\boldsymbol{x})}(\boldsymbol{x}-\boldsymbol{u})h(\boldsymbol{u})d\boldsymbol{u} = h(\boldsymbol{x})$,

where $\mathbf{G}_{\boldsymbol{\lambda}}(\boldsymbol{x})$ is a family of matrices in an appropriate directed set. In the third part of this chapter, using the Riemann Sum, we give a discrete approximation form by a superposition of Gaussian pdfs, and we prove that we can change this superposition at any approximated point of support of pdf. It means, we can achieve better approximation with less points of the partition, i.e., we can decrease the computational cost. In other words, we prove that

$h(\boldsymbol{x}) \approx \sum_{\ell \in \mathcal{L}} p_\ell \, K\left(\boldsymbol{x} - \boldsymbol{u}_\ell \, ; \tilde{\mathbf{C}}(\boldsymbol{x}) \right)$, where $\tilde{\mathbf{C}}(\boldsymbol{x})$ is the localized covariance matrix, and p_ℓ are unknown constants. As we shall see, there are many common points with the Fourier inversion Theorem. The basic deference is that we require (it is a nature condition) positivity of family of approximation to the identity. This additional condition makes it different, and it is the reason we study the approximation of pdfs separately.

Last, in Chapter 4, we deal with estimation of pdfs. We define the delta sequences estimators, $\hat{f}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_m\left(\boldsymbol{x}, \boldsymbol{x}_i \right)$, and we present the most basic of the kernel estimators (which are delta ones). Then, we focus on theorems of asymptotic unbiasedness, ;i.e., $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) - f(\boldsymbol{x}) \xrightarrow{n \to \infty} 0$, and consistency of variation term, i.e., $\hat{f}_n(\boldsymbol{x}) - \mathrm{E}\left(\hat{f}(\boldsymbol{x}) \right) \xrightarrow{P} 0$ in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$. In last years, functional data have been developed in statistics. For this reason, we generalized the above definitions in infinite dimensional spaces, and we present some recent published results in Banach spaces.

At first glance, integral approximation, and estimation of a pdf are different methods. However, the fact that the expected value of a delta estimator can be written in the following form, $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) = \int_{\mathbb{R}^N} K_{m_n}(\boldsymbol{x}, \mathbf{z}) f(\mathbf{z}) \mu(dz)$, allow us to link these two methods. Studying the bias problem, we shall show that $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) \xrightarrow{n \to \infty} f(\boldsymbol{x})$, which is the same problem as the integral approximation of a pdf. On the other hand, if we assume all p_ℓ are equal, then the discrete form which we derive in Chapter 3, looks like with a delta estimator. Finally, as we shall notice, either we refer to the classical analysis, either to the approximation-estimation of pdfs, the properties of families of approximation to the identity play a key role on proving the theorems.

CHAPTER 2

The Importance of Integral Approximation and the Necessity for an Approximation of a Pdf

Table of Contents

2.1. Introduction

In this chapter we present a basic theorem in Harmonic analysis, and an application of a dynamical stochastic system. We have two main objectives. First, we shall show the importance of the convergence, which we shall deal with in the sequel. Second, we shall show the necessity of the approximation of a pdf.

In Section 2.2, we present a theorem from mathematical analysis, the Fourier inversion Theorem. In the proof of this theorem, we use the fact that the convolution of a function f with a family of functions $\left(K_\delta\right)_{\delta>0}$,with some properties (see Def. 1 below), convergences to the function (Theorem 1 below). We shall prove a similar theorem in Chapter 3. But there, we need a new property of these functions (kernels): to be positive.

Someone might ask why we need an approximation of a pdf. We answer this question in Section 2.3. We present an application where a discrete approximation of a pdf is necessary.

2.2. The Theorem of inverse Fourier transform

Definition 1: A family $\left(K_\delta\right)_{\delta>0}$ of functions on $\mathbb{R}^N$ is called ***an approximation to the identity*** if :

(a) For any $\delta>0$, $\int_{\mathbb{R}^N} K_\delta(\boldsymbol{y})\,d\boldsymbol{y}=1$.

(b) There exists a constant $M>0$ such that, for any $\delta>0$ and $\boldsymbol{y}\in\mathbb{R}^N$,

$$\left|K_\delta(\boldsymbol{y})\right|\le\frac{M}{\delta^N}.$$

(c) There exists a constant $M>0$ such that, for any $\delta>0$ and $\boldsymbol{y}\in\mathbb{R}^N$,

$$\left|K_\delta(\boldsymbol{y})\right|\le\frac{M\,\delta}{|\boldsymbol{y}|^{N+1}}. \qquad \blacksquare$$

Definition 2: Let f be a locally integrable function on $\mathbb{R}^N$ (i.e., $f\in L^1_{loc}\left(\mathbb{R}^N\right)$). The ***Lebesgue set***, $\mathrm{Leb}(f)$, of f is defined by

$$\mathrm{Leb}(f)=\left\{\boldsymbol{x}\in\mathbb{R}:\left|f(\boldsymbol{x})\right|<\infty \text{ and } \lim_{\substack{\lambda(B)\to 0\\ \boldsymbol{x}\in B}}\frac{1}{\lambda(B)}\int_B\left|f(\boldsymbol{y})-f(\boldsymbol{x})\right|d\boldsymbol{y}\right\}, \tag{1}$$

where λ denotes the Lebesgue measure, and B an open ball in $\mathbb{R}^N$. $\blacksquare$

Furthermore, it can be proved that $\lambda\left(\mathbb{R}^N \setminus \mathrm{Leb}(f)\right) = 0$ (Giannopoulos (2014), Harmonic Analysis (notes), Chapter 2, Lemma 2.3.5).

Definition 3: If $f \in L^1\left(\mathbb{R}^N\right)$, then its ***Fourier transform*** $\hat{f} : \mathbb{R}^N \to \mathbb{C}$ is defined by

$$\hat{f}(\boldsymbol{\xi}) = \int_{\mathbb{R}^N} f(\boldsymbol{x}) \exp\left(-2\pi i\, \boldsymbol{x} \cdot \boldsymbol{\xi}\right) d\boldsymbol{x} . \quad ■ \tag{2}$$

Lemma 1: Let $f \in L^1\left(\mathbb{R}^N\right)$, and $\boldsymbol{x} \in \mathrm{Leb}(f)$. Then, for any $\delta > 0$ we define the following function

$$\mathcal{A}(\delta) = \frac{1}{\delta^N} \int_{|\boldsymbol{y}| \le \delta} \left| f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x}) \right| d\boldsymbol{y} . \tag{3}$$

Then, the function $\mathcal{A}$ is bounded, continuous and $\lim_{\delta \to 0} \mathcal{A}(\delta) = 0$. ■

Proof: Giannopoulos (2014), Harmonic Analysis (notes), Chapter 2, Lemma 2.4.8. ◄

Theorem 1 (Giannopoulos (2014), Chapter 2, Theorem 2.4.7.)**:** Let $\left(K_\delta\right)_{\delta > 0}$ be an approximation to the identity. Then, for any $f \in L^1\left(\mathbb{R}^N\right)$ the following limiting relation holds true

$$\lim_{\delta \to 0} \int_{\mathbb{R}^N} f(\boldsymbol{y})\, K_\delta(\boldsymbol{x}-\boldsymbol{y})\, d\boldsymbol{y} = f(\boldsymbol{x}), \tag{4}$$

for any $\boldsymbol{x} \in \mathrm{Leb}(f)$, i.e., almost everywhere with respect to Lebesgue measure on $\mathbb{R}^N$. ■

Proof: Let $\delta > 0$. We set $J_k = \left\{ \boldsymbol{y} \in \mathbb{R}^N :\ 2^k \delta < |\boldsymbol{y}| \le 2^{k+1} \delta \right\}$, and we have

$$\begin{aligned}
\left| \int_{\mathbb{R}^N} f(\boldsymbol{y})\, K_\delta(\boldsymbol{x}-\boldsymbol{y})\, d\boldsymbol{y} - f(\boldsymbol{x}) \right| &= \left| \int_{\mathbb{R}^N} f(\boldsymbol{x}-\boldsymbol{y})\, K_\delta(\boldsymbol{y})\, d\boldsymbol{y} - f(\boldsymbol{x}) \right| \\
&\le \int_{\mathbb{R}^N} \left| f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x}) \right| \left| K_\delta(\boldsymbol{y}) \right| d\boldsymbol{y} \\
&\le \int_{|\boldsymbol{y}| \le \delta} \left| f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x}) \right| \left| K_\delta(\boldsymbol{y}) \right| d\boldsymbol{y} \\
&\quad + \sum_{k=0}^{\infty} \int_{J_k} \left| f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x}) \right| \left| K_\delta(\boldsymbol{y}) \right| d\boldsymbol{y}
\end{aligned}$$

$$\leq \frac{M}{\delta^N} \int_{|y| \leq \delta} |f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x})| |K_\delta(\boldsymbol{y})| d\boldsymbol{y}$$

$$+ \sum_{k=0}^{\infty} M\delta \int_{J_k} |f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x})| \frac{1}{|\boldsymbol{y}|^{N+1}} d\boldsymbol{y}$$

$$\leq M\mathcal{A}(\delta) + \sum_{k=0}^{\infty} \frac{M\delta}{(2^k\delta)^{N+1}} \int_{|y| \leq 2^{k+1}\delta} |f(\boldsymbol{x}-\boldsymbol{y}) - f(\boldsymbol{x})| d\boldsymbol{y}$$

$$= M\mathcal{A}(\delta) + \sum_{k=0}^{\infty} \frac{M\delta}{(2^k\delta)^{N+1}} (2^{k+1}\delta)^N \mathcal{A}(2^{k+1}\delta)$$

$$= M\mathcal{A}(\delta) + \sum_{k=0}^{\infty} \frac{2^N M}{2^k} \mathcal{A}(2^{k+1}\delta)$$

$$= M_1\left[\mathcal{A}(\delta) + \sum_{k=0}^{\infty} \frac{1}{2^k} \mathcal{A}(2^{k+1}\delta)\right],$$

by setting $M_1 = 2^N M$.

Now, let $\varepsilon > 0$. We know that $\sum_{k=0}^{\infty} \frac{1}{2^k} < \infty$ hence, we can find $n \in \mathbb{N}$, such that $\sum_{k=n}^{\infty} \frac{1}{2^k} < \varepsilon$.

Further, from the fact that $\lim_{\delta \to 0} \mathcal{A}(\delta) = 0$ we can find $\delta_0 > 0$, such that $\mathcal{A}(2^k\delta) < \frac{\varepsilon}{3}$ for any $\delta < \delta_0$ and $k = 0,1,...,n$. Also, we know that $\|\mathcal{A}\|_\infty < \infty$ (see Lemma 1).

Then, for any $\delta < \delta_0$

$$\left| \int_{\mathbb{R}^N} f(\boldsymbol{y}) K_\delta(\boldsymbol{x}-\boldsymbol{y}) d\boldsymbol{y} - f(\boldsymbol{x}) \right| \leq M_1\mathcal{A}(\delta) + M_1 \sum_{k=0}^{n-1} \frac{1}{2^k} \mathcal{A}(2^{k+1}\delta)$$

$$+ M_1 \sum_{k=n}^{\infty} \frac{1}{2^k} \mathcal{A}(2^{k+1}\delta)$$

$$\leq M_1 \left[\frac{\varepsilon}{3} + \frac{\varepsilon}{3} \sum_{k=0}^{n-1} \frac{1}{2^k} + \| \mathcal{A} \|_\infty \sum_{k=n}^{\infty} \frac{1}{2^k} \right]$$

$$\leq M_1 \left[\frac{\varepsilon}{3} + \frac{2\varepsilon}{3} + \| \mathcal{A} \|_\infty \, \varepsilon \right]$$

$$= M_1 \left[1 + \| \mathcal{A} \|_\infty \right] \varepsilon , \quad \text{for any } \varepsilon > 0 .$$

This is proves Eq. (4), $\lim_{\delta \to 0} \int_{\mathbb{R}^N} f(\boldsymbol{y}) \, K_\delta(\boldsymbol{x} - \boldsymbol{y}) \, d\boldsymbol{y} = f(\boldsymbol{x})$, and the proof is complete.◄

Lemma 2 (Multiplication formula, Giannopoulos (2014), Chapter 2, Theorem 2.4.7)**:** Let f, g be two integrable functions on $\mathbb{R}^N$. Then,

$$\int_{\mathbb{R}^N} \hat{f}(\xi) \, g(\xi) d\xi = \int_{\mathbb{R}^N} f(\boldsymbol{y}) \, \hat{g}(\boldsymbol{y}) d\boldsymbol{y} . \qquad \blacksquare \tag{5}$$

***Proof*:** We shall use the definition of Fourier transform (Def. 3) and the Fubini Theorem (see Billingsley (1995), Chapter 2, Sec. 18, Theorem 18.3)

$$\int_{\mathbb{R}^N} \hat{f}(\xi) \, g(\xi) d\xi = \int_{\mathbb{R}^N} \left(\int_{\mathbb{R}^N} f(\boldsymbol{y}) \exp(-2\pi i \, \boldsymbol{y} \cdot \xi) \, d\boldsymbol{y} \right) g(\xi) d\xi$$

$$= \int_{\mathbb{R}^N} \left(\int_{\mathbb{R}^N} g(\xi) \exp(-2\pi i \, \boldsymbol{y} \cdot \xi) \, d\xi \right) f(\boldsymbol{y}) d\boldsymbol{y}$$

$$= \int_{\mathbb{R}^N} f(\boldsymbol{y}) \, \hat{g}(\boldsymbol{y}) d\boldsymbol{y} . \qquad ◄$$

Lemma 3: Let $\boldsymbol{x} \in \mathbb{R}^N$ and $\delta > 0$.Then, the Fourier transform of function $g_\delta(\xi) = \exp\left(-\pi \, \delta |\xi|^2 \right) \exp(2\pi \, i \, \boldsymbol{x} \cdot \xi)$, $\xi \in \mathbb{R}^N$ given by:

$$\hat{g}_\delta(\boldsymbol{y}) = \frac{1}{\delta^{N/2}} \exp\left(-\frac{\pi}{\delta} |\boldsymbol{x} - \boldsymbol{y}|^2 \right). \qquad \blacksquare \tag{6}$$

***Proof*:** Giannopoulos (2014), Harmonic Analysis (notes), Chapter 3, Lemma 3.2.3. ◄

To prove the following theorem we need to define a particular family of kernels $\left(K_{\delta^2}\right)_{\delta>0}$ given by:

$$K_{\delta^2}(\boldsymbol{y}) = \frac{1}{\delta^N}\exp\left(-\frac{\pi\,|\boldsymbol{y}|^2}{\delta^2}\right), \quad \delta > 0 \text{ and } \boldsymbol{y}\in\mathbb{R}. \tag{7}$$

We shall prove that the above family is an approximation to the identity.

Theorem 2 (Inverse Fourier transform, Giannopoulos (2014), Chapter 3, Theorem 3.2.1)**:** Let the family $\left(K_{\delta^2}\right)_{\delta>0}$ be defined in terms of Eq. (7). If $f\in L^1\left(\mathbb{R}^N\right)$, and $\hat{f}\in L^1\left(\mathbb{R}^N\right)$ then, the following relation holds true:

$$f(\boldsymbol{x}) = \lim_{\delta\to 0}\int_{\mathbb{R}^N} f(\boldsymbol{y})\,K_{\delta^2}(\boldsymbol{x}-\boldsymbol{y})\,d\boldsymbol{y} = \int_{\mathbb{R}^N}\hat{f}(\boldsymbol{\xi})\exp(2\pi\, i\,\boldsymbol{x}\cdot\boldsymbol{\xi})\,d\boldsymbol{\xi}, \tag{8}$$

almost everywhere with respect to Lebesgue measure on $\mathbb{R}^N$. ■

Proof**:** First, we prove that $\left(K_{\delta^2}\right)_{\delta>0}$ is an approximation to the identity. For any $\delta>0$, by changing of variable $\boldsymbol{y}=\delta\,\mathbf{z}$ we derive

$$\int_{\mathbb{R}^N} K_{\delta^2}(\boldsymbol{y})\,d\boldsymbol{y} = \frac{1}{\delta^N}\int_{\mathbb{R}^N}\exp\left(-\frac{\pi\,|\boldsymbol{y}|^2}{\delta^2}\right)d\boldsymbol{y} = \int_{\mathbb{R}^N}\exp\left(-\pi\,|\mathbf{z}|^2\right)d\mathbf{z} = 1. \tag{a}$$

Also, for any $\delta>0$ and $\boldsymbol{y}\in\mathbb{R}^N$ we have

$$0\le K_{\delta^2}(\boldsymbol{y}) = \frac{1}{\delta^N}\exp\left(-\frac{\pi\,|\boldsymbol{y}|^2}{\delta^2}\right)\le\frac{1}{\delta^N}. \tag{b}$$

Last, using the known inequality $\exp(t)\ge\dfrac{t^{N+1}}{(N+1)!}$, $t\ge 0$, at $t=\dfrac{\sqrt{\pi}\,|\boldsymbol{y}|}{\delta}$, we obtain

$$0\le K_{\delta^2}(\boldsymbol{y}) = \frac{1}{\delta^N}\exp\left(-\frac{\pi\,|\boldsymbol{y}|^2}{\delta^2}\right)\le\frac{1}{\delta^N}\,\frac{(N+1)!\,\delta^{N+1}}{\pi^{(N+1)/2}\,|\boldsymbol{y}|^{N+1}} = \frac{M\,\delta}{|\boldsymbol{y}|^{N+1}}. \tag{c}$$

By setting $M=\dfrac{(N+1)!}{\pi^{(N+1)/2}}$ .So, we have proved the three properties of definition 1.

Now, let $\boldsymbol{x}\in\mathbb{R}^N$ and g_{δ^2} be the function of Lemma 3. We know that

$$\hat{g}_{\delta^2}(\boldsymbol{y}) = \frac{1}{\delta^N}\exp\left(-\frac{\pi}{\delta^2}\,|\boldsymbol{x}-\boldsymbol{y}|^2\right) \tag{d}$$

Using Lemma 2, for any $\delta>0$, we obtain

$$\int_{\mathbb{R}^N} \hat{f}(\xi)\exp\left(-\pi\,\delta^2|\xi|^2\right)\exp(2\pi\, i\,\boldsymbol{x}\cdot\xi)\,d\xi = \int_{\mathbb{R}^N} f(\boldsymbol{y})\,K_{\delta^2}(\boldsymbol{x}-\boldsymbol{y})\,d\boldsymbol{y}. \qquad \text{(e)}$$

In the right term of above equation, K_{δ^2} is an approximation to the identity, so by theorem 1 we have

$$\lim_{\delta\to 0}\int_{\mathbb{R}^N} f(\boldsymbol{y})\,K_{\delta^2}(\boldsymbol{x}-\boldsymbol{y})\,d\boldsymbol{y} = f(\boldsymbol{x}) \text{ almost everywhere.} \qquad \text{(f)}$$

For the left term, we use the dominated convergence Theorem (see Billingsley (1995), Chapter 2, Sec. 16, Theorem 16.4)

$$\lim_{\delta\to 0}\int_{\mathbb{R}^N} \hat{f}(\xi)\exp\left(-\pi\,\delta^2|\xi|^2\right)\exp(2\pi\, i\,\boldsymbol{x}\cdot\xi)\,d\xi = \int_{\mathbb{R}^N} \hat{f}(\xi)\exp(2\pi\, i\,\boldsymbol{x}\cdot\xi)\,d\xi. \qquad \text{(g)}$$

Finally, combining Eqs. (f) and (g) we obtain the equation (8)

$$f(\boldsymbol{x}) = \lim_{\delta\to 0}\int_{\mathbb{R}^N} f(\boldsymbol{y})\,K_{\delta^2}(\boldsymbol{x}-\boldsymbol{y})\,d\boldsymbol{y} = \int_{\mathbb{R}^N} \hat{f}(\xi)\exp(2\pi\, i\,\boldsymbol{x}\cdot\xi)\,d\xi. \qquad \blacktriangleleft$$

Stein, Shakarchi (2003), in Chapter 5, Theorem 1.9, prove the same theorem on a Schartz space.

2.3. An example in which the approximation of pdf is necessary

In the sequel of this work we focus on approximation of multidimensional pdfs, which is the basic result of this work which is analyzed in Chapter 3. In this section, we present a problem which shows the necessity of a discrete approximation of any pdf.

