The transfer of distributions by LULU smoothers

Pieter-Willem Butler

Thesis presented in partial fulfilment of the requirements for the degree of
Master of Science

at Stellenbosch University

Promoter: Dr. C. H. Rohwer

December 2008

Declaration

By submitting this thesis electronically, I declare that the entirety of the work contained therein is my own, original work, that I am the owner of the copyright thereof (unless to the extent explicitly stated otherwise) and that I have not previously in its entirety or in part submitted it for obtaining any qualification.

Date: November 27, 2008

Abstract

Keywords: LULU smoother, median, discrete pulse transform, distribution transfer

LULU smoothers is a class of nonlinear smoothers and they are compositions of the maximum and minimum operators. By analogy to the discrete Fourier transform and the discrete wavelet transform, one can use LULU smoothers to create a nonlinear multiresolution analysis of a sequence with pulses. This tool is known as the Discrete Pulse Transform (DPT).

Some research have been done into the distributional properties of the LULU smoothers. There exist results on the distribution transfers of the basic LULU smoothers, which are the building blocks of the discrete pulse transform. The output distributions of further smoothers used in the DPT, in terms of input distributions, has been a challenging problem.

We motivate the use of these smoothers by first considering linear filters as well as the median smoother, which has been very popular in signal and image processing. We give an overview of the attractive properties of the LULU smoothers after which we tackle their output distributions.

The main result is the proof of a recursive formula for the output distribution of compositions of LULU smoothers in terms of a given input distribution.

Opsomming

Sleutelbegrippe: LULU gladstryker, mediaan, diskrete puls transform, verdelingsoordrag

LULU gladstrykers is 'n klas van nie-lineëre gladstrykers. Hulle is saamgestel uit die maksimum en minimum operatore. Analoog aan die diskrete Fourier transform en die diskrete golfie transform, kan 'n mens die LULU gladstrykers gebruik om 'n nie-lineëre multiresolusie analise van 'n sein met pulse te skep. Dit staan bekend as die Diskrete Puls Transform (DPT).

Daar is al navorsing gedoen op die verdelingseienskappe van die LULU gladstrykers. Resultate bestaan vir die verdelingsoordragte van die basiese LULU gladstrykers, wat die boustene van die diskrete puls transform is. Die uittreeverdelings van verdere LULU gladstrykers wat gebruik word in die DPT, in terme van intreeverdelings, is 'n uitdagende probleem.

Ons motiveer die gebruik van hierdie gladstrykers deur eers te kyk na lineëre filters asook die mediaan gladstryker wat baie gewild is in sein- en beeldverwerking. Ons gee 'n oorsig van die aantreklike eienskappe van die LULU gladstrykers waarna ons hul uittreeverdelings aanpak.

Die hoofresultaat is die bewys van 'n rekursiewe formula vir die uittreeverdeling van samestellings van LULU gladstrykers in terme van 'n gegewe intreeverdeling.

Acknowledgements

I would like to thank the following people:

- My promoter, Carl Rohwer, for your invaluable guidance and feedback and for introducing me to LULU theory.
- Hans Eggers, for your encouragement and support through the PMA programme.
- Marcel Wild, for your input and support during the last stretch.
- My family, for your continuous encouragement, financial support and prayer.
- My cell group members, for your prayers, especially Wikus van Zyl and Edmund-Douw Jacobs.
- Steve Kroon, for proofreading the text.
- Inge Lamprecht, for encouragement and companionship during my last year of study.

According to my earnest expectation and hope
that in nothing I shall be ashamed, but with
all boldness, as always, so now also
Christ will be magnified in my body,
whether by life or by death.
Phil 1:20

Contents

Chapter 1

Introduction

Since the advent of computers, the amount of data being processed has increased by orders of magnitude. This created the need to automate the processing of data. The presence of noise is inevitable in most systems and procedures to remove it have been devised. Gaussian noise, or noise from distributions that are "near" to a Gaussian, are generally removed with known linear techniques. Impulsive noise, which is noise consisting of interspersed impulses, cannot adequately be removed using linear procedures.