We start with some basic definitions

Definition 4: Let Ω be a sample space, and $\mathbb{T} = \left[\, t_0, \mathrm{T} \,\right]$ be a real interval. A generic ***random (or stochastic) differential equation*** (RDE) given by

$$\frac{dx(t;\omega)}{dt} = G(x(t;\omega), y(t;\omega), t) \qquad \blacksquare \qquad (9)$$

Where, ω is the sample argument, $t \in \mathbb{T}$, and G is a continuous function which is nonlinear generally. Thus, $y(t;\omega)$ is a known random function which is called the excitation of the system and $x(t;\omega)$ is called the response of the system. ■

Definition 5: Let $\mathscr{X}$ be a separable Banach space, and $\mathrm{P} = \mathrm{P}_{\mathscr{X}}$ be a probability measure defined on it. The ***characteristic functional*** $\mathscr{F}$ of P is a cylinder functional defined on the dual space $\mathscr{U} = \mathscr{X}'$ by the formula:

$$\mathscr{F}(u) = \mathrm{E}^{\omega}\left(\exp\left(i \langle u , x(\omega) \rangle\right)\right) = \int_{\mathscr{X}} \exp(i \langle u , x \rangle) \, \mathrm{P}(dx), \qquad u \in \mathscr{U} . \tag{10}$$

Notice that this is the Fourier transform of the induced probability measure $\mathrm{P}_{\mathscr{X}}$. Also, this integral always exists provided that the corresponding probability measure is well defined. See Vakhania et al. (1987) and Pugachev, Sinitsyn (2001) for its properties. ■

The concept of the characteristic functional was introduced by Kolmogorov in 1935, by means of an amazing and far ahead of its time, two-page article, in Comptes Rendu de l' Académie des Science de Paris.

Furthermore, we shall need the definitions of functional derivatives. Here, we give only the definitions. For more details and properties of them see Athanassoulis (2011), Functional analysis (notes).

Definition 6: Let $\left(\mathrm{X} , \| \ \|_{\mathrm{X}} \right)$, $\left(\mathrm{Y} , \| \ \|_{\mathrm{Y}} \right)$ be two normed spaces, and consider an operator, $F: \mathrm{X} \to \mathrm{Y}$. We shall say that F is ***Frechet differentiable*** at $x_0 \in \mathrm{X}$ if there exists a continuous and linear operator $L: \mathrm{X} \to \mathrm{Y}$, such that

$$\lim_{\substack{x \to x_0 \\ x \neq x_0}} \frac{\left\| F(x) - F(x_0) - L(x - x_0) \right\|_{\mathrm{Y}}}{\left\| x - x_0 \right\|_{\mathrm{X}}} = 0 \tag{11}$$

The operator L is called ***Frechet derivative*** of F at point $x_0 \in \mathrm{X}$, and it will be denoted by $DF(x_0)$ (or $F'(x_0)$). ■

Definition 7: Let X, Y be linear topological spaces, and consider an operator, $F: \mathrm{X} \to \mathrm{Y}$. Also, assume $h \in \mathrm{X}$, and $h \neq 0_{\mathrm{X}}$. We shall say that F has ***Gateaux derivative*** at $x_0 \in \mathrm{X}$, and it will be denoted by $\delta F(x_0 ; h)$, if there exists the following limit

$$\delta F(x_0 ; h) = \left. \frac{d F(x_0 + t h)}{d t} \right|_{t=0} = \lim_{t \to \infty} \frac{F(x_0 + t h) - F(x_0)}{t}, \tag{12}$$

and if $h = 0_{\mathrm{X}}$ then we shall define $\delta F(x_0 ; 0_{\mathrm{X}}) = 0$. ■

Definition 8: Let X be a linear function space, which is equipped with a metric, and consider an operator, $F: \mathrm{X} \to \mathbb{R}$ (or $\mathbb{C}$). We shall say that F has ***Volterra derivative*** at $x(\cdot) \in \mathrm{X}$, and it will be denoted by $F'(x(\cdot))$, if there exists the following limit

$$\lim_{\gamma} \frac{F\left(x(\cdot) + h_{\gamma}^{\xi}(\cdot) \right) - F\left(x(\cdot) \right)}{\int_{D_{\gamma}^{(\xi)}} h_{\gamma}^{\xi}(t)\,dt} \quad \text{as } \gamma \text{ increases,} \tag{13}$$

where $D_{\gamma}^{(\xi)}$ is the support of function h_{γ}^{ξ}.

See Athanassoulis (2011), Functionals and functional differentiation (notes), for more details and clarifications. ■

Now, assume a dynamical system which is described by the following RDE and the initial condition:

$$\frac{dx(t;\omega)}{dt} + \mu x(t;\omega) + k x^3(t;\omega) = y(t;\omega), \tag{14a}$$

$$x(t_0;\omega) = x_0(\omega), \tag{14b}$$

Where μ, k are deterministic constants, $x_0(\omega)$ is a random variable with known characteristic function $\varphi_0(u),\ u \in \mathbb{R}$. The excitation is a real-valued random function with sample space a separable Banach space $\mathscr{Y}$, probability measure P_y, and known characteristic functional $\mathscr{F}_y(v)$, $v \in \mathscr{V} = \mathscr{Y}'$. We denote $\mathscr{X}$ the sample space of random function $x(t;\omega)$, its probability measure P_x, and the dual space of by $\mathscr{U} = \mathscr{X}'$. In this problem, we assume that $\mathscr{Y} = C^k(\mathrm{I}),\ \mathrm{I} \subseteq \mathbb{R}$ for some $k \in \mathbb{N} \cup \{0\}$ and $\mathscr{X} = C^{k+1}(\mathrm{I})$. Last, we assume that the above probability measures and the joint one, P_{xy}, are well defined. See Skorokhod (1969, 2005), Chapter 2 and Spiliotis (notes) (2012), for more details about probability measures in infinite dimensional spaces.

Our aim is to derive a new equation for the corresponding pdfs. This has first introduced by Athanassoulis, Sapsis (2008). In Sections 4 and 5, someone can study the whole proof. Here, we shall give a description of this work.

The joint response-excitation characteristic functional is given by

$$\mathscr{F}_{xy}(u,v) = \int_{\mathscr{X}\times\mathscr{Y}} \exp\left(i\left(\langle u, x\rangle + \langle v, y\rangle\right)\right) \mathrm{P}_{xy}(dx,dy), \quad u \in \mathscr{U}, \quad v \in \mathscr{V}, \tag{15}$$

Now, let us consider the Volterra u – partial derivative of $\mathscr{F}_{xy}$ at time t

$$\frac{\delta \mathscr{F}_{xy}(u,v)}{\delta u(t)} = \int_{\mathscr{X}\times\mathscr{Y}} i\, x(t)\exp\big(i\,(\langle u\,,x\rangle + \langle v\,,y\rangle)\big)\,\mathrm{P}_{xy}(dx,dy)\,, \tag{16}$$

and we differentiate with respect to t, obtaining

$$\frac{d}{dt}\frac{\delta \mathscr{F}_{xy}(u,v)}{\delta u(t)} = \int_{\mathscr{X}\times\mathscr{Y}} i\, x'(t)\exp\big(i\,(\langle u\,,x\rangle + \langle v\,,y\rangle)\big)\,\mathrm{P}_{xy}(dx,dy)\,. \tag{17}$$

Similarly,

$$\frac{\delta \mathscr{F}_{xy}(u,v)}{\delta v(t)} = \int_{\mathscr{X}\times\mathscr{Y}} i\, y(t)\exp\big(i\,(\langle u\,,x\rangle + \langle v\,,y\rangle)\big)\,\mathrm{P}_{xy}(dx,dy)\,. \tag{18}$$

We computing the three-fold Volterra $u-$partial derivative of $\mathscr{F}_{xy}$ at time instants $t_1,t_2,t_3 \in \mathrm{I}$,and then we set $t_1 = t_2 = t_3 = t$

$$\frac{\delta^3 \mathscr{F}_{xy}(u,v)}{\delta u(t_1)\delta u(t_2)\delta u(t_3)} = \int_{\mathscr{X}\times\mathscr{Y}} i\, x(t_1)x(t_1)x(t_3)\exp\big(i\,(\langle u\,,x\rangle + \langle v\,,y\rangle)\big)\,\mathrm{P}_{xy}(dx,dy)\,. \tag{19}$$

Now, combining Eqs. (16 – 19) we obtain the following differential equation for the characteristic functional

$$\frac{d}{dt}\frac{\delta \mathscr{F}_{xy}(u,v)}{\delta u(t)} + \mu\frac{\delta \mathscr{F}_{xy}(u,v)}{\delta u(t)} - k\,\mu\frac{\delta^3 \mathscr{F}_{xy}(u,v)}{\delta u(t)^3} = \frac{\delta \mathscr{F}_{xy}(u,v)}{\delta v(t)}\,, \tag{20a}$$

with initial condition

$$\mathscr{F}_{xy}\big(\upsilon\delta\ \cdot - t_0\ \ ,0\big) = \varphi_0(\upsilon),\ \ \upsilon\in\mathbb{R}\,, \tag{20b}$$

where δ denotes the Dirac delta generalized function.

Continuing, we apply Eq. (20a), to the pair

$$u = \upsilon\cdot\delta\big(\cdot - t\big),\ v = \nu\cdot\delta\big(\cdot - s\big) \text{ for fixed } t,s\,. \tag{21}$$

For the first term of Eq. (20a) we have

$$\frac{d}{dt}\frac{\delta \mathscr{F}_{xy}(u,v)}{\delta u(t)}\Bigg|_{\substack{u=\upsilon\cdot\delta(\cdot - t)\\ v=\nu\cdot\delta(\cdot - s)}} =$$

$$= \int_{\mathscr{X}\times\mathscr{Y}} i\, x'(t) \exp(i\,\upsilon\, x(t) + i\,\nu\, y(s))\, \mathrm{P}_{xy}(dx, dy) \tag{22a}$$

$$= \frac{1}{\upsilon}\frac{\partial}{\partial t} \int_{\mathscr{X}\times\mathscr{Y}} \exp(i\,\upsilon\, x(t) + i\,\nu\, y(s))\, \mathrm{P}_{xy}(dx, dy) \tag{22b}$$

$$= \frac{1}{\upsilon}\frac{\partial}{\partial t} \int_{\mathbb{R}}\int_{\mathbb{R}} \exp(i\,\upsilon\, x + i\,\nu\, y)\; f_{x(t)y(s)}(x, y)\, dx\, dy \tag{22c}$$

$$= \frac{1}{\upsilon}\frac{\partial \varphi_{x(t)y(s)}(\upsilon, \nu)}{\partial t}. \tag{22d}$$

And taking the limit $s \to t$, we conclude to

$$\lim_{s \to t} \frac{d}{dt} \frac{\delta \mathscr{F}_{xy}(u, v)}{\delta u(t)} \Bigg|_{\substack{u = \upsilon\cdot\delta(\cdot - t) \\ v = \nu\cdot\delta(\cdot - s)}} = \frac{1}{\upsilon}\frac{\partial \varphi_{x(t)y(s)}(\upsilon, \nu)}{\partial t}\Bigg|_{s=t}. \tag{23a}$$

In Eq. (22c), we applied the projection Theorem. Athanassoulis, Sapsis (2008) give an analytic discussion for this theorem in Section 2. They give the general form in Eqs. (2.5) and (2.6).

With similar way, we obtain

$$\frac{\delta \mathscr{F}_{xy}(\upsilon\cdot\delta(\cdot - t), \nu\cdot\delta(\cdot - t))}{\delta u(t)} = \frac{\partial \varphi_{x(t)y(s)}(\upsilon, \nu)}{\partial \upsilon}. \tag{23b}$$

$$\frac{\delta^3 \mathscr{F}_{xy}(\upsilon\cdot\delta(\cdot - t), \nu\cdot\delta(\cdot - t))}{\delta u(t)^3} = \frac{\partial^3 \varphi_{x(t)y(s)}(\upsilon, \nu)}{\partial \upsilon^3}. \tag{23c}$$

$$\frac{\delta \mathscr{F}_{xy}(\upsilon\cdot\delta(\cdot - t), \nu\cdot\delta(\cdot - t))}{\delta v(t)} = \frac{\partial \varphi_{x(t)y(s)}(\upsilon, \nu)}{\partial \nu}. \tag{23d}$$

Combining Eqs. (23a-d) and (20a,b), we derive

$$\frac{1}{u}\frac{\partial \varphi_{x(t)y(s)}(u, v)}{\partial t}\Bigg|_{s=t} + \mu \frac{\partial \varphi_{x(t)y(\mathrm{t})}(u, v)}{\partial u} - k \frac{\partial^3 \varphi_{x(t)y(\mathrm{t})}(u, v)}{\partial u^3} = \frac{\partial \varphi_{x(t)y(\mathrm{t})}(u, v)}{\partial v},$$

$$\varphi_{x(t_0)y(\mathrm{t}_0)}(\upsilon, 0) = \varphi_{x(t_0)}(\upsilon) = \varphi_0(\upsilon),\ \upsilon \in \mathbb{R}. \tag{24a,b}$$

Finally, implying the inverse Fourier transform, as we see in previous section, to (24a, b) we derive

$$\frac{\partial f_{x(t)y(s)}(a,\mathrm{b})}{\partial t}\bigg|_{s=t} + \frac{\partial}{\partial a}\Big[\left(\mu a + k a^3\right) f_{x(t)y(s)}(a,\mathrm{b})\Big] + \frac{\partial}{\partial a}\Big[b\, f_{x(t)y(s)}(a,\mathrm{b})\Big] = 0,$$

$$\int_{\mathbb{R}} f_{x(t_0)y(t_0)}(a,\mathrm{b})\, d\,\mathrm{b} = f_{x(t_0)}(a), \quad a \in \mathbb{R}. \tag{25a,b}$$

Further, the marginal compatibility condition

$$\int_{\mathbb{R}} f_{x(t)y(t)}(a,\mathrm{b})\, d\,a = f_{y(t)}(\mathrm{b}), \quad \mathrm{b} \in \mathbb{R} \;, t \geq t_0, \tag{25c}$$

and the constitutive conditions

$$\int_{\mathbb{R}\times\mathbb{R}} f_{x(t)y(s)}(a,\mathrm{b})\, d\,a\; d\,\mathrm{b} = 1, \qquad t,s \geq t_0, \tag{25d}$$

$$f_{x(t)y(s)}(a,\mathrm{b}) \geq 0 \;\text{ for any }\; a,\mathrm{b} \in \mathbb{R} \;\text{ and }\; t,s \geq t_0. \tag{25e}$$

Venturi et al. (2011), in Section 2, conclude to the same equation (25a,b) by using the following functional integral representation of a pdf

$$f_{x(t)y(t)}(a,\mathrm{b}) = \mathrm{E}^{\omega}\left(\delta(\alpha - x(t;\omega))\,\delta(\mathrm{b} - y(s;\omega))\right)$$

$$= \int_{\mathscr{X}\times\mathscr{Y}} \delta(\alpha - x(t;\omega))\,\delta(\mathrm{b} - y(s;\omega))\,\mathrm{P}_{xy}(dx,dy).$$

To solve the system (25a-e) it is obvious that we need a discrete approximation form of the joint response-excitation pdf. Then, we can apply a numerical method. Athanassoulis, Sapsis (2008), in Sections 8, 9 discuss about this approximation pdf. They approximate the particular pdf by a superposition with Gaussian pdfs (kernels). The following chapter gives a rigorous mathematical construction for this approximation, in particular, we shall prove that any pdf can be approximated by a Gaussian (or more general) pdfs different at any approximated point.

2.4. References

Athanassoulis, G.A., 2011, Functional analysis (notes).

Athanassoulis G.A., 2011, Functionals and functional differentiation (notes).

Athanassoulis, G.A., Sapsis, T.P., 2008, New partial differential equations governing the joint, response–excitation, probability distributions of nonlinear systems, under general stochastic excitation. Probabilistic Engineering Mechanics, 23, pp. 289–306.

Billingsley, P., 1995, Probability and measure, edition 3. John Wiley & Sons, Inc., New York.

Giannopoulos, A., 2014, Harmonic Analysis (notes)

Pugachev, V.V.S., Sinitsyn, I.I.N., 2001. Stochastic Systems: Theory and Applications, World Scientific.

Skorokhod, A.V., Gichman, I.I, 1969, Introduction to the theory of random processes. W.B Saunders Company.

Skorokhod A.V., 2005, Basic principles and applications of probability theory. Springer.

Spiliotis, I., 2002, Probability measures in infinite dimensional spaces (notes).

Vakhania, N.N., Tarieladze, V.I. & Chobanyan, S.A., 1987. Probability Distributions on Banach spaces, Dordrecht: D.Reidel Publ. Co.

Venturi, D., et al., 2012, A computable evolution equation for the joint response-excitation probability density function of stochastic dynamical systems. Proceedings of the Royal Society A: Mathematical, Physical and Engineering Sciences, 468(2139), pp.759–783.

CHAPTER 3

Integral Approximations of Multidimensional Pdfs

Table of Contents

3.1. Introduction

In this chapter we present the main results of our work. We deal with a method of approximation theory, the integral approximation of multidimensional probability density functions. In particular, we prove that any probability density function (pdf) can be approximated by superposition of a family of approximation to the identity (or delta family).

As we have noticed in few applied articles, researchers conclude in equations with unknowns pdfs. To find them, they approximate pdfs by a superposition of Gaussian pdfs (kernels), and then, apply numerical methods. In addition, they can change this superposition at any approximated point. In practice, this works effectively (see Brox (2007), Sec. 2, who deals with the human motion tracking and Athanassoulis, Sapsis (2008), Sec 8, 9, who deal with random dynamical systems), but there is not a rigorous and strict mathematical theory in the literature, which allows this.

Our aim is exactly this. We want to conclude to a discrete approximation form of a multidimensional pdf, which is useful in applied works but it is equally useful and interesting as a mathematical construction. To prove this, we construct a particular space, in which we prove some helpful lemmata, and then, the main theorem (Theorem2, integral approximation). Finally, in the last part of this chapter, we derive the discrete approximation. First of all, we try to give the geometric notion of the functions which we use. Thus, we prove that we can use a different delta family at any approximated point, which is the most important, and new result in integral approximation of pdfs.

Specifically, in Sections 3.2 and .3.3, we provide some basic results from analysis and algebra, and introduce the definition of delta family giving a short discussion, correspondingly. Then, we prove the main result of this chapter and work (Section 3.4), that is any pdf generates a delta family. Finally in Section 3.5, we show that any probability density function can be approximated by a superposition of Gaussian pdfs, and we can change it at any approximated point. Last, we give some properties for the corresponding covariance matrix, and correlations coefficients.

3.2. Preliminaries

In this section, we present some basic results from linear algebra, and real analysis which, we shall use in the sequel.

Definition 1: Let $\mathbf{A}$ be a $\left(N \times N \right)$ real matrix (i.e., $\mathbf{A} \in \mathbb{R}^{N \times N}$). $\mathbf{A}$ is called ***diagonalizable*** if it is similar to a diagonal matrix, i.e.,, $\mathbf{\Lambda} = \mathbf{Q}^{-1} \cdot \mathbf{A} \cdot \mathbf{Q}$. Where, $\mathbf{\Lambda} = \mathrm{diag}\left(\lambda_1, \lambda_2, \ldots, \lambda_N \right)$ is the diagonal matrix of eigenvalues of $\mathbf{A}$, and, the nonsingular matrix $\mathbf{Q}$ is the matrix of eigenvectors of $\mathbf{A}$. ■

Definition 2: We say that matrix, $\mathbf{A} \in \mathbb{R}^{N \times N}$ is ***positive definite*** and we write $\mathbf{A} > 0$ if

$$\boldsymbol{x}^T \cdot \mathbf{A} \cdot \boldsymbol{x} = \sum_{i=1}^{N} \sum_{j=1}^{N} \mathrm{A}_{ij}\, x_i\, x_j > 0 \quad \forall\ \boldsymbol{x} \in \mathbb{R}^N \setminus \{\mathbf{0}\} \qquad ■$$

In the following lemma, we collect and prove some auxiliary matrix-theoretic results, which we shall use in proofs of Lemmata 3 and 4.

Lemma 1: Let $\mathbf{A} \in \mathbb{R}^{N \times N}$ be a symmetric (i.e., $\mathbf{A}^T = \mathbf{A}$), and positive definite matrix. Then,

(a) $\mathbf{A}$ is diagonalizable, i.e., $\mathbf{\Lambda} = \mathbf{Q}^{-1} \cdot \mathbf{A} \cdot \mathbf{Q}$ and eigenvectors of $\mathbf{A}$, $\boldsymbol{q}_i, 1 \le i \le N$, can be chosen to form an orthonormal basis on $\mathbb{R}^N$, i.e., $\mathbf{Q}^{-1} = \mathbf{Q}^T$.

(b) The eigenvalues of $\mathbf{A}$ are real and positive (i.e., $\lambda_i > 0$, $1 \le i \le N$).

(c) The quadratic region $\mathrm{E}_{\boldsymbol{A}} = \left\{ \boldsymbol{x} \in \mathbb{R}^N : \boldsymbol{x}^T \cdot \boldsymbol{A} \cdot \boldsymbol{x} \le 1 \right\}$ is a real ellipsoidal in $\mathbb{R}^N$, centered at $\mathbf{0}$ with lengths of semiaxes $s_i = \dfrac{1}{\sqrt{\lambda_i}}$. The eigenvalues determine the lengths of semiaxes, and the eigenvectors (i.e., the columns, $\boldsymbol{q}_i$, $1 \le i \le N$, of matrix $\mathbf{Q}$) determine the directions of semiaxes. ■

***Proof*: (a)** Βάρσος,Δεριζιώτης,Εμμανουήλ,Μαλιάκας,Μελάς,Τατέλη, (2009). “Μια εισαγωγή στη γραμμική άλγεβρα” Chapter 4, Theorem 4.4.2.

(b) First, we show that $\lambda_i, 1 \le i \le N$ are real numbers. By definition of eigenvalues we have $\mathbf{A} \cdot \boldsymbol{q}_i = \lambda_i \boldsymbol{q}_i$, $1 \le i \le N$ and $\boldsymbol{q}_i \in \mathbb{C}^N - \left\{ \mathbf{0} \right\}$. Then,

$$\bar{\boldsymbol{q}}_i^{\,T} \cdot \mathbf{A} \cdot \boldsymbol{q}_i = \bar{\boldsymbol{q}}_i^{\,T} \cdot \left(\mathbf{A} \cdot \boldsymbol{q}_i \right) = \lambda_i \bar{\boldsymbol{q}}_i^{\,T} \cdot \boldsymbol{q}_i = \lambda_i \sum_{j=1}^{N} \left| \boldsymbol{q}_{ij} \right|^2$$

Also, we can obtain

$$\bar{\boldsymbol{q}}_i^{\,T} \cdot \mathbf{A} \cdot \boldsymbol{q}_i = \overline{\left(\mathbf{A} \cdot \boldsymbol{q}_i \right)}^{\,T} \cdot \boldsymbol{q}_i = \overline{\left(\lambda_i \, \boldsymbol{q}_i \right)}^{\,T} \cdot \boldsymbol{q}_i = \bar{\lambda}_i \sum_{j=1}^{N} \left| \, \boldsymbol{q}_{i\,j} \, \right|^2$$

So, we have $\bar{\lambda}_i = \lambda_i \;\Rightarrow\; \lambda_i \in \mathbb{R}$ and $\boldsymbol{q}_i \in \mathbb{R}^N - \left\{ \mathbf{0} \right\}$, $1 \le i \le N$. Thus, the fact that $\mathbf{A}$ is positive definite implies $\boldsymbol{q}_i^T \cdot \mathbf{A} \cdot \boldsymbol{q} > 0 \Rightarrow \lambda_i \sum_{j=1}^{N} \left| \, \boldsymbol{q}_{i\,j} \, \right|^2 > 0 \Rightarrow \lambda_i > 0, \; 1 \le i \le N$.

(c) $\boldsymbol{x}^T \cdot \mathbf{A} \cdot \boldsymbol{x} \le 1 \Leftrightarrow \left(\mathbf{Q}^T \cdot \boldsymbol{x} \right)^T \cdot \mathbf{\Lambda} \cdot \mathbf{Q}^T \cdot \boldsymbol{x} \le 1 \Leftrightarrow \boldsymbol{y}^T \cdot \mathbf{\Lambda} \cdot \boldsymbol{y} \le 1 \Leftrightarrow \sum_{i=1}^{N} \left(\dfrac{y_i}{\sqrt{1/\lambda_i}} \right)^2 \le 1$

by setting $\boldsymbol{y} = \mathbf{Q}^T \cdot \boldsymbol{x}$, this is., $\boldsymbol{y} \in \mathcal{E}_{\mathbf{Q}} \left(\mathbf{0} \, ; \dfrac{1}{\sqrt{\lambda_1}}, \dfrac{1}{\sqrt{\lambda_2}}, \ldots, \dfrac{1}{\sqrt{\lambda_N}} \right)$.

See Sharipov (1996), Linear algebra and multidimensional geometry, Chapter IV, for more details about quadratic forms. ◄

The formula of change of variables in multidimensional integrals (for invertible transformations) will be the basic technical tool in what follows. It is recalled by the following.

Theorem 1 (Change of variables in multidimensional integrals): Let D and Ω be domains in $\mathbb{R}^N$, $\boldsymbol{x} \in \mathrm{D}$, $\boldsymbol{u} \in \Omega$, and $\boldsymbol{x} = \boldsymbol{T}(\boldsymbol{u})$ a transformation having Ω as its domain of definition and D as its range. The transformation $\boldsymbol{T}$ is assumed to be

- One-to one, and thus ***invertible***,
- C^1 in Ω,
- with ***non-zero*** (Frechet) ***derivative***

$$D\boldsymbol{T}(\boldsymbol{u}) \equiv \frac{D\boldsymbol{T}(\boldsymbol{u})}{D\boldsymbol{u}} \equiv \frac{D\boldsymbol{x}}{D\boldsymbol{u}} = \begin{pmatrix} \dfrac{\partial x_1}{\partial u_1} & \cdots & \dfrac{\partial x_1}{\partial u_N} \\ \vdots & \ddots & \vdots \\ \dfrac{\partial x_N}{\partial u_1} & \cdots & \dfrac{\partial x_N}{\partial u_N} \end{pmatrix}.$$

Then, for every function f, integrable in D, we have

$$\int_{\mathrm{D}} f(\boldsymbol{x}) \, d\boldsymbol{x} = \int_{\Omega = \boldsymbol{T}^{-1}(D)} f(\boldsymbol{T}(\boldsymbol{u})) \left| \det \left(\frac{D\boldsymbol{x}}{D\boldsymbol{u}} \right) \right| d\boldsymbol{u},$$

where $\det \left(\dfrac{D\boldsymbol{x}}{D\boldsymbol{u}} \right)$ is the Jacobian of the derivative of $\boldsymbol{T}$. ■

In addition, we shall give some definitions about generalized convergence; see McShane (1955), Section 3.

Definition 3: Let Λ be a nonempty set and $\prec$ be a binary relation with properties

D.1 $(\forall \lambda \in \Lambda)\ \lambda \prec \lambda$ (reflexivity)

D.2 $(\forall\ \lambda, \mu, \nu \in \Lambda)\ \lambda \prec \mu$ and $\mu \prec \nu \ \Rightarrow\ \lambda \prec \nu$ (transitivity)

D.3 $(\forall \lambda_1, \lambda_2 \in \Lambda)(\exists \lambda_3 \in \Lambda)\ \lambda_1 \prec \lambda_3$ and $\lambda_2 \prec \lambda_3$ (every pair λ_1, λ_2 in Λ has supremum)

The relation $\prec$ is called a ***direction*** in Λ and the pair $(\Lambda, \prec)$ is called a ***directed set***. ■

Definition 4: Let $(\Lambda, \prec)$ a directed set, and X a general set (e.g., topological/metric space). Every function $x: \Lambda \to X$ is called ***net*** in X and we write $(x_\lambda)_{\lambda \in \Lambda}$. ■

A net is the generalization of a sequence. Every sequence is a net with the directed set $(\mathbb{N}, \leq)$. ■

Definition 5: Let X be a metric (topological) space, $(\Lambda, \prec)$ a directed set, $(x_\lambda)_{\lambda \in \Lambda}$ a net in X, and $x \in X$. We say that the $(x_\lambda)_{\lambda \in \Lambda}$ **converges to** x (in the direction $\prec$) and we write $x_\lambda \xrightarrow{\prec} x$ (or $\lim\limits_{\lambda,\prec} x_\lambda = x$) if for every region $\mathrm{U} = \mathrm{U}(x)$ of x, there is $\lambda_0 = \lambda_0(\mathrm{U}) \in \Lambda$: $x_\lambda \in \mathrm{U}(x)$ for every $(\forall\ \lambda \in \Lambda)$: $\lambda_0 \prec \lambda$. ■

Notational conventions: To make the notation more compact (in order to save space in writing equations and proofs) the index vector $(\lambda_1, \lambda_2, \ldots, \lambda_N)$ will be usually denoted by $\boldsymbol{\lambda}$, and the scaled vector $\left(\frac{\delta}{\lambda_1}, \frac{\delta}{\lambda_2}, \ldots, \frac{\delta}{\lambda_N}\right)$ will be symbolically represented by $\frac{\delta}{\boldsymbol{\lambda}}$:

$$\left(\frac{\delta}{\lambda_1}, \frac{\delta}{\lambda_2}, \ldots, \frac{\delta}{\lambda_N}\right) \overset{\text{def}}{=} \frac{\delta}{\boldsymbol{\lambda}} \tag{1a}$$

Also, we define

$$\lambda_1 > \mu_1 \wedge \lambda_2 > \mu_2 \wedge \ldots \wedge \lambda_N > \mu_N \ \overset{\text{def}}{\Leftrightarrow}\ \boldsymbol{\lambda} > \boldsymbol{\mu} \tag{1b}$$

and

$$\lambda_1 \downarrow 0, \lambda_2 \downarrow 0, \ldots, \lambda_N \downarrow 0 \ \overset{\text{def}}{\Leftrightarrow}\ \boldsymbol{\lambda} \downarrow \mathbf{0} \quad ■ \tag{1c}$$

3.3. Approximation of functions by integral operators. A few comments on a huge subject

Definition 6: Let $C_b\ \mathbb{R}^N \to \mathbb{R}\ = C_b(\mathbb{R}^N)$ be the space of bounded continuous functions defined on $\mathbb{R}^N$, $(S,\prec)$ be a directed set (e.g., $S = \left(\mathbb{R}^+,\leq\right)$ or a set of positive definite matrices ordered by decreasing norm), and $\left(K_s\right)_{s\in S}$ be a family of locally integrable functions $K_s : \mathbb{R}^N \to \mathbb{R}$, with the property

$$\forall s \in S \qquad \int_{\mathbb{R}^N} K_s(\boldsymbol{u})\, d\boldsymbol{u} = 1. \tag{2}$$

The family $\left(K_s\right)_{s\in S}$ is called an ***approximate reproducing kernel family*** (**ARK family**), or a ***delta family*** (Bandyopadhyay (2002), Sec. 12.4), or an ***approximation to the identity*** (Stein 1993, Sec. 6) if

$$\lim_{s,\prec} \int_{\mathbb{R}^N} K_s(\boldsymbol{x}-\boldsymbol{u})\, h(\boldsymbol{u})\, d\boldsymbol{u} = h(\boldsymbol{x}), \quad \forall h \in C_b(\mathbb{R}^N) \ \text{ and } \ \forall \boldsymbol{x} \in \mathbb{R}^N, \tag{3}$$

We shall also (loosely speaking) refer to the functions K_s as ***the kernel*** (***functions***). ■

Kernel functions having the limiting reproducing property (3) first appeared in Fourier (Harmonic) Analysis, in connection with the problem of summation of Fourier series. The Dirichlet kernel, the Fejer kernel and the Poisson kernel are three well-known examples. See e.g. Katznelson (1968), Korner (1988), Duoandikoetxea (2001); the multidimensional case is studied in detail by Grafakos (2008), Classical Fourier Analysis (Chapter 3).

The functions K_s of an ARK family may be positive (non-negative([1])) or of alternating sign. In Harmonic Analysis the Dirichlet kernel is of alternating sign, while the Fejer's and Poisson's kernels are of positive type. In this section we shall study only **positive kernels**. The amazing fact is that positivity, in conjunction with condition (2), suffice to ensure the validity of the limiting reproducing property (3). Informally speaking,

"Every probability density function generates an* ARK *family."

3.4. Delta families generated by probability density functions through general linear transformations

Consider any function $K : \mathbb{R}^N \to \mathbb{R}$, such that

- $K(\boldsymbol{x}) \geq 0$, and (4a)

([1]) In the sequel we shall use the term ***positive*** as a synonym to the term ***non-negative***.

- $$\int_{\mathbb{R}^N} K(\boldsymbol{u})\, d\boldsymbol{u} = 1, \tag{4b}$$

that is, **any pdf over** $\mathbb{R}^N$. The support of $K : \mathbb{R}^N \to \mathbb{R}$ may be the whole $\mathbb{R}^N$ or some proper subset $A \subseteq \mathbb{R}^N$. The ***Gaussian pdf*** $f_{Gauss} : \mathbb{R}^N \to \mathbb{R}$ is a standard example with support $\mathbb{R}^N$. Its general form is given by the formula

$$f_{Gauss}(\boldsymbol{x}) = f_{Gauss}(\boldsymbol{x};\boldsymbol{\mu},\mathbf{C}) = \frac{1}{\sqrt{(2\pi)^N \cdot \det(\mathbf{C})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})^{\mathrm{T}} \cdot \mathbf{C}^{-1} \cdot (\boldsymbol{x}-\boldsymbol{\mu})\right\},$$

where $\boldsymbol{\mu}$ is the mean vector and $\mathbf{C}$ is the covariance matrix (assumed nonsingular). The Gaussian kernel K, which will be extensively used in the sequel, is defined by

$$K_{Gauss}(\boldsymbol{x}) = f_{Gauss}(\boldsymbol{x};\mathbf{0},\mathbf{C}) = \frac{1}{\sqrt{(2\pi)^N \cdot \det(\mathbf{C})}} \exp\left\{-\frac{1}{2}\boldsymbol{x}^{\mathrm{T}} \cdot \mathbf{C}^{-1} \cdot \boldsymbol{x}\right\}, \tag{5a}$$

or

$$K_{Gauss}(\boldsymbol{x}-\boldsymbol{u}) = f_{Gauss}(\boldsymbol{x};\boldsymbol{u},\mathbf{C}) = \frac{1}{\sqrt{(2\pi)^N \cdot \det(\mathbf{C})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^{\mathrm{T}} \cdot \mathbf{C}^{-1} \cdot (\boldsymbol{x}-\boldsymbol{u})\right\}, \tag{5b}$$

Given a function K satisfying Eq. (4a,b), i.e. given any pdf defined over $\mathbb{R}^N$, we define the family

$$K_{\mathbf{G}}(\boldsymbol{x}) = \frac{1}{|\det(\mathbf{G})|} K\left(\mathbf{G}^{-1} \cdot \boldsymbol{x}\right), \tag{6}$$

where $\mathbf{G} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ (i.e., $\mathbf{G}$ is a $(N \times N)$ real matrix) with $\det(\mathbf{G}) \neq 0$ (i.e., invertible).

Each kernel $K_{\mathbf{G}}$, as defined by Eq. (6), satisfies conditions (4a) and (4b), that is, each $K_{\mathbf{G}}(\boldsymbol{x})$ is a pdf as well. The condition $K_{\mathbf{G}}(\boldsymbol{x}) \geq 0$ is obvious; to prove condition (4b) use will be made of the formula of change of variables in multi-dimensional integrals, as stated in Theorem 1 of Sec. 3.2.

Lemma 2: Let $K_{\mathbf{G}}$ be defined in terms of Eq. (6), through a pdf K. Then, for any $\mathbf{G} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ with $\det(\mathbf{G}) \neq 0$,

$$\int_{\mathbb{R}^N} K_{\mathbf{G}}(\boldsymbol{x})\, d\boldsymbol{x} = 1 = \int_{\mathbb{R}^N} K_{\mathbf{G}}(\boldsymbol{x}-\boldsymbol{u})\, d\boldsymbol{u}. \quad \blacksquare \tag{7a,b}$$

Proof of Eq. (7a)**:** To calculate the integral $\int K_{\mathbf{G}}(\boldsymbol{x})\, d\boldsymbol{x} = \frac{1}{|\det(\mathbf{G})|} \int K\left(\mathbf{G}^{-1} \cdot \boldsymbol{x}\right) d\boldsymbol{x}$,

we shall use the change of variables

$$\begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_N \end{pmatrix} \equiv \boldsymbol{x} = \mathbf{T}(\boldsymbol{v}) = \mathbf{G}\cdot\boldsymbol{v} \quad \text{,where} \quad \boldsymbol{v} \in \mathbb{R}^N \tag{8a}$$

(anisotropic linear dilatation with rotation). This transformation possesses all properties listed in Theorem 1 of Sec. 3.2. Further, in this case we have

$$\boldsymbol{x} = \mathbf{G}\cdot\boldsymbol{v} \;\Rightarrow\; x_i = \sum_{k=1}^{N} G_{ik}\, v_k \,, \qquad 1 \le i \le N \quad \Rightarrow$$

$$\frac{\partial x_i}{\partial v_j} = \mathrm{G}_{i1}\cdot 0 + \ldots + \mathrm{G}_{ij}\cdot 1 + \ldots + \mathrm{G}_{iN}\cdot 0 = \mathrm{G}_{ij} \,, \qquad 1 \le i \le N$$

That means

$$\frac{D\boldsymbol{x}}{D\boldsymbol{v}} = \mathbf{G} \quad \Rightarrow \tag{8b}$$

$$\det\left(\frac{D\boldsymbol{x}}{D\boldsymbol{v}}\right) = \det(\mathbf{G}) \quad \text{and} \quad \mathbf{T}^{-1}(\mathbb{R}^N) = \mathbb{R}^N \tag{8c,d}$$

Thus, applying Theorem 1, we obtain

$$\int_{\mathbb{R}^N} K_{\mathbf{G}}(\boldsymbol{x})\,d\boldsymbol{x} = \frac{1}{\left|\det(\mathbf{G})\right|}\int_{\mathbb{R}^N} K\left(\mathbf{G}^{-1}\cdot\boldsymbol{x}\right)d\boldsymbol{x} =$$

$$= \frac{1}{\left|\det(\mathbf{G})\right|}\int_{\mathbf{T}^{-1}(\mathbb{R}^N)} K(\boldsymbol{v})\left|\det\left(\frac{D\boldsymbol{x}}{D\boldsymbol{v}}\right)\right| d\boldsymbol{v} =$$

$$= \frac{1}{\left|\det(\mathbf{G})\right|}\int_{\mathbb{R}^N} K(\boldsymbol{v})\left|\det(\mathbf{G})\right| d\boldsymbol{v} = \int_{\mathbb{R}^N} K(\boldsymbol{v})\,d\boldsymbol{v} = 1.$$

This proves Eq. (7a).

To prove Eq. (7b), the integral $\int K_{\mathbf{G}}(\boldsymbol{x}-\boldsymbol{u})\,d\boldsymbol{u}$ is calculated by using the transformation

$$\boldsymbol{u} = T_1(\boldsymbol{w}) = -\boldsymbol{w} + \boldsymbol{x}\,, \tag{9a}$$

having the properties

$$\left|\det\left(\frac{D\boldsymbol{u}}{D\boldsymbol{w}}\right)\right| = 1 \quad \text{and} \quad \mathbf{T}_1^{-1}(\mathbb{R}^N) = \mathbb{R}^N\,. \tag{9b,c}$$

Applying again Theorem 1, we obtain

$$\int_{\mathbb{R}^N} K_{\mathbf{G}}(\boldsymbol{x}-\boldsymbol{u})\,d\boldsymbol{u} = \int_{\mathbf{T}^{-1}(\mathbb{R}^N)=\mathbb{R}^N} K_{\mathbf{G}}(\boldsymbol{x}-T_1(\boldsymbol{w}))\left|\det\left(\frac{D\boldsymbol{u}}{D\boldsymbol{w}}\right)\right| d\boldsymbol{w} = ,$$

$$= \int_{\mathbb{R}^N} K_{\mathbf{G}}(\boldsymbol{w})\,d\boldsymbol{w} = 1$$

in accordance with the first part of the proof. Thus, the proof of Lemma 2 has been complete. ◄

Lemma 2 holds true for any matrix $\mathbf{G} \in \mathbb{R}^{\mathrm{N}\times\mathrm{N}}$ with the only property $\det(\mathbf{G}) \neq 0$. In continuing, in Lemmata 3, 4 and Theorem 2 we need a structure of convergence for the net $K_{\mathbf{G}}$.

For any proper rotation matrix $\mathbf{Q}$, (i.e., $\mathbf{Q}^{-1} = \mathbf{Q}^{\mathrm{T}}$, $\det(\mathbf{Q}) = +1$), we define the family of matrices $L_{\mathbf{Q}}$ as:

$$L_{\mathbf{Q}} = \left\{ \mathbf{G}_{\lambda} = \boldsymbol{\Lambda}\cdot\mathbf{Q} : \quad \boldsymbol{\Lambda} = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N) = \mathrm{diag}(\boldsymbol{\lambda}),\ \boldsymbol{\lambda} > \mathbf{0} \right\} \tag{10}$$

Notice that if $\mathbf{G}_{\lambda} \in L_{\mathbf{Q}}$, then $\det(\mathbf{G}_{\lambda}) = \det(\boldsymbol{\Lambda}\cdot\mathbf{Q}) = \det(\boldsymbol{\Lambda})\det(\mathbf{Q}) = \det(\boldsymbol{\Lambda})$, and $\det(\boldsymbol{\Lambda}) = \lambda_1^2\ \lambda_2^2 \ldots \lambda_N^2 > 0$.

Now, we are in a position to define a direction in $L_{\mathbf{Q}}$.