Modern data gathering and transmission is generally digital. The data can then be viewed as n-dimensional data arrays. Since these can in principle be considered as n 1-dimensional sequences, we shall concern ourselves with these. Such sequences are generally finite, in which case we can append zeros on both sides, and consider all the sequences to be absolutely summable. We shall assume the elements are ordered, so that we can call them time series and consider them to be indexed from $-\infty$ to ∞ by integers. Such sequences can be considered to be in a vector space, with the usual definition of addition and a scalar product.

The study of non-linear smoothers, which is used in the removal of impulsive noise, has been slow due to a lack of definitions. Boodram et al[2] tried to compute linear parts of a non-linear smoother. This has some merit but looking back it was not very fruitful. Velleman[22] considered such a non-linear theory difficult to envisage. In Galagher[7] the concepts of constant neighbourhood, edge, impulse and oscillation are defined.

The concept of monotonicity is a basic one when nonlinear smoothing is studied.

Definition. We define $\mathcal{M}_n$ to be the set of all n-monotone sequences. A sequence is n-monotone if for each j, $\{x_j, x_{j+1}, \ldots, x_{j+n+1}\}$ is monotone (non-increasing or non-decreasing).

These sets are nested, so that

$$\mathcal{M}_\infty \subset \cdots \subset \mathcal{M}_{n+1} \subset \mathcal{M}_n \subset \cdots \subset \mathcal{M}_0 = \mathbb{R}^{\mathbb{Z}}$$

where $\mathcal{M}_\infty$ is the set of all monotone sequences.

Definition. An n-blockpulse is a blockpulse of width n. It is upward if it is non-negative and downward if is is nonpositive.

Definition. An n-impulse is a sequence x such that $x_i = 0$ for $i \notin \{k + 1, \ldots, k+n\}$ for some k.

Figure 1.1 shows blockpulses and an impulse. By definition the set $\mathcal{M}_n$ cannot contain any blockpulses of width less than n. These properties of local monotonicity define a concept of smoothness.

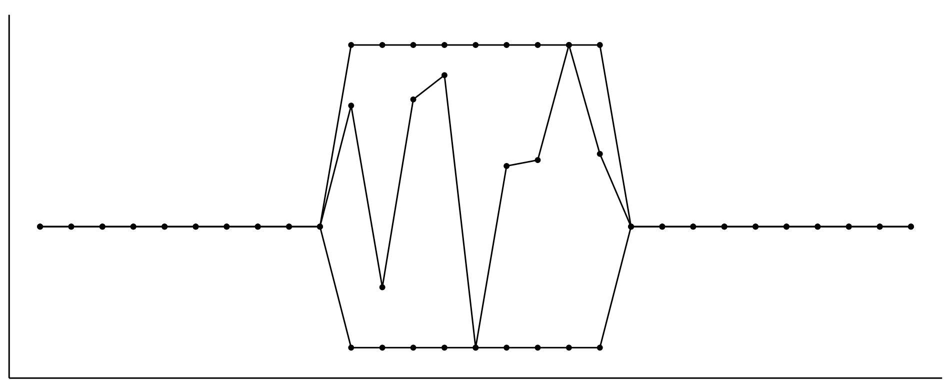


Figure 1.1: An impulse in between an upward and downward blockpulse

In order for smoothers to be axis and scale independent it is natural to choose the following basic axioms, following Mallows[11], but relaxing one axiom to allow operators like LULU operators while maintaining scale invariance.

Definition. A smoother P is an operator $\mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$ with the following properties

1. $PE = EP$.
2. $P(x + c) = P(x) + c$.
3. $P(cx) = cP(x)$, for all non-negative c.

where E is the shift operator with $(Ex)_i = x_{i+1}$. Axioms 1 and 2 mean the smoother is shift invariant relative to the horizontal and vertical axes, respectively. Axiom 3 means the smoother is scale invariant.