Definition 7 (direction in $L_{\mathbf{Q}}$)**:** We define the **binary relation** $\prec$ in $L_{\mathbf{Q}}$ as:

$\mathbf{G}_{\lambda_1} \prec \mathbf{G}_{\lambda_2} \iff \boldsymbol{\lambda}_2 \leq \boldsymbol{\lambda}_1$ (i.e., in this direction, $\boldsymbol{\lambda} \downarrow \mathbf{0}$; see Eq. (1b,c)). ■

We shall show that the above relation is a direction in $L_{\mathbf{Q}}$, for any $\mathbf{Q}$ ■

***Proof*:** We will prove the three properties of the direction; see Def. 3 of Sec. 3.2.

D1) For $\mathbf{G}_{\lambda} \in L_{\mathbf{Q}}$: $\boldsymbol{\lambda}_{\mathbf{G}_{\lambda}} \leq \boldsymbol{\lambda}_{\mathbf{G}_{\lambda}} \Rightarrow \mathbf{G}_{\lambda} \prec \mathbf{G}_{\lambda}$.

D2) For $\mathbf{G}_{\lambda_1}, \mathbf{G}_{\lambda_2}, \mathbf{G}_{\lambda_3} \in L_{\mathbf{Q}}$ with $\mathbf{G}_{\lambda_1} \prec \mathbf{G}_{\lambda_2}$ and $\mathbf{G}_{\lambda_2} \prec \mathbf{G}_{\lambda_3} \Rightarrow \boldsymbol{\lambda}_2 \leq \boldsymbol{\lambda}_1$ and $\boldsymbol{\lambda}_3 \leq \boldsymbol{\lambda}_2$ so we obtain $\boldsymbol{\lambda}_3 \leq \boldsymbol{\lambda}_2 \leq \boldsymbol{\lambda}_1 \Rightarrow \boldsymbol{\lambda}_3 \leq \boldsymbol{\lambda}_1 \Rightarrow \mathbf{G}_{\lambda_1} \prec \mathbf{G}_{\lambda_3}$.

D3) For $\mathbf{G}_{\lambda_1}, \mathbf{G}_{\lambda_2} \in L_{\mathbf{Q}}$, let $\boldsymbol{\lambda}_1 = (\lambda_1, \lambda_2, \ldots, \lambda_N)$ and $\boldsymbol{\lambda}_2 = (\mu_1, \mu_2, \ldots, \mu_N)$, we set $\boldsymbol{\lambda}_3 = \left(\min(\lambda_1,\mu_1), \min(\lambda_2,\mu_2), \ldots, \min(\lambda_N,\mu_N)\right) \equiv (\nu_1, \nu_2, \ldots, \nu_N)$, and, $\boldsymbol{\Lambda}_3 = \mathrm{diag}(\nu_1, \nu_2, \ldots, \nu_N)$. Then, by setting $\mathbf{G}_{\lambda_3} = \boldsymbol{\Lambda}_3\cdot\mathbf{Q} \in L_{\mathbf{Q}}$, we obtain $\boldsymbol{\lambda}_3 \leq \boldsymbol{\lambda}_1$

and $\boldsymbol{\lambda}_3 \leq \boldsymbol{\lambda}_2 \;\Rightarrow\; \mathbf{G}_{\lambda_1} \prec \mathbf{G}_{\lambda_3}$ and $\mathbf{G}_{\lambda_2} \prec \mathbf{G}_{\lambda_3}$. ◀

So, the $\left(L_{\mathbf{Q}} , \prec \right)$ is directed set which we shall use for the convergence in Lemmata 3,4, and, when the net $K_{\mathbf{G}_\lambda} :(L , \prec) \to \mathbb{R}$ converges to $k \in \mathbb{R}$, we shall write $\lim\limits_{\mathbf{G}_\lambda , \prec} K_{\mathbf{G}_\lambda} = k$.

Lemmata 3 and 4, presented below, establish the important fact that the probability mass of $K_{\mathbf{G}_\lambda}$ is finally (as $\boldsymbol{\lambda} \downarrow \mathbf{0}$) concentrated within a ball of arbitrarily small radius.

Lemma 3: Let $B(\mathbf{0},\delta) = \{ \boldsymbol{x} \in \mathbb{R}^N : \| \boldsymbol{x} \|_2 < \delta \}$, the open ball centered at $\mathbf{0}$ with radius δ, and $K_{\mathbf{G}_\lambda}$ be defined in terms of a pdf K, by means of Eq. (6) and, $\mathbf{G}_\lambda \in L_{\mathbf{Q}}$ for some $\mathbf{Q}$. Then, for any given $\delta > 0$ and for any proper rotation matrix $\mathbf{Q}$, the following limiting relations, in the direction $\prec$ (see Def. 7), hold true:

$$\lim_{\mathbf{G}_\lambda , \prec} \int_{B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x} = 1, \tag{11a}$$

$$\lim_{\mathbf{G}_\lambda , \prec} \int_{\mathbb{R}^N - B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x} = 0 \qquad ■ \tag{11b}$$

Proof of Eq. (11a)**:** To calculate the integral, use will be made of the transformation $\boldsymbol{x} = \boldsymbol{T}(\boldsymbol{v})$ $\left(\boldsymbol{x} = \mathbf{G}_\lambda \cdot \boldsymbol{v} \right)$, introduced in the proof of Lemma 2, Eq. (7a). In order to specify the inverse image $\boldsymbol{T}^{-1}\left(B(\mathbf{0},\delta) \right)$, we observe that

$$\boldsymbol{x} \in B(\mathbf{0},\delta) \Leftrightarrow \| \boldsymbol{x} \|_2^2 \leq \delta^2 \Leftrightarrow \| \mathbf{G}_\lambda \cdot \boldsymbol{v} \|_2^2 \leq \delta^2 \Leftrightarrow \boldsymbol{v}^T \cdot \mathbf{G}_\lambda{}^T \cdot \mathbf{G}_\lambda \cdot \boldsymbol{v} \leq \delta^2 \Leftrightarrow$$

$$\Leftrightarrow \boldsymbol{v}^T \cdot \mathbf{Q}^{\mathrm{T}} \cdot \boldsymbol{\Lambda}^2 \cdot \mathbf{Q} \cdot \boldsymbol{v} \leq \delta^2 \Leftrightarrow \boldsymbol{v}^T \cdot \frac{1}{\delta^2} \mathbf{Q}^{\mathrm{T}} \cdot \boldsymbol{\Lambda}^2 \cdot \mathbf{Q} \cdot \boldsymbol{v} \leq 1 \Leftrightarrow$$

$$\Leftrightarrow \boldsymbol{v} \in \mathcal{E}_{\mathbf{Q}} \left(\mathbf{0} ; \frac{\delta}{\sqrt{\lambda_1^2}}, \frac{\delta}{\sqrt{\lambda_2^2}}, \ldots , \frac{\delta}{\sqrt{\lambda_N^2}} \right) \equiv \mathcal{E}_{\mathbf{Q}} \left(\mathbf{0} ; \frac{\delta}{\lambda_1}, \frac{\delta}{\lambda_2}, \ldots , \frac{\delta}{\lambda_N} \right)$$

$$\equiv \mathcal{E}_{\mathbf{Q}} \left(\mathbf{0} ; \delta / \boldsymbol{\lambda} \right)$$

that is,

$$\mathbf{T}^{-1}\left(B(\mathbf{0},\delta) \right) = \mathcal{E}_{\mathbf{Q}} \left(\mathbf{0} ; \frac{\delta}{\lambda_1}, \frac{\delta}{\lambda_2}, \ldots , \frac{\delta}{\lambda_N} \right) \equiv \mathcal{E}_{\mathbf{Q}} \left(\mathbf{0} ; \delta / \boldsymbol{\lambda} \right), \tag{12a}$$

the ***ellipsoidal with semiaxes*** $\delta/\lambda_1, \delta/\lambda_2, \ldots, \delta/\lambda_N$, and the orthogonal columns of $\mathbf{Q}$, $\left\{ \boldsymbol{q}_i\ ,\ 1 \le i \le N \right\}$, determine the orientation of ellipsoidal.

Taking into account Eqs. (8a,b) and (12), and applying the formula of change of variables in multidimensional integrals, we calculate the integral appearing in Eq. (11a) as follows:

$$\begin{aligned}
\int_{B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x} &= \frac{1}{\det(\mathbf{G}_\lambda)} \int_{B(\mathbf{0},\delta)} K\left(\mathbf{G}_\lambda^{-1}\cdot \boldsymbol{x}\right) d\boldsymbol{x} \\
&= \frac{1}{\det(\mathbf{G}_\lambda)} \int_{\mathbf{T}^{-1}(B(\mathbf{0},\delta))} K(\boldsymbol{v}) \left| \det\left(\frac{D\boldsymbol{x}}{D\boldsymbol{v}} \right) \right| d\boldsymbol{v} \\
&= \frac{1}{\det(\mathbf{G}_\lambda)} \int_{\mathcal{E}_{\mathbf{Q}}(\mathbf{0};\delta/\boldsymbol{\lambda})} K(\boldsymbol{v}) \det(\mathbf{G}_\lambda)\, d\boldsymbol{v} \\
&= \int_{\mathcal{E}_{\mathbf{Q}}(\mathbf{0};\delta/\boldsymbol{\lambda})} K(\boldsymbol{v})\, d\boldsymbol{v},
\end{aligned}$$

that is,

$$\int_{B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathcal{E}_{\mathbf{Q}}(\mathbf{0};\delta/\boldsymbol{\lambda})} K(\boldsymbol{v})\, d\boldsymbol{v}\,. \tag{12b}$$

Taking the limit of both sides of the above equation in the direction $\prec$ (i.e., $\boldsymbol{\lambda} \downarrow \mathbf{0}$), and observing that $\lim_{\mathbf{G}_\lambda,\prec} \mathcal{E}_{\mathbf{Q}}(\mathbf{0};\delta/\boldsymbol{\lambda}) = \mathbb{R}^N$ (it holds true because $\mathbf{Q}$ is orthogonal), we obtain

$$\begin{aligned}
\lim_{\mathbf{G}_\lambda,\prec} \int_{B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x} &= \lim_{\mathbf{G}_\lambda,\prec} \int_{\mathcal{E}(\mathbf{0};\delta/\boldsymbol{\lambda})} K(\boldsymbol{v})\, d\boldsymbol{v} = \\
&= \int_{\lim_{\mathbf{G}_\lambda,\prec} \mathcal{E}(\mathbf{0};\delta/\boldsymbol{\lambda})} K(\boldsymbol{v})\, d\boldsymbol{v} = \int_{\mathbb{R}^N} K(\boldsymbol{v})\, d\boldsymbol{v} = 1\,,
\end{aligned}$$

which proves Eq. (11a).

To prove Eq. (11b) we observe that

$$\int_{\mathbb{R}^N - B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x} = \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x} - \int_{B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x})\, d\boldsymbol{x}.$$

Taking the limit of both sides of the above equation in the direction $\prec$, and using Eqs. (7a) and (11a), we obtain Eq. (11b). The proof of the Lemma 3 is complete. ◄

Remark 1: In proof of the above lemma, we observe that we used only the property that the columns of $\mathbf{Q}$, $\{ \boldsymbol{q}_i \, , \; 1 \le i \le N \}$, are orthogonal in $\mathbb{R}^N$ and it is independent which are exactly. It is a crucial observation for the proof of Theorem 2 below.

Lemma 4: Let $B(\boldsymbol{x}, \delta) = \left\{ \boldsymbol{u} \in \mathbb{R}^N : \| \boldsymbol{u} - \boldsymbol{x} \|_2 < \delta \right\}$, the open ball centered at $\boldsymbol{x}$ with radius δ, and $K_{\mathbf{G}_\lambda}$ be defined in terms of a pdf K, by means of Eq. (6) and $\mathbf{G}_\lambda \in L_{\mathbf{Q}}$, for some $\mathbf{Q}$. Then, for any given $\delta > 0$, and for any proper rotation matrix $\mathbf{Q}$, the following limiting relations, in the direction $\prec$, hold true:

$$\lim_{\mathbf{G}_\lambda, \prec} \int_{B(\boldsymbol{x}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \, d\boldsymbol{u} = 1, \tag{13a}$$

$$\lim_{\mathbf{G}_\lambda, \prec} \int_{\mathbb{R}^N - B(\boldsymbol{x}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \, d\boldsymbol{u} = 0. \quad \blacksquare \tag{13b}$$

Proof of Eq. (13a)**:** To calculate the integral, we shall use the transformation $\boldsymbol{u} = \boldsymbol{T}_1(\boldsymbol{w}) = -\boldsymbol{w} + \boldsymbol{x}$, written also $(u_n = -w_n + x_n)$, introduced in the second part of the proof of Lemma 2, Eq. (9a). In order to specify the inverse image $\boldsymbol{T}_1^{-1}(B(\boldsymbol{x}, \delta))$, we observe that

$$\boldsymbol{u} \in B(\boldsymbol{x}, \delta) \;\Leftrightarrow\; \sum_{n=1}^{N} \left(u_n - x_n \right)^2 \le \delta^2 \;\Leftrightarrow\; \sum_{n=1}^{N} w_n^2 \le \delta^2 \;\Leftrightarrow\; \boldsymbol{w} \in B(\mathbf{0}, \delta),$$

that is, $\boldsymbol{T}_1^{-1}(B(\boldsymbol{x}, \delta)) = B(\mathbf{0}, \delta)$. Using this result and Eq. (9b), the application of Theorem 1 leads to the following calculations:

$$\int_{B(\boldsymbol{x}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \, d\boldsymbol{u} = \int_{\boldsymbol{T}_1^{-1}(B(\boldsymbol{x}, \delta))} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{T}_1(\boldsymbol{w})) \left| \det\left(\frac{D\boldsymbol{u}}{D\boldsymbol{w}} \right) \right| d\boldsymbol{w}$$

$$= \int_{B(\mathbf{0}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{w}) \, d\boldsymbol{w}$$

Thus, the desired conclusion follows by taking the limit of the first and the last terms of the above equation in the direction $\prec$, and invoking Lemma 3, Eq. (11a).

Eq. (13b) is simply obtained from Eq. (13a), exactly as Eq. (11b) has been derived by Eq. (11a), in Lemma 3. ◄

Remark 2: It is important to keep in mind that the limiting equations in Lemmata 2 and 3 are valid for <u>any given radius</u> δ of the balls $B(\mathbf{0}, \delta)$ and $B(\boldsymbol{x}, \delta)$.

We are now in a position to state and prove the ***main result of this section***:

Theorem 2 [Every pdf generates a delta family]: Let $K: \mathbb{R}^N \to \mathbb{R}$ be a pdf over $\mathbb{R}^N$, and $K_{\mathbf{G}_\lambda}(\boldsymbol{x}) = \frac{1}{\det(\mathbf{G}_\lambda)} K\left(\mathbf{G}_\lambda^{-1} \cdot \boldsymbol{x}\right)$, and $\mathbf{G}_\lambda \in L_{\mathbf{Q}}$, for some $\mathbf{Q}$.

Then, for any function $h \in C_b(\mathbb{R}^N)$, and for any proper rotation matrix $\mathbf{Q}$, we have

$$\lim_{\mathbf{G}_\lambda, \prec} \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u})\, h(\boldsymbol{u})\, d\boldsymbol{u} = h(\boldsymbol{x}). \qquad \blacksquare \qquad (14)$$

Proof: Since $K_{\mathbf{G}_\lambda} \in L^1(\mathbb{R}^N)$ and h is bounded over $\mathbb{R}^N$, the integral $\int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{u})\, h(\boldsymbol{u})\, d\boldsymbol{u}$ exists. Also, from the boundedness of h, we conclude that there exists a constant $M_h > 0$ such that

$$\sup\left\{ |h(\boldsymbol{u})|, \boldsymbol{u} \in A \subseteq \mathbb{R}^N \right\} \leq \sup\left\{ |h(\boldsymbol{u})|, \boldsymbol{u} \in \mathbb{R}^N \right\} \leq M_h. \qquad \text{(a)}$$

Further, using Eq. (7b) and the positivity of $K_{\mathbf{G}_\lambda}$, we obtain

$$\left| \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u})\, h(\boldsymbol{u})\, d\boldsymbol{u} - h(\boldsymbol{x}) \right| = \left| \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u})\, h(\boldsymbol{u})\, d\boldsymbol{u} - \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u})\, h(\boldsymbol{x})\, d\boldsymbol{u} \right| =$$

$$= \left| \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \big(h(\boldsymbol{u}) - h(\boldsymbol{x}) \big)\, d\boldsymbol{u} \right| \leq \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right| d\boldsymbol{u} \qquad \text{(b)}$$

Decomposing the last integral in two parts, over the ball $B(\boldsymbol{x}, \delta)$ and its complement $\mathbb{R}^N - B(\boldsymbol{x}, \delta)$, we obtain

$$\int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right| d\boldsymbol{u} =$$

$$= \int_{B(\boldsymbol{x}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right| d\boldsymbol{u} + \int_{\mathbb{R}^N - B(\boldsymbol{x}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u}) \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right| d\boldsymbol{u} \leq$$

$$\leq \sup\left\{ \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right|, \boldsymbol{u} \in B(\boldsymbol{x}, \delta) \right\} \int_{B(\boldsymbol{x}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u})\, d\boldsymbol{u} +$$

$$+ \sup\left\{ \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right|, \boldsymbol{u} \in \mathbb{R}^N - B(\boldsymbol{x}, \delta) \right\} \int_{\mathbb{R}^N - B(\boldsymbol{x}, \delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x} - \boldsymbol{u})\, d\boldsymbol{u}. \qquad \text{(c)}$$

We shall now show that both terms of the last member of (c) become arbitrarily small, which essentially proves our assertion. First, on the basis of inequality (a), we obtain

$$\sup\left\{ \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right|,\ \boldsymbol{u} \in \mathbb{R}^N - B(\boldsymbol{x},\delta) \right\} \le 2\, M_h . \tag{d}$$

Besides, since h is continuous at (any) $\boldsymbol{x} \in \mathbb{R}^N$, for any given $\varepsilon/2 > 0$, it is possible to find a $\delta = \delta(\varepsilon, \boldsymbol{x}) > 0$, such that

$$\sup\left\{ \left| h(\boldsymbol{u}) - h(\boldsymbol{x}) \right|,\ \boldsymbol{u} \in B(\boldsymbol{x},\delta) \right\} < \frac{\varepsilon}{2} . \tag{e}$$

Combining Eqs. (b), (c), (d) and (e), we obtain

$$\left| \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\, h(\boldsymbol{u})\, d\boldsymbol{u} - h(\boldsymbol{x}) \right| < \frac{\varepsilon}{2} \int_{B(\boldsymbol{x},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\, d\boldsymbol{u} + 2\, M_h \int_{\mathbb{R}^N - B(\boldsymbol{x},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\, d\boldsymbol{u} . \tag{f}$$

Further, because of the Eq. (13b), we can find a matrix $\mathbf{G}_{\lambda_0} = \mathbf{G}_{\lambda_0(\varepsilon)} \in L_{\mathbf{Q}}$, $\boldsymbol{\lambda}_0$ **is the same for every** $\mathbf{Q}$, such that, for any $\mathbf{G}_\lambda \in L_{\mathbf{Q}}$: $\mathbf{G}_{\lambda_0} \prec \mathbf{G}_\lambda$,

$$\int_{\mathbb{R}^N - B(\boldsymbol{x},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\, d\boldsymbol{u} < \frac{\varepsilon}{4\, M_h} . \tag{g}$$

Finally, using (g), and the fact that

$$\int_{B(\boldsymbol{x},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\, d\boldsymbol{u} \le \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\, d\boldsymbol{u} = 1 \quad \text{(Lemma 2)},$$

inequality (f) implies that, for any $\mathbf{G} \in L_{\mathbf{Q}}$: $\mathbf{G}_{\lambda_0} \prec \mathbf{G}$, we have

$$\left| \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\, h(\boldsymbol{u})\, d\boldsymbol{u} - h(\boldsymbol{x}) \right| < \frac{\varepsilon}{2} \cdot 1 + 2\, M_h \cdot \frac{\varepsilon}{4\, M_h} = \varepsilon .$$

The last inequality is equivalent with Eq. (14). The proof of Theorem 2 is thus completed. ◄

The essential in Theorem 2 is that $\boldsymbol{\lambda}_0$ is independent from the proper rotation matrix $\mathbf{Q}$. It means we can find the same λ_0 for every $\mathbf{Q}$, such that, In. (g) holds true. We can see this observing (see Lemmata 3 and 4) that

$$\int_{\mathbb{R}^N - B(\boldsymbol{x},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\,d\boldsymbol{u} = \int_{\mathbb{R}^N - B(\mathbf{0},\delta)} K_{\mathbf{G}_\lambda}(\boldsymbol{w})\,d\boldsymbol{w} = \int_{\mathbb{R}^N - \mathcal{E}_{\mathbf{Q}}(\mathbf{0};\delta/\lambda)} K(\boldsymbol{v})\,d\boldsymbol{v}$$

Now, combining the Remark 1, which we made in Lemma 3, and the above equality, there exists $\lambda_0 = \lambda_0(\varepsilon)$, (same for every $\mathbf{Q}$), such that, for any $\lambda < \lambda_0$:

$$\int_{\mathbb{R}^N - \mathcal{E}_{\mathbf{Q}}(\mathbf{0};\delta/\lambda)} K(\boldsymbol{v})\,d\boldsymbol{v} < \frac{\varepsilon}{4\,M_h} .$$

Equivalently, there exists $\mathbf{G}_{\lambda_0} = \mathbf{\Lambda}_0 \cdot \mathbf{Q}$, $\mathbf{\Lambda}_0 = \operatorname{diag}(\lambda_0)$, such that, for any $\mathbf{G}_\lambda \in L_{\mathbf{Q}}$: $\mathbf{G}_{\lambda_0} \prec \mathbf{G}_\lambda$ In. (g) holds true, and λ_0 is same for every $\mathbf{Q}$.

Observing this detail, in proof of Theorem 2, we can assume that $\mathbf{Q}$ is generally different at any point $\boldsymbol{x} \in \operatorname{supp}(h)$, i.e., $\mathbf{Q} = \mathbf{Q}(\boldsymbol{x}) = \mathbf{Q}_{\boldsymbol{x}}$, and we can now state **the general and main result of this work**:

$$\lim_{\mathbf{G}_{\lambda}(x),\prec} \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda(x)}(\boldsymbol{x}-\boldsymbol{u})\,h(\boldsymbol{u})\,d\boldsymbol{u} = h(\boldsymbol{x}) \qquad (14)'$$

3.5. The delta family generated by a general (correlated) Gaussian pdf

Notational conventions: In the sequel, by writing $\mathbf{Q}, \mathbf{G}_\lambda$ and $\tilde{\mathbf{C}}$ we shall mean that matrices are dependent from the point $\boldsymbol{x}$, i.e., $\mathbf{Q} = \mathbf{Q}(\boldsymbol{x}), \mathbf{G}_\lambda = \mathbf{G}_\lambda(\boldsymbol{x})$ and $\tilde{\mathbf{C}} = \tilde{\mathbf{C}}(\boldsymbol{x})$, and we shall write the vector $\boldsymbol{x}$ when we want to emphasize .

Let $\mathbf{C}$ be a nonsingular (invertible) covariance matrix and

$$K(\boldsymbol{x}-\boldsymbol{u}) = \frac{1}{\sqrt{(2\pi)^N \cdot \det(\mathbf{C})}} \exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^{\mathrm{T}} \cdot \mathbf{C}^{-1} \cdot (\boldsymbol{x}-\boldsymbol{u})\right\}, \qquad (15)$$

be the Gaussian pdf centered at $\boldsymbol{u}$, with covariance matrix $\mathbf{C}$. The latter is not assumed diagonal; it may be a complete (nonsingular, positive-definite) matrix. Then, in accordance to Theorem 2 of Sec. 3.4, the kernel family

$$K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u}) = \frac{1}{\det(\mathbf{G}_\lambda)} \frac{1}{\sqrt{(2\pi)^N \cdot \det(\mathbf{C})}} \exp\left\{-\frac{1}{2}\left(\mathbf{G}_\lambda^{-1} \cdot (\boldsymbol{x}-\boldsymbol{u})\right)^{\mathrm{T}} \cdot \mathbf{C}^{-1} \cdot \mathbf{G}_\lambda^{-1} \cdot (\boldsymbol{x}-\boldsymbol{u})\right\}, \qquad (16)$$

is a delta family, producing the approximation to the identity

$$\lim_{\mathbf{G}_\lambda,\prec} \int_{\mathbb{R}^N} K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u})\,f(\boldsymbol{u})\,d\boldsymbol{u} = f(\boldsymbol{x}), \qquad (17)$$

for any continuous pdf $f : \mathbb{R}^N \to \mathbb{R}$. For our applications it is more convenient to reformulate Eq. (15) so that the matrix $\mathbf{G}_\lambda$ does not affect directly the variables $\boldsymbol{x}$ and $\boldsymbol{u}$. This becomes possible with the aid of some auxiliary matrix-theoretic results, which are collected and proved in the following lemma.

Lemma 5: Let $\boldsymbol{x} = x_1, x_2, \ldots, x_N$, $\mathbf{C} \in \mathbb{R}^{N\times N}$, $\det(\mathbf{C}) \neq 0$ and $\mathbf{G}_\lambda = \boldsymbol{\Lambda}\cdot\mathbf{Q}$, where $\Lambda = \mathrm{diag}(\lambda_1, \lambda_2, \ldots, \lambda_N)$ with $\lambda_i > 0$, $1 \le i \le N$, and, $\mathbf{Q}$ is a proper rotation matrix, i.e., orthogonal with $\det(\mathbf{Q}) = +1$. Then,

(a) We have $\left((\boldsymbol{\Lambda}\cdot\mathbf{Q})^{-1}\cdot\boldsymbol{x}\right)^T\cdot\mathbf{C}^{-1}\cdot(\boldsymbol{\Lambda}\cdot\mathbf{Q})^{-1}\cdot\boldsymbol{x} = \boldsymbol{x}^T\cdot\boldsymbol{\Lambda}^{-1}\cdot\left(\mathbf{Q}\cdot\mathbf{C}^{-1}\cdot\mathbf{Q}^T\right)\cdot\boldsymbol{\Lambda}^{-1}\cdot\boldsymbol{x}$

where $\boldsymbol{\Lambda}^{-1}\cdot\left(\mathbf{Q}^T\cdot\mathbf{C}^{-1}\cdot\mathbf{Q}\right)\cdot\boldsymbol{\Lambda}^{-1} = \left(\dfrac{1}{\lambda_i\,\lambda_j}\displaystyle\sum_{l=1}^{N}\sum_{k=1}^{N} Q_{ik}^T\, C_{kl}^{-1}\, Q_{lj}\right)_{N\times N}$.

(b) The inverse of the matrix of $\tilde{\mathbf{C}} = \boldsymbol{\Lambda}\cdot\left(\mathbf{Q}^T\cdot\mathbf{C}\cdot\mathbf{Q}\right)\cdot\boldsymbol{\Lambda}$, where

$\boldsymbol{\Lambda}\cdot\left(\mathbf{Q}^T\cdot\mathbf{C}\cdot\mathbf{Q}\right)\cdot\boldsymbol{\Lambda} = \left(\lambda_i\,\lambda_j\displaystyle\sum_{l=1}^{N}\sum_{k=1}^{N} Q_{ik}^T\, C_{kl}^{-1}\, Q_{lj}\right)_{N\times N}$, is the matrix

$\tilde{\mathbf{C}}^{-1} = \boldsymbol{\Lambda}^{-1}\cdot\left(\mathbf{Q}^T\cdot\mathbf{C}^{-1}\cdot\mathbf{Q}\right)\cdot\boldsymbol{\Lambda}^{-1}$

(c) If $\tilde{\mathbf{C}} = \boldsymbol{\Lambda}\cdot\left(\mathbf{Q}^T\cdot\mathbf{C}\cdot\mathbf{Q}\right)\cdot\boldsymbol{\Lambda}$ then, $\det(\tilde{\mathbf{C}}) = \lambda_1^2\,\lambda_2^2\,\ldots\,\lambda_N^2\cdot\det(\mathbf{C})$. ■

***Proof*:** The proof of the above lemma is obvious by combining, the known results from linear algebra, if $\mathbf{A}, \mathbf{B} \in \mathbb{R}^{N\times N}$, and $\boldsymbol{\Lambda}$ is a diagonal matrix, then $(\mathbf{A}\cdot\mathbf{B})^T = \mathbf{B}^T\cdot\mathbf{A}^T$, $(\mathbf{A}\cdot\mathbf{B})^{-1} = \mathbf{B}^{-1}\cdot\mathbf{A}^{-1}$, $\boldsymbol{\Lambda}^T = \boldsymbol{\Lambda}$, and $\det(\mathbf{A}\cdot\mathbf{B}) = \det(\mathbf{A})\cdot\det(\mathbf{B})$, and the properties of a proper rotation matrix, $\mathbf{Q}^{-1} = \mathbf{Q}^T$ and $\det(\mathbf{Q}) = +1$. ◀

Theorem 3 [Reformulation of the Gaussian kernel in terms of unscaled variables $\boldsymbol{x}$ and $\boldsymbol{u}$]: Let $\mathbf{C} \equiv (C_{ij})_{N\times N}$ be a covariance matrix, and $\tilde{\mathbf{C}} = \boldsymbol{\Lambda}\cdot\left(\mathbf{Q}^T\cdot\mathbf{C}\cdot\mathbf{Q}\right)\cdot\boldsymbol{\Lambda}$ be the corresponding covariance matrix. Then, the Gaussian kernel, Eq. (16), can be written in terms of unscaled variables $\boldsymbol{x}$ and $\boldsymbol{u}$, and the scaled covariance $\tilde{\mathbf{C}}(\boldsymbol{x})$, in the form

$$K_{\mathbf{G}_\lambda}(\boldsymbol{x}-\boldsymbol{u}) = K\left(\boldsymbol{x}-\boldsymbol{u};\tilde{\mathbf{C}}\right) \tag{18a}$$

where,

$$K\left(\boldsymbol{x}-\boldsymbol{u};\tilde{\mathbf{C}}\right) = \frac{1}{\sqrt{(2\pi)^N\cdot\det(\tilde{\mathbf{C}})}}\exp\left\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{u})^{\mathrm{T}}\cdot\tilde{\mathbf{C}}^{-1}\cdot(\boldsymbol{x}-\boldsymbol{u})\right\}$$

and $\tilde{\mathbf{C}} = \mathbf{\Lambda} \cdot \left(\mathbf{Q}^T \cdot \mathbf{C} \cdot \mathbf{Q} \right) \cdot \mathbf{\Lambda}$. ■ (18b,c)

***Proof*:** By setting $\mathbf{G}_\lambda = \mathbf{\Lambda} \cdot \mathbf{Q}$ in Eq. (16) and using (a),(b),(c) of Lemma 5 we easily derive Eqs. (18a,b,c). ◄

The correlation coefficients of the $\lambda-$dependent localized covariance matrix

It is interesting to see the change of the correlation coefficients after application the linear transformation. The initial correlation coefficients of the covariance matrix $\mathbf{C}$ and the corresponding of the $\lambda-$dependent localized covariance matrix $\tilde{\mathbf{C}} = \mathbf{\Lambda} \cdot \left(\mathbf{Q}^T \cdot \mathbf{C} \cdot \mathbf{Q} \right) \cdot \mathbf{\Lambda}$ given by the following equations

$$\rho_{ij} = \frac{C_{ij}}{\sqrt{C_{ii}} \sqrt{C_{jj}}} \tag{19a}$$

$$\tilde{\rho}_{ij} = \frac{\tilde{C}_{ij}}{\sqrt{\tilde{C}_{ii}} \sqrt{\tilde{C}_{jj}}} = \frac{\sum_{l=1}^{N} \sum_{k=1}^{N} Q_{ik}^T \, C_{kl} \, Q_{lj}}{\left(\sum_{l=1}^{N} \sum_{k=1}^{N} Q_{ik}^T \, C_{kl} \, Q_{li} \right)^{1/2} \left(\sum_{l=1}^{N} \sum_{k=1}^{N} Q_{jk}^T \, C_{kl} \, Q_{lj} \right)^{1/2}}, \tag{19b}$$

We observe that $\tilde{\rho}_{ij}$ are independent from the scaled variable λ but they are dependent from the rotation matrix $\mathbf{Q}$, as we expected. This means that $\tilde{\rho}_{ij} = \tilde{\rho}_{ij}(\boldsymbol{x})$. Notice that, generally, $\rho_{ij} \neq \tilde{\rho}_{ij}$, but in special case that $\mathbf{Q}$ is the identity matrix we can easily show that $\rho_{ij} = \tilde{\rho}_{ij}$.

From the integral approximate representation to the approximation by means of linear superposition of Gaussian kernel with localized covariance matrices:

$$\int_{\mathbb{R}^N} K\left(\boldsymbol{x} - \boldsymbol{u} \,; \tilde{\mathbf{C}}(\boldsymbol{x}) \right) f(\boldsymbol{u}) \, d\boldsymbol{u} \approx f(\boldsymbol{x}) \tag{20}$$

Now, applying the definition of Riemann integral we can find a partition of the support of f, such that,

$$\int_{\mathbb{R}^N} K\left(\boldsymbol{x} - \boldsymbol{u} \,; \tilde{\mathbf{C}}(\boldsymbol{x}) \right) f(\boldsymbol{u}) \, d\boldsymbol{u} \approx \sum_{\ell \in \mathcal{L}} K\left(\boldsymbol{x} - \boldsymbol{u}_\ell \,; \tilde{\mathbf{C}}(\boldsymbol{x}) \right) f(\boldsymbol{u}_\ell) \, \Delta \boldsymbol{u}_\ell \,. \tag{21}$$

Where $\mathcal{L}$ is the set of points of the partition. Combining Eqs. (20) and (21) we also obtain

$$\sum_{\ell \in \mathcal{L}} K\left(\boldsymbol{x} - \boldsymbol{u}_\ell \,;\tilde{\mathbf{C}}(\boldsymbol{x}) \right) f(\boldsymbol{u}_\ell)\Delta \boldsymbol{u}_\ell \;\approx\; f(\boldsymbol{x}). \tag{22}$$

By setting $p_\ell = f(\boldsymbol{u}_\ell)\Delta \boldsymbol{u}_\ell$, we obtain the following discrete approximation form of a pdf f,

$$f(\boldsymbol{x}) \approx \sum_{\ell \in \mathcal{L}} p_\ell \, K\left(\boldsymbol{x} - \boldsymbol{u}_\ell \,;\tilde{\mathbf{C}}(\boldsymbol{x}) \right) \qquad \boldsymbol{u}_\ell \in \operatorname{supp}(f). \tag{24}$$

It is enough to choose $\boldsymbol{\lambda} \le \boldsymbol{\lambda}_0$ at any point $\boldsymbol{x}$, where $\boldsymbol{\lambda}_0$ is that we found in In. (14g). Thus ,by (b) of Lemma 5, the $(i\,j)$ element of $\tilde{\mathbf{C}}$ is equal with

$$\tilde{\mathbf{C}}_{ij}(\boldsymbol{x}) = \lambda_i \lambda_j \left(\left(\mathbf{Q}(\boldsymbol{x})\right)^T \cdot \mathbf{C} \cdot \mathbf{Q}(\boldsymbol{x}) \right)_{i\,j}. \tag{23}$$

So, we choose $\tilde{\mathbf{C}}$, such that, it has “little” elements to its diagonal and “free” nondiagonal ones.

3.6. Notes

As we see, we chose a particular family of matrices (Eq. (10)) to prove the main lemmata and theorems in Sections 3.4 and 3.5. Someone could observe that we can use other families as well. For instance, the diagonalizable matrices

$$L_{\mathbf{Q}} = \left\{ \mathbf{G}_{\boldsymbol{\lambda}} = \mathbf{Q}^T \cdot \boldsymbol{\Lambda} \cdot \mathbf{Q} : \quad \boldsymbol{\Lambda} = \operatorname{diag}\left(\lambda_1, \lambda_2, \dots, \lambda_N\right) = \operatorname{diag}(\boldsymbol{\lambda}),\ \boldsymbol{\lambda} > \mathbf{0} \right\},$$

or

$$L_{\mathbf{Q}} = \left\{ \mathbf{G}_{\boldsymbol{\lambda}} = \mathbf{Q}^T \cdot \boldsymbol{\Lambda} : \quad \boldsymbol{\Lambda} = \operatorname{diag}\left(\lambda_1, \lambda_2, \dots, \lambda_N\right) = \operatorname{diag}(\boldsymbol{\lambda}),\ \boldsymbol{\lambda} > \mathbf{0} \right\}.$$

In that case, we can prove the basic theory with similar way, but we cannot write a “clear” discrete form as Eq. (24). In particular, in these cases we cannot distinguish the localized parameter $\boldsymbol{\lambda}$ from the localized covariance matrix, as we did in Theorem 3 (Eq. (18c)). In some cases, we can avoid it with a change of variable, but then, it becomes more complex.

In addition, an interesting case is when $\mathbf{G}_{\boldsymbol{\lambda}}$ is a diagonal matrix with positive determinant, which is a special case of Eq. (10) ($\mathbf{Q}$ is the identity matrix). Then, it is easily to see that the ellipsoidal is moved to a fixed orientation for all approximated points, that is the usual orthogonal system in $\mathbb{R}^N$. Another way to see this, is that the correlation coefficients of the initial covariance matrix are same with the correlation coefficients of the $\boldsymbol{\lambda}$–dependent localized covariance matrix, this means there is not rotation. In this case, we have only dilatation.

Finally, we would like to change the localized covariance matrix at any point $\boldsymbol{u}_\ell \in \operatorname{supp}(f)$ (see Eq. (24)); i.e., to change the proper rotation matrix at any point $\boldsymbol{u}_\ell \in \operatorname{supp}(f)$. Although researchers apply it to empirical, we cannot prove it with our mathematical theory.

We need stronger conditions such that, the transformation $\boldsymbol{x} = \mathbf{T}(\boldsymbol{v}) = \mathbf{G}(\boldsymbol{x})\cdot\boldsymbol{v}$ (see Eq. (8a)) is bijection for all $\boldsymbol{x}, \boldsymbol{v} \in \mathbb{R}^N$, which is too limitative.

3.7. References

Brox, T., et al., 2007, Nonparametric density estimation with adaptive, anisotropic kernels for human motion tracking. Springer-Vergal, LNCS 4814, pp. 152-160.

Duoandikoetxea, J., 2001, Fourier analysis. Graduate Studies in Mathematics, American Mathematical Society.

Grafakos, L., 2008, Classical Fourier Analysis. Springer.

Katznelson, Y., 1968, An introduction to harmonic analysis, Cambridge University Press, Cambridge.

Korner, W.T., 1988, Fourier analysis. Cambridge University Press, Cambridge.

McShane, E.J., 1955, A theory of limits.

Sharipov, R.A., 1996, Course of Linear Algebra and Multidimensional Geometry, Textbook (2004).

Stein,E.M, 1993, Harmonic analysis real-variable methods. Princeton University Press, New Jersey.

Susarla, V., Walter, G., 1981, Estimation of a multivariate density function using delta sequences. The Annals of Statistics, Vol. 9, No.2, pp. 347-355.

Tanaka, A., et al., 2007, Integrated Kernels and Their Properties.

Vidal-Sanz, M,J., Delgado, A.M., 2004, Universal consistency of delta estimators. Ann. Inst. Statist. Math., Vol. 56, No. 4, pp. 791-818.

Vidal-Sanz, M,J., 2005, Pointwise universal consistency of nonparametric density estimators. Bernoulli, 11, pp. 971-985.

Walter, G., Blum, J., 1979, Probability density estimation using delta sequences, The Annals of Statistics, Vol. 7, No.2, pp. 328-340.

Βάρσος,Δεριζιώτης,Εμμανουήλ,Μαλιάκας,Μελάς,Τατέλη, (2009). “Μια εισαγωγή στη γραμμική άλγεβρα.

CHAPTER 4

Multidimensional Kernel Density Estimation and its Connection with Integral Approximation of Multidimensional Pdfs

Table of Contents

4.1. Introduction

In this chapter we present a method of multidimensional density estimation by delta-sequences. In Section 4.2, we give some definitions and lemmata from real analysis, and probability theory, which we shall use in the rest of chapter.

Then, in Section 4.3, we give the definitions of some general delta-sequence estimators (histogram, orthogonal series, and kernel estimators), which have attracted the attention of many researchers. The class of locally adaptive density estimators has developed, only recently. We introduce them, in Section 4.4, and we present the most important of them.

In the following two Sections, we prove the asymptotic unbiasedness, and the consistency of a delta estimator in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$ (Section 4.5), and we generalize the previous results in a Banach space (Section. 4.6).

4.2. Preliminaries

In this section, we present some definitions and lemmata from analysis and measure theory (especially, in probability theory) which we shall use in the rest of this chapter.

Definition 1: Let T be an operator on a general space of functions, $\mathrm{X} \neq \varnothing$. T is ***linear*** if satisfies:

(a) $\mathrm{T}(x+y) = \mathrm{T}(x) + \mathrm{T}(y), \quad x, y \in \mathrm{X}$

(b) $\mathrm{T}(\lambda x) = \lambda \mathrm{T}(x), \quad x \in \mathrm{X}$ and $\lambda \in \mathbb{R}$

If T, instead of (a), satisfies $\mathrm{T}(x+y) \leq \mathrm{T}(x) + \mathrm{T}(y), \quad x, y \in \mathrm{X}$ then, it is a ***sublinear*** operator. ■

Definition 2: Let a set $\mathrm{X} \neq \varnothing$. A set $\mathrm{G} \subseteq \mathrm{X}$ is called G_δ if it can be written as a countable intersection of open subsets of X. ■

Definition 3: Let $(\mathrm{X}, \mathcal{F}, \mu)$ be a measure space, μ is a $\sigma-$***finite measure*** if there exists a sequence $\{\mathrm{F}_n\} \in \mathcal{F}$, such that $\bigcup_{n=1}^{\infty} \mathrm{F}_n = \mathrm{X}$, and $\mu(\mathrm{F}_n) < \infty$, $\forall\, n \in \mathbb{N}$. ■

Definition 4: Let $(\mathrm{X}, \mathcal{F}, \mu)$ be a measure space, a nonempty set, $\mathrm{E} \in \mathcal{F}$ is said to be an atom of $\mathcal{F}$ if $\forall\, \mathrm{F} \in \mathcal{F} : \mathrm{F} \subset \mathrm{E} \Rightarrow \mathrm{F} = \varnothing$.Thus, μ is a **diffuse measure** if there are no atoms in $\mathcal{F}$. ■

Definition 5: Let μ,ν be two measures in measurable space $\left(\mathrm{X},\mathcal{F}\right)$. We say that ν is ***absolutely continuous with respect to*** μ (or dominated by μ**)**, and we write $\nu \ll \mu$, if for every set $\mathrm{F}\in\mathcal{F}$ **:** $\mu\left(\mathrm{F}\right)=0 \Rightarrow \nu\left(\mathrm{F}\right)=0$. Further, with the additional property that both measures are $\sigma-$finite, Radon-Nikodym Theorem states that there exists a $\mu-$ measurable function f taking values in $\left[0,\infty\right]$, such that for any measurable set A (i.e., $\mathrm{A}\in\mathcal{F}$) the following relation holds true

$$\nu\left(\mathrm{A}\right)=\int_{\mathrm{A}} f\, d\mu .$$

Furthermore, f is called a ***density function*** of ν or ***Radon-Nikodym derivative*** with respect to μ, and is denoted $f=\dfrac{d\nu}{d\mu}$. ■

Definition 6: Let $\left(\mathrm{X},\mathcal{F},\mu\right)$ be a measure space and f**:** $\mathrm{X}\to\mathbb{R}$ be a measurable function on set $\mathrm{X}\neq\varnothing$. In addition, we define the set of essential upper bounds of f as $\mathrm{U}_f^{ess}=\left\{a\in\mathbb{R}:\mu\left(f^{-1}\left(a,\infty\right)\right)=0\right\}$. Then, the ***essential supremum*** is defined as $\operatorname{ess\,sup} f=\inf \mathrm{U}_f^{ess}$, and if $\mathrm{U}_f^{ess}=\varnothing$ we define $\operatorname{ess\,sup} f=\infty$. Also, we can define the essential infimum similarly. ■

Below, in Lemmata 1-4, we present some basic inequalities in probability theory.

Lemma 1 (c_r Inequality)**:** Let $\left(\Omega,\mathcal{F},P\right)$ be a probability space, and X , Y**:** $\Omega\to\mathbb{R}$ be two random variables. Then, the following inequality, which is known as c_r inequality, holds true

$$\mathrm{E}\left(\,|X+Y|^r\,\right)\le c_r\left(\mathrm{E}\left(\,|X|^r\,\right)+\mathrm{E}\left(\,|Y|^r\,\right)\right)$$

where $c_r=1$, if $0<r\le 1$ and $c_r=2^{r-1}$, if $r>1$. ■

***Proof*:** if $0<r\le1$: then function $|x|^r, x\ge 0$ is a concave and increasing. Thus,

$$|x+y|^r-|x|^r=\int_x^{x+y} r\,t^{r-1}\,dt=\int_0^y r\left(x+s\right)^{r-1}ds$$

$$\le\int_0^y r\,s^{r-1}\,ds\le|y|^r .$$

If $r>1$: then $|x|^r$ is a convex function and following inequality holds true for any x,y

$$\left|\frac{x+y}{2}\right|^r \le \frac{|x|^r + |y|^r}{2}.$$

Now, taking expectations in above two cases, and the result follows. ◄

Lemma 2 (Markov Inequality)**:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $X : \Omega \to \mathbb{R}$ be a random variable. Then

$$P(X \ge a) \le \frac{\mathrm{E}(X)}{a}. \qquad \blacksquare$$

***Proof*:** Fist, we write $X \ge \alpha\, \mathrm{I}(X \ge \alpha)$, where I is the indicator function then, we derive

$$\mathrm{E}(X) \ge \mathrm{E}(\alpha\, \mathrm{I}(X \ge \alpha)) = \alpha P(X \ge \alpha) \Rightarrow P(X \ge a) \le \frac{\mathrm{E}(X)}{a}. \qquad ◄$$

Lemma 3 (Jensen Inequality)**:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $g : \mathbb{R} \to \mathbb{R}$ be a convex function, i.e., $\lambda g(x) + (1-\lambda) g(y) \ge \lambda g(\lambda x + (1-\lambda) y)$, $x, y \in \mathbb{R}$ and $\lambda \in (0,1)$. Also, let $f : \Omega \to \mathbb{R}$ be a measurable function. Assume that f, $g(f)$ are integrable then

$$g\left(\int f\, dP\right) \le \int g(f)\, dP. \qquad \blacksquare$$

***Proof*:** Durrett (2010), Section 1.5, Theorem 1.5.1. ◄

Lemma 4 (Hoeffding Inequality)**:** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and X_i, $1 \le i \le n$ be independent real valued random variables. Also, let $a_i, b_i \in \mathbb{R}$, $1 \le i \le n$, and assume that $X_i \in [a_i, b_i]$ with probability one. Then, for any $\varepsilon > 0$ the following inequality holds true

$$P\left(\left|\frac{1}{n}\sum_{i=1}^{n} (X_i - \mathrm{E}(X_i))\right| > \varepsilon\right) \le 2\exp\left\{\frac{-2n^2\varepsilon^2}{\sum_{i=1}^{n}(a_i - b_i)^2}\right\} \qquad \blacksquare$$

***Proof*:** Györfi, Kohler, Kryzak, Walk (2002), Appendix A.2, Lemma A.3, and Hoeffding (1963), Theorem 2. ◄

Lemma 5 (Borel-Cantelli)**: (a)** Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\{A_n\}_{n\in\mathbb{N}}$ be a sequence in $\mathcal{F}$, and assume that $\sum_{n=1}^{\infty} P(A_n) < \infty$. Then, $P(\limsup A_n) = 0$.

(b) Let $(\Omega, \mathcal{F}, P)$ be a probability space, and $\{A_n\}_{n\in\mathbb{N}}$ be a sequence of independent events in $\mathcal{F}$, and assume that $\sum_{n=1}^{\infty} P(A_n) = \infty$. Then, $P(\limsup A_n) = 1$. ■

Proof**: (a)** Assume the random variable $X = \sum_{n=1}^{\infty} \mathrm{I}(A_n)$ then, $\limsup A_n = \{X = \infty\}$. Thus

$$\mathrm{E}(X) = \sum_{n=1}^{\infty} \mathrm{E}(\mathrm{I}(A_n)) = \sum_{n=1}^{\infty} P(A_n) < \infty.$$

Also, we know that if $X \geq 0$ and $\mathrm{E}(X) < \infty$ then $P\{X = \infty\} = 0$. It implies $P(\limsup A_n) = 0$.

(b) $P\left(\{\limsup A_n\}^c\right) = P\left(\bigcup_{n=1}^{\infty} \bigcap_{k=n}^{\infty} A_k\right) = \lim_{n\to\infty} P\left(\bigcap_{k=n}^{\infty} A_k^c\right)$, the last equation holds true because the sequence $\left\{\bigcap_{k=n}^{\infty} A_k^c\right\}$ is increasing for any $n \geq 1$. Now, for $n \geq 1$ we have

$$P\left(\bigcap_{k=n}^{\infty} A_k^c\right) = \lim_{m\to\infty} P\left(\bigcap_{k=n}^{m} A_k^c\right) = \lim_{m\to\infty} \prod_{k=n}^{m} P\left(A_k^c\right) = \lim_{m\to\infty} \prod_{k=n}^{m} \left(1 - P\left(A_k\right)\right)$$

$$\leq \lim_{m\to\infty} \prod_{n=k}^{m} \exp\left\{-P\left(A_k\right)\right\} = \lim_{m\to\infty} \exp\left\{\sum_{n=k}^{m} -P\left(A_k\right)\right\}$$

$$= \exp\left\{\sum_{n=k}^{\infty} -P\left(A_k\right)\right\} = 0.$$

Finally, $P\left(\{\limsup A_n\}^c\right) = 0 \Rightarrow 1 - P\left(\{\limsup A_n\}\right) = 0 \Rightarrow P\left(\{\limsup A_n\}\right) = 1$. ◄

4.3. Delta-sequence estimators

There are a lot of examples of delta-sequence estimators in the literature. Histograms, orthogonal series, and kernel density estimators are well-known examples. Below, we give the definition and some examples of a delta sequence estimator.

Definition 7: A sequence of functions $K_m : \mathbb{R}^N \times \mathbb{R}^N \rightarrow \mathbb{R}$ is called a ***delta-sequence*** if $\forall f \in C(\mathbb{R}^N)$ (i.e., for each continuous function on $\mathbb{R}^N$) and $\boldsymbol{x} \in \mathbb{R}^N$

$$\int_{\mathbb{R}^N} K_{m_n}(\boldsymbol{x},\boldsymbol{y})\, f(\boldsymbol{x})\, d\boldsymbol{y} \rightarrow f(\boldsymbol{x}), \quad \text{as} \quad n \rightarrow \infty, \qquad \blacksquare \tag{1}$$

where m_n belongs in a directed set. We shall refer to the functions K_{m_n} as the kernels (functions).

Definition 8: Let $\{\boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n\}$ be independent observations from an unknown distribution F on $\mathbb{R}^N$ with probability density function (pdf) f. A ***delta-sequence estimator*** can be written in the form

$$\hat{f}_n(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i). \qquad \blacksquare \tag{2}$$

Definition 9: Let $\{x_1, x_2, \ldots, x_n\}$ be independent observations from an unknown distribution F on $\mathbb{R}$ with pdf f, and $\mathrm{I}_j(x) : \mathbb{R} \rightarrow \{0,1\}$ be the indicator function for the bin $[(j-1)h, jh)$, $j \in \mathbb{N}$ and $h > 0$ (i.e., $\mathrm{I}_j(x) = 1$, if $x \in [(j-1)h, jh)$ and $\mathrm{I}_j(x) = 0$, otherwise). Then the ***Histogram estimator***, with equal binwidths and in one dimension, given by

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=-\infty}^{\infty} \mathrm{I}_j(x)\,\mathrm{I}_j(x_i). \qquad \blacksquare \tag{3}$$

Definition 10: Let $\{x_1, x_2, \ldots, x_n\}$ be independent observations from an unknown distribution F on $\mathbb{R}$ with pdf f, and $\{\Phi_j(x)\}_{j=1}^{m}$ be a orthonormal system on $[a,b] \subseteq \mathbb{R}$ consisting of eigenfunctions of a compact operator on $L^2[a,b]$. Then the ***Orthogonal series estimator*** given by

$$\hat{f}_n(x) = \frac{1}{n}\sum_{i=1}^{n}\sum_{j=1}^{m} \Phi_j(x)\,\Phi_j(x_i). \qquad \blacksquare \tag{4}$$

Definition 11: Let $\{ \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \}$ be a $N-$variate random sample, and $K_m : \mathbb{R}^N \to \mathbb{R}$ be a kernel with the properties $K_m \geq 0$ and $\int_{\mathbb{R}^N} K_m(\boldsymbol{x})\, d\boldsymbol{x} = 1$, then a ***kernel estimator*** given by

$$\hat{f}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_m(\boldsymbol{x} - \boldsymbol{x}_i). \qquad \blacksquare \tag{5}$$

For more details for theses estimators see Nolan, Marron (1989), Sec. 1, Walter, Blum (1979), Sec. 1.

4.4. Locally adaptive density estimators

Let $\{ \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \}$ be a $N-$variate random sample with unknown pdf f. The simplest multivariate kernel density estimator given by

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n \cdot h^N} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h} \right) = \frac{1}{n} \sum_{i=1}^{n} K_h(\boldsymbol{x} - \boldsymbol{x}_i), \tag{6a}$$

by setting $K_h(\boldsymbol{x}) = \frac{1}{h^N} K\left(\frac{\boldsymbol{x}}{h} \right).$ (6b)

Where $K : \mathbb{R}^N \to \mathbb{R}$ is a N-variate density function, and $h > 0$ is the smoothing parameter (bandwidth or window width), which controls the size of kernel, and in this case is held constant for all evaluation points $\boldsymbol{x} \in \mathbb{R}^N$. This estimator is called ***fixed***. □

It is more interesting the smoothing parameter (bandwidth) based on the evaluation point or on the random sample. These estimators are called <u>*locally adaptive density estimators*</u>.

So, we can generalize Eqs. (6a,b), assuming that the bandwidth is based on the point $\boldsymbol{x} \in \mathbb{R}^N$, i.e., $h_{\boldsymbol{x}} = h(\boldsymbol{x})$, and in this case we obtain the ***balloon*** estimator.

$$\hat{f}_B(\boldsymbol{x}) = \frac{1}{n \cdot h_{\boldsymbol{x}}^N} \sum_{i=1}^{n} K\left(\frac{\boldsymbol{x} - \boldsymbol{x}_i}{h_{\boldsymbol{x}}} \right) = \frac{1}{n} \sum_{i=1}^{n} K_{h_{\boldsymbol{x}}}(\boldsymbol{x} - \boldsymbol{x}_i). \tag{7}$$

The balloon estimator was first introduced by Loftsgaarden, Quesenberry (1965) in the form of the $k-th$ nearest neighbor estimator, which can be written in the form of Eq. (8) by using a constant kernel and setting $h(\boldsymbol{x}) = d_k(\boldsymbol{x})$ where $d_k(\boldsymbol{x})$ returns the distance to the $k-th$ nearest data point to $\boldsymbol{x}$. □

With analogous way, we can assume that the smoothing parameter is based on the random sample $\{ \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \}$. We define the ***sample point estimator*** which is equal with

$$\hat{f}_S(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{h_{\boldsymbol{x}_i}^N} K\left(\frac{\boldsymbol{x}-\boldsymbol{x}_i}{h_{\boldsymbol{x}_i}}\right) = \frac{1}{n}\sum_{i=1}^{n} K_{h_{\boldsymbol{x}_i}}\left(\boldsymbol{x}-\boldsymbol{x}_i\right). \tag{8}$$

This type of estimator was introduced by Breiman et al. (1977). □

As we saw the smoothing parameter h controls the size of kernel. One constant smoothing parameter corresponds to spherically symmetric kernels. We can generalize this parameter to a vector $\boldsymbol{h} = (h_1, h_2, \ldots, h_N)$, $h_i > 0$, $0 \le i \le N$, and more general to a symmetric and positive definite matrix $\mathbf{H}$ that is analogous to the covariance matrix of K. With this way, we generalize the above estimators. (i.e., Eqs. (6a), (7), and (8))

The ***generalized fixed estimator*** given by

$$\hat{f}(\boldsymbol{x}) = \frac{1}{n \cdot |\mathbf{H}|^{1/2}} \sum_{i=1}^{n} K\left(\mathbf{H}^{-1/2} \cdot (\boldsymbol{x}-\boldsymbol{x}_i)\right) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}}\left(\boldsymbol{x}-\boldsymbol{x}_i\right), \tag{9a}$$

by setting $K_{\mathbf{H}}(\boldsymbol{x}) = \dfrac{1}{|\mathbf{H}|^{1/2}} K_{\mathbf{H}(\boldsymbol{x})}\left(\mathbf{H}^{-1/2} \cdot \boldsymbol{x}\right).$ (9b)

where, $|\cdot|$ indicates the determinant of a matrix.

The ***generalized balloon estimator*** given by

$$\hat{f}_B(\boldsymbol{x}) = \frac{1}{n \cdot |\mathbf{H}(\boldsymbol{x})|^{1/2}} \sum_{i=1}^{n} K\left(\mathbf{H}(\boldsymbol{x})^{-1/2} \cdot (\boldsymbol{x}-\boldsymbol{x}_i)\right) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}(\boldsymbol{x})}\left(\boldsymbol{x}-\boldsymbol{x}_i\right). \tag{10}$$

And the ***generalized sample-point estimator*** given by

$$\hat{f}_S(\boldsymbol{x}) = \frac{1}{n}\sum_{i=1}^{n} \frac{1}{|\mathbf{H}(\boldsymbol{x}_i)|^{1/2}} K\left(\mathbf{H}(\boldsymbol{x}_i)^{-1/2} \cdot (\boldsymbol{x}-\boldsymbol{x}_i)\right) = \frac{1}{n}\sum_{i=1}^{n} K_{\mathbf{H}(\boldsymbol{x}_i)}\left(\boldsymbol{x}-\boldsymbol{x}_i\right) \tag{11}$$

For the latter, we can assume that we divide the support of the density in m bins, then the ***binned sample-point estimator*** given by

$$\hat{f}_{BS}(\boldsymbol{x}) = \frac{1}{n}\sum_{j=1}^{m} \frac{n_j}{|\mathbf{H}(\boldsymbol{t}_j)|^{1/2}} K\left(\mathbf{H}(\boldsymbol{t}_j)^{-1/2} \cdot (\boldsymbol{x}-\boldsymbol{t}_j)\right) = \frac{1}{n}\sum_{j=1}^{m} n_j K_{\mathbf{H}(\boldsymbol{t}_j)}\left(\boldsymbol{x}-\boldsymbol{t}_j\right), \tag{12}$$

where n_j is the number of data points in the $j-th$ bin, (i.e., $\sum_{j=1}^{m} n_j = n$), $\boldsymbol{t}_j$ is the center of the $j-th$ bin, and $\mathbf{H}(\boldsymbol{t}_j)$ is the smoothing matrix associated with the $j-th$ bin. In general, an equally spaced mesh of points is laid down over the support of the density to define the bins, although other binning rules such as the linear binning defined in Hall, Wand (1996) could be considered. □

Generally, we can summarize all the above estimators, defining three families of matrices, depending on the form of matrix $\mathbf{H}$, for Eqs. (9a), (10), (11), and (12).

$\mathcal{H}_1 = \left\{ \mathbf{H} = h^2 \cdot \mathbf{I}_N , h > 0 \right\}$, which correspond to spherically symmetric kernels with only one smoothing parameter.

$\mathcal{H}_2 = \left\{ \mathbf{H} = \mathrm{diag}\left(h_1^2, h_2^2, \ldots, h_N^2 \right), \boldsymbol{h} > 0 \right\}$, which correspond to ellipsoidal kernels with N smoothing parameters.

$\mathcal{H}_3 = \left\{ \mathbf{H} \in \mathbb{R}^{N\times N}, \mathbf{H} \text{ is symmetric and positive difinite} \right\}$, which is the most general and allows ellipsoidal kernels of arbitrary orientation with $N\,(\mathrm{N}+1)\,/\,2$ parameters.
Sain (2002) in Sec 1 and 3 gives a discussion about the above three classes of matrices and numerical comparison among of the estimators. □

A lot of researchers focused on finding the optimal choice for the smoothing parameters that minimize the Mean Integrated Square Error (MISE) for the above estimators

$$MISE = \mathrm{E} \int_{\mathbb{R}^N} \left(\hat{f}(x) - f(x) \right)^2 dx = \int_{\mathbb{R}^N} \mathrm{E} \left(\hat{f}(x) - f(x) \right)^2 dx = \int_{\mathbb{R}^N} MSE\, dx,$$

For more details about MISE, Asymptotic MISE (AMISE), and the optimal selection of bandwidth see Want, Jones (1994,1995), Shirahata, Chu (1992), Sain, Scott (1996), Sain (2002).

4.5. Pointwise consistency of delta estimators in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$

In this section we assume that P is a probability measure in $\left(\mathbb{R}^N, \mathfrak{B}^N \right)$ which is absolutely continuous with respect to a $\sigma-$finite measure μ and $f = \dfrac{dP}{d\mu}$ is the Radon-Nikodym derivate. In addition, we assume that $f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$. Supposing that μ is the Lebesgue measure, then f is the usual probability density function (pdf) but it is instructive to assume a general $\sigma-$finite measure μ; see Vidal-Sanz (2005), Sec. 1.

Let $\left\{ \boldsymbol{x}_1, \boldsymbol{x}_2, \ldots, \boldsymbol{x}_n \right\}$ be a random sample from P, a delta estimator of f can be written in the following form (see Def. 7 of Sec. 4.3)

$$\hat{f}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{m_n}\left(\boldsymbol{x}, \boldsymbol{x}_i\right), \tag{14}$$

where the smoothing sequence m_n belongs to a directed set $\left(L, \prec \right)$, and K_{m_n} is a net in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$. For example, m_n may be a sequence of positive definite matrices ordered by decreasing a norm in kernel estimation of multivariate densities. In this case, Eq. (14) take the form:

$$\hat{f}_n(\boldsymbol{x}) = \frac{1}{n} \sum_{i=1}^{n} \frac{1}{\left| m_n \right|} K\left(\left(m_n\right)^{-1} \cdot \left(\boldsymbol{x} - \boldsymbol{x}_i \right) \right). \tag{15}$$

Definition 12: Let μ be a σ-finite measure in $\left(\mathbb{R}^N, \mathfrak{B}^N \right)$, and P a probability measure satisfying $P \ll \mu$, i.e., P is absolutely continuous with respect to μ; see Def.5 of Sec. 4.2. We say that a delta estimator $\hat{f}_n$ is ***strongly (or weakly) consistent*** almost everywhere (a.e.) with respect to μ, if $\left| \hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}) \right| \xrightarrow{n \to \infty} 0$ almost surely (or in probability), for almost every $\boldsymbol{x} \in \mathbb{R}^N$ with respect to the measure μ. We say that the convergence is universal when it holds for every P such that, $P \ll \mu$. ■

Using triangular inequality we derive

$$\left| \hat{f}_n(\boldsymbol{x}) - f(\boldsymbol{x}) \right| \le \left| \mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) - f(\boldsymbol{x}) \right| + \left| \hat{f}_n(\boldsymbol{x}) - \mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) \right|. \tag{16}$$

The deterministic term, $\mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) - f(\boldsymbol{x})$, is called ***bias term***, and the stochastic term, $\hat{f}_n(\boldsymbol{x}) - \mathrm{E}\left(\hat{f}(\boldsymbol{x}) \right)$, is called ***variation term***. In the sequel, we study the convergence to zero of each term separately.

4.5.1. Convergence of bias term

In this section we study bias problem, we want to conclude that the expected value of $\hat{f}_n(\boldsymbol{x})$ convergences to f almost everywhere, for any $f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$.

The expected value of $\hat{f}_n(\boldsymbol{x})$ is easily calculated and reads as follow

$$a_n(f)(\boldsymbol{x}) = \mathrm{E}\left(\hat{f}_n(\boldsymbol{x}) \right) = \int_{\mathbb{R}^N} K_{m_n}\left(\boldsymbol{x}, \mathbf{z} \right) f(\mathbf{z}) \mu(dz). \tag{17}$$

Eq. (17) defines the linear operator $a_n : L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right) \to C\left(\mathbb{R}^N \right)$.

Definition 13: Let a_n be a linear operator on $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$. We say that a_n is ***bounded in measure***, if for every $\varepsilon > 0$, there exists $\delta > 0$, such that

$$\sup_{\| f \|_{L_1} < 1} \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \left| a_n(f)(\boldsymbol{x}) \right| > \delta \right\} \right) \leq \varepsilon \tag{18a}$$

Thus, a sequence of linear operators, $\{a_n\}_{n \in \mathbb{N}}$, is uniformly bounded in measure if the maximal operator $a^M(f)(\boldsymbol{x}) = \sup_{n \in \mathbb{N}} \left| a_n(f)(\boldsymbol{x}) \right|$ is such that, for every $\varepsilon > 0$, there exists $\delta > 0$ such that

$$\sup_{\| f \|_{L_1} < 1} \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M(f)(\boldsymbol{x}) > \delta \right\} \right) \leq \varepsilon . \tag{18b}$$

Notice that maximal operator is a sublinear operator and not a linear one. ■

Bellow, we shall present a type result of one of fundamental theorems in functional analysis, the uniform boundedness principle.

Theorem 1 (Pointwise uniform boundedness, Vidal-Sanz (2005), Sec. 2, Theorem 1)**:** Let $\{a_n\}_{n \in \mathbb{N}}$ be a sequence of linear operators on $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$ each of them being bounded in measure. Then, only one of the following statements holds true:

(a) $\{a_n\}_{n \in \mathbb{N}}$ is uniformly bounded in measure.

(b) For every $\varepsilon > 0$, there exists a set $C_\varepsilon \subseteq L_1 \equiv L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$, which is G_δ and dense in L_1, such that for every $f \in C_\varepsilon$: $\mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M(f)(\boldsymbol{x}) = \infty \right\} \right) > \varepsilon$

■

To prove this theorem, we shall need the following lemma

Lemma 6: For any $\varepsilon > 0$ and $\delta > 0$ we define the set

$$V_\varepsilon^\delta = \left\{ f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right) : \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M(f)(\boldsymbol{x}) > \delta \right\} \right) > \varepsilon \right\}. \tag{a}$$

Then, V_ε^δ is an open subset in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$. Hence, for any sequence $(\delta_k)_{k \in \mathbb{N}}$, $\delta_k > 0$, $\left\{ V_\varepsilon^{\delta_k} \right\}_{k \in \mathbb{N}}$ is a sequence of open sets for any $\varepsilon > 0$. ■

***Proof of Lemma 6*:** Vidal-Sanz (2005), Sec. 2, Theorem 2. ◄

***Proof of Theorem 1*:** Let us first assume that **(b)** is not hold true. Assume there exists a $k \in \mathbb{N}$, such that, $V_{\varepsilon}^{\delta_k}$ is not dense in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$. Then, there exists an $f_0 \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$ and $r > 0$ such that for $f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$: $\| f \| \le r \Rightarrow$ $\left(f_0 + f\right) \notin V_{\varepsilon}^{\delta_k}$, where $\| \cdot \|$ indicates the norm of $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$. So, we have $\mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M\left(f_0 + f\right)(\boldsymbol{x}) > \delta_k \right\} \right) \le \varepsilon$. In addition, we can write $f = \left(f_0 + f\right) - f_0$, and using the fact that the maximal operator is sublinear (see Def. 1) we obtain

$$\mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M(f)(\boldsymbol{x}) > 2\delta_k \right\} \right) \le \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M\left(f_0 + f\right)(\boldsymbol{x}) > \delta_k \right\} \right) + \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M\left(f_0\right)(\boldsymbol{x}) > \delta_k \right\} \right) \le 2\varepsilon$$

Finally, $$\sup_{\| f \| \le r} \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \left| a_n(f)(\boldsymbol{x}) \right| > \delta \right\} \right) \le \frac{2\varepsilon}{r} \qquad \text{(b)}$$

That is, a^M uniformly bounded in measure (see Def. 13) with $\varepsilon' = \dfrac{2\varepsilon}{r}$ and $\delta = 2\delta_k$, i.e., **(a)** holds true.

Now, if every $V_{\varepsilon}^{\delta_k}$ is dense in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$, the application of Baire theorem (see Rudin (1987), Chapter 5, Theorem 5.6) implies that $C = \bigcap_{k=1}^{\infty} V_{\varepsilon}^{\delta_k}$ is a dense and, G_{δ} set (see Def. 2). If $f \in C$, then for any $\varepsilon > 0$ and $\delta_k > 0$ we have

$\mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : a^M(f)(\boldsymbol{x}) > \delta_k \right\} \right) > \varepsilon$. In other words, **(b)** holds true, and the same time **(a)** does not hold. ◄

Our aim is to show that the linear operators $\left\{a_n\right\}_{n \in \mathbb{N}}$ is approximate identity in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$. The following theorem shows that it is enough to prove this property only for a dense subset of $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$.

Theorem 2 (Pointwise approximation, Vidal-Sanz (2005), Sec. 2, Theorem 2)**:** Let $\left\{a_n\right\}_{n \in \mathbb{N}}$ be a sequence of linear operators in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$, and assume that

(a) the sequence $\left\{a_n\right\}_{n \in \mathbb{N}}$ is uniformly bounded in measure

(b) there exists a dense set, $G \subseteq L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$, such that for every $\tilde{f} \in G$ **:**

$$a_n\left(\tilde{f}\right) \xrightarrow{n\to\infty} a\left(\tilde{f}\right) \text{ a.e.}$$

Then $\left\{a_n\right\}_{n\in\mathbb{N}}$ is an approximation to the identity in the a.e sense, i.e., $a_n(f) \xrightarrow{n\to\infty} a(f)$ a.e., for every $f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$. ■

***Proof*:** Assume that there exists a dense set, $G \subseteq L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$, such that for every $\tilde{f} \in G$ and $\delta > 0$

$$\lim_{n\to\infty} \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})\right| > \delta \right\} \right) = 0 \qquad \text{(a)}$$

Also, by assumption G is dense so, for any $\varepsilon > 0$ and $f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$ there exists $\tilde{f} \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu\right)$, such that, $\left\| f - \tilde{f} \right\| \leq \varepsilon$ (b)

Now, using the triangular inequality we derive:

$$\sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - f(\boldsymbol{x})\right| \leq \sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - a_{n'}\left(\tilde{f}\right)(\boldsymbol{x})\right|$$
$$+ \sup_{n'>n} \left| a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})\right| + \left| \tilde{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right| \qquad \text{(c)}$$

for $\boldsymbol{x} \in \mathbb{R}^N$, and $n \in N$. Observing that if $\sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - f(\boldsymbol{x})\right| > \delta$ then,

$$\sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - a_{n'}\left(\tilde{f}\right)(\boldsymbol{x})\right| + \sup_{n'>n} \left| a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x})\right| + \left| \tilde{f}(\boldsymbol{x}) - f(\boldsymbol{x})\right| > \delta,$$

by Eq. (c), and each term is positive. Hence, at least one of them must be higher than $\frac{\delta}{3}$.

So, $\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - f(\boldsymbol{x})\right| > \delta \right\}$

$$\subseteq \left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - a_{n'}\left(\tilde{f}\right)(\boldsymbol{x})\right| > \frac{\delta}{3} \right\}$$

$$\bigcup\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x}) \right| > \frac{\delta}{3} \right\}$$

$$\bigcup\left\{ \boldsymbol{x} \in \mathbb{R}^N : \left| \tilde{f}(\boldsymbol{x}) - f(\boldsymbol{x}) \right| > \frac{\delta}{3} \right\} \tag{d}$$

It follows

$$\mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - f(\boldsymbol{x}) \right| > \delta \right\} \right)$$

$$\begin{aligned} &\le \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) \right| > \frac{\delta}{3} \right\} \right) \\ &+ \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x}) \right| > \frac{\delta}{3} \right\} \right) \\ &+ \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \left| \tilde{f}(\boldsymbol{x}) - f(\boldsymbol{x}) \right| > \frac{\delta}{3} \right\} \right) \end{aligned} \tag{e}$$

Using the assumption **(a),** we can show that first term of (d) can be arbitrary small, i.e.,

$$\mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) \right| > \frac{\delta}{3} \right\} \right) \le \varepsilon_1, \tag{f}$$

for $\varepsilon_1 > 0$ and ε_1 can be arbitrary small.

Then, we have

$$\mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}(f)(\boldsymbol{x}) - f(\boldsymbol{x}) \right| > \delta \right\} \right)$$

$$\begin{aligned} &\le \varepsilon_1 + \mu\left(\left\{ \boldsymbol{x} \in \mathbb{R}^N : \sup_{n'>n} \left| a_{n'}\left(\tilde{f}\right)(\boldsymbol{x}) - \tilde{f}(\boldsymbol{x}) \right| > \frac{\delta}{3} \right\} \right) \\ &+ \frac{3\left\| \tilde{f}(\boldsymbol{x}) - f(\boldsymbol{x}) \right\|}{\delta} \end{aligned} \tag{g}$$

And finally,

$$\mu\left(\left\{\boldsymbol{x}\in\mathbb{R}^N : \sup_{n'>n}\left|a_{n'}(f)(\boldsymbol{x})-f(\boldsymbol{x})\right|>\delta\right\}\right)$$
$$\leq \varepsilon_1+\mu\left(\left\{\boldsymbol{x}\in\mathbb{R}^N : \sup_{n'>n}\left|a_{n'}\left(\tilde{f}\right)(\boldsymbol{x})-\tilde{f}(\boldsymbol{x})\right|>\frac{\delta}{3}\right\}\right) \qquad \text{(h)}$$
$$+\frac{3\varepsilon}{\delta}$$

In latter inequality $\varepsilon,\varepsilon_1$ are arbitrary small and the second term tend to zero by In. (a). So, a.e approximation follows. ◄

Remark 1: The assumption **(a)** of Theorem 2 is necessary, and we cannot assume weaker condition, i.e., all a_n are bounded in measure. To see this, we assume that $\{a_n\}_{n\in\mathbb{N}}$ is an approximation to the identity in $L_1\left(\mathbb{R}^N,\mathfrak{B}^N,\mu\right)$, and all a_n are bounded in measure but the uniform boundedness in measure is not satisfied. It follows that $\{a_n\}_{n\in\mathbb{N}}$ is an approximation to the identity in every dense $C\subseteq L_1\left(\mathbb{R}^N,\mathfrak{B}^N,\mu\right)$. Then, using Theorem 1, we can prove that there exists a dense set C_ε, such that for any $\varepsilon>0$ and $f\in C_\varepsilon$:

$$\mu\left(\left\{\boldsymbol{x}\in\mathbb{R}^N : \lim_{n\in\mathbb{N}}\left|a_n(f)(\boldsymbol{x})-(f)(\boldsymbol{x})\right|=\infty\right\}\right)>\varepsilon,$$

which contradicts the a.e approximation property; see Pointwise approximation, Vidal-Sanz (2005), Sec. 2, Theorem 2.

Now, we are in position to state sufficient conditions for pointwise approximation of expected value of a delta estimator.

Theorem 3 (Sufficient conditions for pointwise approximation, Vidal-Sanz (2005), Sec. 2, Theorem 3)**:** Let $a_n(f)(\boldsymbol{x})=\mathrm{E}\left(\hat{f}_n(\boldsymbol{x})\right)=\int_{\mathbb{R}^N}K_{m_n}(\boldsymbol{x},\mathbf{z})f(\mathbf{z})\mu(dz)$. We define $|a_n|=\int\left|K_{m_n}(\boldsymbol{x},\mathbf{z})\right|f(\mathbf{z})\mu(d\mathbf{z})$. Assume that

(a) the sequence $|a_n|$ is uniformly bounded in measure.

(b) $\int_{\mathbb{R}^N}K_{m_n}(\boldsymbol{x},\mathbf{z})\mu(d\mathbf{z})\xrightarrow{n\to\infty}1$ a.e.

(c) For every $\delta>0$ there exists M_δ, such that $\sup_n\int_{\|\boldsymbol{x}-\mathbf{z}\|\leq\delta}\left|K_{m_n}(\boldsymbol{x},\mathbf{z})\right|\mu(d\mathbf{z})<M_\delta$ a.e.

(d) For every $\delta > 0$ $\displaystyle\int_{\| \mathbf{x}-\mathbf{z} \| > \delta} \left| K_{m_n}(\mathbf{x},\mathbf{z}) \right| \mu(d\mathbf{z}) \xrightarrow{n \to \infty} 0$ a.e.

Then, $a_n(f) \xrightarrow{n \to \infty} a(f)$ a.e. for every $f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$ ■

Proof: From the fact that $\left\{ \left| a_n \right| \right\}_{n \in \mathbb{N}}$ is uniform bounded in measure we can prove that $\left\{ a_n \right\}_{n \in \mathbb{N}}$ is uniform bounded in measure. This is observing,

$$a^M(f)(\mathbf{x}) = \sup_{n \in \mathbb{N}} \left| a_n(f)(\mathbf{x}) \right| \le \sup_{n \in \mathbb{N}} \int \left| K_{m_n}(\mathbf{x},\mathbf{z}) f(\mathbf{z}) \right| \mu(d\mathbf{z}) = |a|^M(|f|)(\mathbf{x}),$$

and then for any $\delta > 0$

$$\mu\left(\left\{ \mathbf{x} \in \mathbb{R}^N : a^M(f)(\mathbf{x}) > \delta \right\} \right) \le \mu\left(\left\{ \mathbf{x} \in \mathbb{R}^N : |a|^M(|f|)(\mathbf{x}) > \delta \right\} \right), \qquad \text{(a)}$$

which proves that $\left\{ a_n \right\}_{n \in \mathbb{N}}$ is uniformly bounded in measure.

In the sequel, we shall give the idea of proof. It is proved the approximation property for any $f \in L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$ with an a.e. identical elements in $C_c\left(\mathbb{R}^N \right)$ (the set of continuous and compactly supported functions on $\mathbb{R}^N$). Then, as $C_c\left(\mathbb{R}^N \right)$ is dense in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \mu \right)$, applying Theorem 2 the result follows. ◄

Remark 2: The idea and the manipulations of proof of Theorem 3 are similar with Theorem 2 in Chapter 3, and for this reason we omit the last part of proof. This is, why we work in set of continuous compactly supported functions, and every continuous function defined on a compact set is bounded.

Remark 3: In Theorems 2 and 3 we assume a general delta estimator, and we proved them under some assumptions. Now, we shall give some examples of delta estimators which satisfy these assumptions.

Example 1: Let λ be the Lebesgue measure, and I_C be the indicator function of a set, $C \subseteq \mathbb{R}^N$. Consider the kernel estimator

$$\hat{f}_{m_n}(\mathbf{x}) = \frac{1}{n} \sum_{i=1}^{n} K_{m_n}(\mathbf{x} - \mathbf{x}_i) = \frac{1}{n} \frac{1}{\det(m_n)} \sum_{i=1}^{n} K\left(m_n^{-1}(\mathbf{x} - \mathbf{x}_i) \right)$$

in $L_1\left(\mathbb{R}^N, \mathfrak{B}^N, \lambda \right)$. If there exist a closed interval $C \subseteq \mathbb{R}^N$, and positive constants c_1, c_2, such that $c_1 \mathrm{I}_C(\boldsymbol{u}) \le |K(\boldsymbol{u})| \le c_2 \mathrm{I}_C(\boldsymbol{u})$, then the operators $a_n(f)(\mathbf{x}) = \mathrm{E}\left(\hat{f}_n(\mathbf{x}) \right)$ (see Eq. (17)) are uniformly bounded in measure. Now, let $\| \cdot \|$ be a matrix norm, such that

$\| A\cdot B\| \le \| A\| \| B\|, \quad A,B\in\mathbb{R}^{N\times N}$. Then, using the Theorem of change of variables we obtain:

$$\int_{\| \boldsymbol{x}-\mathbf{z}\| >\delta} \left| K_{m_n}(\boldsymbol{x}-\mathbf{z})\right| d\mathbf{z} \le \frac{1}{\delta \det(m_n)} \int_{\mathbb{R}^N} \| \boldsymbol{x}-\mathbf{z}\| K\left(m_n{}^{-1}(\boldsymbol{x}-\mathbf{z})\right) d\mathbf{z}$$

$$\le \frac{\| m_n\|}{\delta} \int_C \| \boldsymbol{u}\| K(\boldsymbol{u}) d\boldsymbol{u} \xrightarrow{\| m_n\| \to 0} 0.$$

Now, the approximation property follows applying Theorem 3.

4.5.2. Convergence of variation term

In the following lemma we shall show the universal weak consistency of variation term.

Lemma 7 (Universal pointwise weak consistency of variation term, Vidal-Sanz (2005), Sec. 3, Preposition 1)**:** Assume that for any probability measure, P, such that $f=\frac{dP}{d\mu}$ $\in L_1\left(\mathbb{R}^N,\mathfrak{B}^N,\mu\right)$, and $\mathrm{E}\left(\left|K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)\right|^r\right)=o\left(n^{r-1}\right)$ for some $r>1$ almost everywhere with respect to μ. Then, $\mathrm{E}\left(\left|\hat{f}_n(\boldsymbol{x})-\mathrm{E}\left(\hat{f}(\boldsymbol{x})\right)\right|^r\right)\xrightarrow{n\to\infty} 0$, and $\left|\hat{f}_n(\boldsymbol{x})-\mathrm{E}\left(\hat{f}(\boldsymbol{x})\right)\right|\xrightarrow{P} 0$ a.e. with respect to μ (19a,b), and the result hold universally in P. ■

***Proof*:** Applying the c_r equation (Lemma1) we have

$$\mathrm{E}\left(\left|\hat{f}_n(\boldsymbol{x})-\mathrm{E}\left(\hat{f}(\boldsymbol{x})\right)\right|^r\right)=\mathrm{E}\left(\left|\frac{1}{n}\sum_{i=1}^{n}\left[K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)-\mathrm{E}\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)\right)\right]\right|^r\right)$$

$$\le \frac{2^{r-1}}{n^r}\sum_{i=1}^{n}\mathrm{E}\left(\left|\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)-\mathrm{E}\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)\right)\right)\right|^r\right)$$

$$\le \frac{2^{r}}{n^r}\sum_{i=1}^{n}\mathrm{E}\left(\left|\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)\right)\right|^r\right) \xrightarrow{n\to\infty} 0, \qquad \text{(a)}$$

which proves Eq. (19a).

Now, applying Markov inequality (Lemma 2) we prove Eq. (19b).

$$P\left(\left|\ \hat{f}_n(\boldsymbol{x})-\mathrm{E}\left(\hat{f}(\boldsymbol{x})\right)\right|>\varepsilon\right)\le\frac{\mathrm{E}\left(\left|\ \hat{f}_n(\boldsymbol{x})-\mathrm{E}\left(\hat{f}(\boldsymbol{x})\right)\right|\right)}{\varepsilon}\xrightarrow{n\to\infty}0. \qquad \text{(b)}$$

◄

Theorem 4 (Universal pointwise strong consistency of variation term, Vidal-Sanz (2005), Sec. 3, Theorem 4)**:** Assume that for any probability measure, P, such that $f=\frac{dP}{d\mu}\in L_1\left(\mathbb{R}^N,\mathfrak{B}^N,\mu\right)$, and $\sum_{n=1}^{\infty}\exp\left\{\frac{-n}{M_n(\boldsymbol{x})^2}\right\}<\infty$ a.e. with respect to μ, where $M_n(\boldsymbol{x})=\operatorname{ess\,sup}_{\mathbf{z}}\left|K_{m_n}(\boldsymbol{x},\mathbf{z})\right|$. Then universal pointwise convergence (see Def. 12) is satisfied a.e. universally in P. ■

Proof**:** First, by assumption, we know that $K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)\in\left[-M_n(\boldsymbol{x}),M_n(\boldsymbol{x})\right]$**,** and applying the Hoeffding Inequality (Lemma 4) we derive

$$P\left(\left|\frac{1}{n}\sum\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)-\mathrm{E}\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)\right)\right)\right|>\varepsilon\right)\le 2\exp\left\{\frac{-2n^2\varepsilon^2}{\sum_{i=1}^{n}\left(2M_n(\boldsymbol{x})\right)^2}\right\}$$

$$=2\exp\left\{\frac{-n\varepsilon^2}{2M_n(\boldsymbol{x})^2}\right\}. \qquad \text{(a)}$$

Further, by Borrel-Canteli Lemma (Lemma 5a), we obtain

$$P\left(\limsup_{n}\left|\frac{1}{n}\sum\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)-\mathrm{E}\left(K_{m_n}(\boldsymbol{x},\boldsymbol{x}_i)\right)\right)\right|>\varepsilon\right)=0, \qquad \text{(b)}$$

which is equivalent with universal pointwise convergence (see Def. 12). ◄

Example 2: Consider the kernel estimator, and assume that kernel K has a global maximum at point, $u=u_0$. Then, we have $M_n(\boldsymbol{x})=\sup_{\mathbf{z}}\left|K_{m_n}(\boldsymbol{x}-\mathbf{z})\right|=\frac{K(u_0)}{\det(m_n)}$, and we require $\sum_{n=1}^{\infty}\exp\left\{-n\det(m_n)^2\right\}<\infty$. A sufficient condition for latter inequality to holds true is $\frac{n\det(m_n)^2}{\log(n)}\to\infty$. Applying Theorem 4 the universal pointwise strong consistency follows.

4.6. Estimation in infinite dimensional spaces

In this section we present the main points of some recent works about estimation in infinite dimensional spaces; see Prakasa-Rao (2010), Niang (2004), Ferraty, Vieu (2008). Before continuing, it is necessary to give a description of the concept of data in an infinite dimensional space.

There are different fields of applied sciences (environmetrics, chemometrics, biometrics, medicine, econometrics, engineering, etc.) where the collected data are curves. These data are called functional. A functional observation can be expressed by a random family $\{X(t), t \in \mathrm{T}\}$ where T be a linearly ordered set of indices (e.g., $\mathrm{T} = \mathbb{R}^{+}$ or $\mathrm{T} = \mathbb{N}$). We give the following general definitions for functional variable /data and functional dataset.

Definition 14. A random variable $\mathcal{X}$ is called ***functional variable*** if it takes values in an infinite dimensional space. An observation (value of a functional variable) χ of $\mathcal{X}$ is called a functional data. ■

Definition 15. A ***functional dataset*** $\{\chi_1, \chi_2, \dots, \chi_n\}$ is the observation of n functional variables $\{\mathcal{X}_1, \mathcal{X}_2, \dots, \mathcal{X}_n\}$. ■

As we can see from the above definitions the notion of functional variable covers a larger area than curves analysis. In particular, a functional variable can be a random surface or any other more complicated mathematical object. Examples of functional data are the stochastic processes with continuous sample paths on a finite interval associated with the supremum norm or stochastic processes whose sample paths are square integrable on the real line. Spaces, such as the previous, are separable Banach spaces. For more details about functional data see Ferraty, Vieu (2006), and Ramsay, Silverman (2005,2002).

Unlike to the finite dimensional spaces (i.e., $\mathbb{R}^{N}$), there is no analog of the Lebesgue measure on an infinite dimensional Banach space. The density function of a random element, if it exists, is related to the dominating measure with respect to which the density function of the Radon–Nikodym derivative is computed.
For this reason, it is necessary to introduce new conditions in the topological structure, and to the delta sequences to obtain uniform results of a delta sequence estimator in a Banach space (or in a general infinite dimensional space).

Let $(\Omega, \mathcal{F}, P)$ be a probability space, and E is an infinite dimensional separable Banach space (e.g., the continuous functions on the interval $[0,1]$ endowed with the supremum norm). In additional, let $\mathfrak{B}$ be the σ-algebra of Borel subsets of E. Suppose now that X is a random element defined on $(\Omega, \mathcal{F}, P)$ taking values in $(E, \mathfrak{B})$, and that it has a density function f with respect to a σ-finite measure μ on $(E, \mathfrak{B})$, such that $0 < \mu(B) < \infty$ for every open ball $B \subseteq E$. Now, let $\{X_1, X_2, \dots, X_n\}$ be independent and identically distributed random elements (i.i.d) as X. Also, suppose C is a compact subset of E with the property that for any $r_n > 0$ there exists $t_k \in E$, $1 \le k \le d_n$

$$C \subseteq \bigcup_{k=1}^{d_n} \mathbf{B}\left(t_k, r_n\right), \tag{20a}$$

and

$$\text{there exists } a_n > 0 \text{ and } c > 0 \text{ such that } d_n \cdot \left(r_n\right)^{a_n} = c. \tag{20b}$$

The symbol $\mathbf{B}\left(t_k, r_n\right)$ denotes the open ball in E with center t_k and radius r_n. This is the new topological condition that we introduce in an infinite dimensional space. We assume that a compact set C can be covered by a finite number of balls (Eq. (20a)) and a geometrical link between the number d_n of balls and the radius r_n of each of them (Eq. (20b)).

Note that the condition (20a,b) is trivially satisfied in an Euclidian space ($\mathbb{R}^N$), by choosing $a_n = N$ (also, it holds true for infinite dimensional projection-based semi-metric spaces; see Ferrati, Vieu (2008) Sec.3 for the proof) but to obtain analogous uniform results in a general infinite dimensional space it is necessary to assume that a compact set can be written as we saw above. Ferrati, Vieu (2008), Sec. 1, observe this detail, and give an extensive discussion. Also, Prakasa-Rao (2010), Sec. 1, includes this in a Bannach space.

Now, we are in position to give the delta-sequence definition in Banach space and two additional conditions which we need in order to prove some theorems.

Let $\| \cdot \|$ denotes the norm of Banach space E.

(G1) We assume that for every $\varepsilon > 0$, there exists $\gamma > 0$ such that, if $x \in C$, $y \in E$: $\|x-y\| \le \gamma$ then $\left| f(x) - f(y) \right| \le \varepsilon$.

$$\text{Then, it follows that there exists } M > 0 \text{ such that } \sup_{x \in C} f(x) \le M < \infty. \tag{21}$$

Definition 16 (Prakasa-Rao (2010) Sec.1)**:** Let E be a Banach space. A sequence of nonnegative functions $\left\{ \delta_m(x,y),\, m \in \mathbb{N} \right\}$ defined on $E \times E$ is said to be a ***delta sequence*** with respect to the measure μ if it satisfies the following conditions

(G2) For every γ, $0 < \gamma \le \infty$, $\displaystyle \lim_{m \to \infty} \sup_{x \in C} \left| \int_{\mathbf{B}[x,\gamma]} \delta_m(x,y)\, \mu(dy) - 1 \right| = 0$,

where $\mathbf{B}[x,\gamma] = \left\{ y \in E : \|x-y\| \le \gamma \right\}$,

(G3) there exists a constant c_0 such that $\displaystyle \sup_{x \in C,\, y \in E} \delta_m(x,y) \le c_0 \cdot s_m < \infty$

where $0 < s_m \to \infty$ as $m \to \infty$ and $\lim\limits_{m\to\infty} \dfrac{m}{s_m \cdot \log(m)} = \infty$,

(G4) there exist $c > 0$, $\beta_1 > 0, \beta_2 > 0$ such that

$$\left| \delta_m(x_1, y) - \delta_m(x_2, y) \right| \le c \cdot (s_m)^{\beta_2} \cdot \| x_1 - x_2 \|^{\beta_1}, \; x_1, x_2, y \in E,$$

(G5) for any $\gamma > 0$, $\lim\limits_{m\to\infty} \sup\limits_{(x,y)\in C\times(E-\mathbf{B}[x,\gamma])} \delta_m(x,y)\cdot\| x-y \| = 0.$ ■

(G6) Finally, we suppose that $d_n = n^a$, for $a > 0$ and

$$(r_n)^{\beta_1} \cdot (s_m)^{\beta_2} < \left(\frac{s_m \cdot \log(m)}{m} \right)^{1/2}.$$

Now, a ***delta sequence estimator*** on a Banach space can be written in the following form

$$\hat{f}_n(x) = \frac{1}{n} \sum_{i=1}^{n} \delta_m(x, X_i). \tag{22}$$

Note that m might depend on n such that $m \to \infty$ as $n \to \infty$.

Theorem 5 (Prakasa-Rao (2010), Sec. 2. Theorem 1)**:** Suppose $m \to \infty$ and there exists $0 < p < 1$ such that $n^p < m < n$ for large n.Under the conditions (G1) – (G6) the following limiting relation holds true:

$$\lim_{n\to\infty} \sup_{x\in C} \left| \hat{f}_n(x) - f(x) \right| = 0 \quad a.s. \quad ■ \tag{23}$$

***Proof*:** Due to the complexity of this proof, we give the reference where an interested reader could study. The completed proof there is in Prakasa-Rao (2010), Sec. 2. Theorem 1. ◄

Theorem 6 (Prakasa-Rao (2010), Sec. 2. Theorem 1)**:** Suppose the conditions (G1) ,and (G3)–(G6). In addition, the following condition holds true:

(G2)′ there exists $0 \le \gamma \le \infty$, such that $\sup\limits_{x\in C} \left| \int\limits_{\mathbf{B}[x,\gamma]} \delta_m(x,y)\, \mu(dy) - 1 \right| = O(D_m)$,

where, $D_m = \sup\left\{ \| x-y \|, x \in C, \mathrm{y} \in E, \delta_m(x,y) > 0 \right\} = o(1)$, as $m \to \infty$. Also, suppose that f is Lipschitzian, i.e., there exists $K > 0$ such that

$$\left| f(x) - f(y) \right| \le K \| x - y \| \text{ , for any } x \in C \text{ and } y \in E \text{ .}$$

Then, with probability one, $\sup_{x \in C} \left| \hat{f}_n(x) - f(x) \right| = O\left(D_m\right) + O\left(\sqrt{\frac{s_m \cdot \log(m)}{m}}\right)$, (24)

■

Proof: Since (G2)′ implies (G2) applying Theorem 5, we get

$$\sup_{x \in C} \left| \hat{f}_n(x) - \mathrm{E}\left(\hat{f}_n(x)\right) \right| = O\left(\sqrt{\frac{s_m \cdot \log(m)}{m}}\right), \text{ a.s}$$

So it is sufficient to prove

$$\sup_{x \in C} \left| \mathrm{E}\left(\hat{f}_n(x)\right) - f(x) \right| = O\left(D_m\right). \tag{a}$$

We have

$$\begin{aligned} \mathrm{E}\left(\hat{f}_n(x)\right) - f(x) &= \int_E \delta_m(x,y) f(y) \mu(dy) - f(x) \\ &= \int_E \delta_m(x,y)\left(f(y) - f(x)\right)\mu(dy) \\ &\quad + \int_E \delta_m(x,y) f(x) \mu(dy) - f(x) \\ &= J + f(x)\left[\int_E \delta_m(x,y)\mu(dy) - 1\right], \end{aligned} \tag{b}$$

where $J = \int_E \delta_m(x,y)\left(f(y) - f(x)\right)\mu(dy)$.

Hence,

$$\left| \mathrm{E}\left(\hat{f}_n(x)\right) - f(x) - J \right| = O\left(D_m\right), \text{ by condition (G2)}'. \tag{c}$$

Thus, f is Lipschitzian, so there exists $K \in \mathbb{R}$, such that

$$\left| f(x) - f(y) \right| \le K \| x - y \|, \text{ for any } x \in C \text{ and } y \in E. \tag{d}$$

Then,

$$|J| \leq \int_E \delta_m(x,y) \left| f(y) - f(x) \right| \mu(dy)$$

$$\leq K D_m \int_E \delta_m(x,y)\, \mu(dy)$$

$$\leq O\left(D_m \right), \text{ by condition (G2)}'. \tag{e}$$

Now, combining (c) and (e) we derive

$$\left| \mathrm{E}\left(\hat{f}_n(x) \right) - f(x) \right| = O\left(D_m \right). \qquad \blacktriangleleft$$

Dabo-Niang (2004) gives a definition of naive kernel estimator in case that μ is a $\sigma-$ finite and diffuse measure.

Definition 17: Let μ be a σ-finite and diffuse measure, and a sequence r_n which satisfies $r_n > 0$, $\lim_{n\to\infty} r_n = 0$, $\lim_{n\to\infty} n \cdot \mu\left(\mathbf{B}\left[x, r_n \right] \right) = +\infty$ (21a,b,c). Let $\mathbf{B}\left[x, r_n \right]$ be the closed ball with center x, and radius r_n, and $\mathrm{I}(A)$ denotes the indicator function of a set A. Then, the ***naive kernel estimator*** given by

$$\hat{f}_n(x) = \frac{1}{n \cdot \mu\left(\mathbf{B}\left[x, r_n \right] \right)} \sum_{i=1}^{n} \mathrm{I}\left(X_i \in \mathbf{B}\left[x, r_n \right] \right), \quad \text{for } x \in E \qquad \blacksquare \tag{25}$$

Then she proves the following theorem

Theorem 7 (Dabo-Niang (2004), Sec. 1, Theorem 1)**:** If the density function f is continuous at $x \in E$ and $\hat{f}_n$ is the naive kernel estimator, then $\lim_{n\to\infty} \mathrm{E}\left(\hat{f}_n(x) - f(x) \right)^2 = 0$.

$\blacksquare$ (26)

Proof: First, we observe that for any $x \in E$ we have

$$\left| \mathrm{E}\left(f_n(x) \right) - f(x) \right| = \frac{1}{\mu\left(\mathrm{B}_{r_n}^x\right)} \left| \int_{\mathrm{B}_{r_n}^x} f(y)\, \mu(dy) - f(x) \right|$$

(a)

$$\leq \frac{1}{\mu\left(\mathrm{B}_{r_n}^x\right)} \int_{\mathrm{B}_{r_n}^x} \left| f(y) - f(x) \right| \mu(dy)$$

Also, by assumption f is continuous at $x \in E$, so $\forall\ \varepsilon > 0$ there exists $\delta_\varepsilon > 0$, such that $\| y - x \| < \delta_\varepsilon \Rightarrow \left| f(y) - f(x) \right| < \varepsilon$. (b)

Thus, if $r_n \to 0$, then, $\forall\ \varepsilon' > 0$ there exists $n_{\varepsilon'}$, such that $n \geq n_{\varepsilon'}:\ 0 < r_n < \varepsilon'$. Let $\varepsilon' = \delta_\varepsilon$, then, $\forall\ \varepsilon > 0$ there exists n_ε such that, $\forall\ n \geq n_\varepsilon$ and $y \in \mathrm{B}_{r_n}^x$ we have

$$\| y - x \| \leq r_n < \delta_\varepsilon \Rightarrow \left| f(y) - f(x) \right| < \varepsilon \qquad \text{(c)}$$

and finally we obtain

$$\frac{1}{\mu\left(\mathrm{B}_{r_n}^x\right)} \int_{\mathrm{B}_{r_n}^x} \left| f(y) - f(x) \right| \mu(dy) < \varepsilon. \qquad \text{(d)}$$

That is, $\lim_{n \to \infty} \mathrm{E}\left(\hat{f}_n(x) - f(x) \right) = 0$. (e)

And obtaining the variation

$$\mathrm{V}\left(f_n(x) \right) = \mathrm{E}\left(\hat{f}_n(x) - f(x) \right)^2$$

(f)

$$= \frac{\int_{\mathrm{B}_{r_n}^x} f(y)\, \mu(dy)}{n\left(\mu\left(\mathrm{B}_{r_n}^x\right) \right)^2} - \frac{\mathrm{E}^2\left(f_n(x) \right)}{n} \leq \frac{\int_{\mathrm{B}_{r_n}^x} f(y)\, \mu(dy)}{n\left(\mu\left(\mathrm{B}_{r_n}^x\right) \right)^2} = \frac{\mathrm{E}\left(f_n(x) \right)}{n\, \mu\left(\mathrm{B}_{r_n}^x\right)}$$

Now, if $r_n \to 0$ and $n\, \mu\left(\mathrm{B}_{r_n}^x\right) \to \infty$, then $V\left(f_n(x) \right) \to 0$. And the proof is complete. ◄

4.7. References

Breiman, L., Meisel, W., Purcell, E., 1977. Variable kernel estimates of multivariate densities. Technometrics, 19, pp. 135–144.

Dabo-Niang, S., 2004, Kernel density estimator in an infinite-dimensional space with a rate of convergence. Applied Mathematics Letters, 7, pp. 381-386.

Durrett, R., 2010, Probability: Theory and examples.

Ferraty, F., Vieu, P., 2006, Nonparametric functional data analysis: Theory and practice. Springer Series in Statistics, New York.

Ferraty, F., Vieu, P., 2008, Non-parametric models for functional data, with application in regression, time-series prediction and curve discrimination. Journal of Nonparametric Statistics, Vol. 20, No. 2, pp. 187-189.

Györfi, L., Kohler, M., Krzyzak, A., Walk, H., 2002, A distribution free theory of nonparametric regression. Springer Series in Statistics.

Hall, P., Wand, P.M., 1996, On the accuracy of binned kernel density estimators. Journal of Multivariate Analysis, 56, pp. 165-184.

Hoeffding, W., 1963, 58, Probability inequalities for sums of bounded random. Journal of American Statistical Association, pp. 13-30.

Loftsgaarden, D.O., Quesenberry, C.P., 1965, A nonparametric estimate of a multivariate density function. Ann. Math. Statist., 36, pp. 1049–1051.

Nolan, D., Marron, S., 1989, Uniform consistency of automatic and location-adaptive delta-sequence estimators. Probability Theory and related fields, 80, pp. 619-630.

Prakasa-Rao, B.S.L, 2010, Nonparametric density estimation for functional data by delta sequences. Brazilian Journal of Probability and Statistics, Vol. 24, No. 3, pp. 468–478.

Ramsay, O.J., Silverman, W.S., 2002, Applied functional data analysis: Methods and case studies. Springer Series in Statistics.

Ramsay, O.J., Silverman, W.S., 2005, Functional data analysis, edition 2, Springer Series in Statistics.

Rudin, W., 1974, Real and complex analysis, edition 2. New York: McGraw-Hill.

Sain, R,S., Scott, W.D., 1996, On locally adaptive density estimation.

Sain, R.S., 2002, Multivariate locally adaptive density estimation. Computational Statistics & Data Analysis, 39, pp. 165–186.

Shirahata, S., In-Sun Chu, 1992, Integrated squared error of kernel type estimator of distribution function. Ann. Inst. Statist. Math. Vol. 44, No. 3, pp. 579-591.

Vidal-Sanz, M,J., Delgado, A.M., 2004, Universal consistency of delta estimators. Ann. Inst. Statist. Math., Vol. 56, No. 4, pp. 791-818.

Vidal-Sanz, M,J., 2005, Pointwise universal consistency of nonparametric density estimators. Bernoulli, 11, pp. 971-985.

Walter, G., Blum, J., 1979, Probability density estimation using delta sequences, The Annals of Statistics, Vol. 7, No.2, pp. 328-340.

Wand, M.P., Jones, M.C., 1994, Multivariate plug-in bandwidth selection. Computational Statistics, pp. 98-116.

Wand, M.P., Jones, M.C., 1995. Kernel Smoothing. Chapman & Hall, London.