1.1 Norms

Suppose you have a time series and you want to fit a curve to the data so that it is optimal in some sense. This leads to approximation theory, which is well developed for the linear case arising from the choice of the ℓ_2-norm. The usual p-norms are defined by

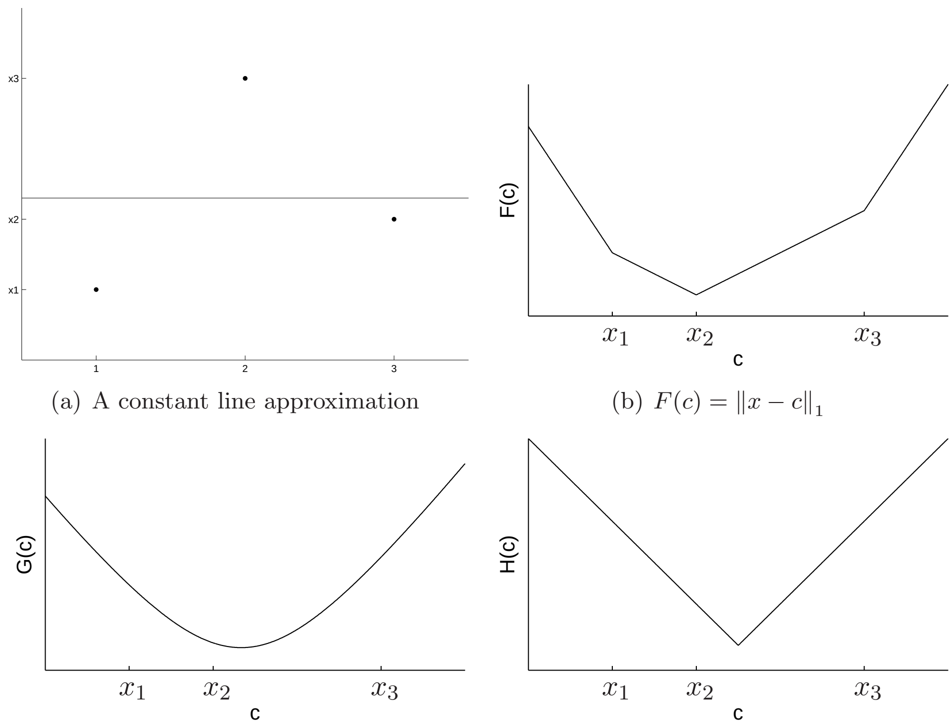
$$\|x\|_p = (|x_1|^p + |x_2|^p + \cdots)^{1/p}, \;\; x \in \mathbb{R}^n$$

The usual norms are $\|x\|_1$ (the absolute norm), $\|x\|_2$ (the Euclidean norm) and $\|x\|_\infty$ (the maximum norm). A vector space with a norm is called a normed vector space. The normed vector space associated with $\|x\|_p$ is denoted by ℓ_p.

The appropriateness of a choice of norm in the presence of noise is made clear by the following illustration. Suppose we have a sequence $x = (x_1, x_2, x_3)$ which are measurements of an underlying quantity that is essentially constant and we want to approximate x with a constant value c that is optimal in some norm. This situation is shown in figure 1.2(a). We will now find the optimal solution for these three values of p. We may assume that $x_1 \leq x_2 \leq x_3$ without loss off generality.

First let $p = 1$. We want to minimize with regards to $c \in \mathbb{R}$ the function

$$\begin{aligned} F(c) &= \|x - c\|_1 \\ &= |x_1 - c| + |x_2 - c| + |x_3 - c| \end{aligned}$$



(a) A constant line approximation

(b) $F(c) = \|x - c\|_1$

(c) $G(c) = \|x - c\|_2$

(d) $H(c) = \|x - c\|_\infty$

Figure 1.2: Minimizing in different norms

The graph of F is shown in Figure 1.2(b). The minimum of $F(c)$ is at $c = x_2 = \text{median}\{x_1, x_2, x_3\}$, where the median is the middle ordered value.

Now let $p = 2$. We want to minimize

$$\begin{aligned} G(c) &= \|x - c\|_2 \\ &= \sqrt{(x_1 - c)^2 + (x_2 - c)^2 + (x_3 - c)^2} \end{aligned}$$

Its graph is shown in Figure 1.2(c). We minimize G^2 by the usual method of linear regression and find that G is minimized by $c = \frac{1}{3}(x_1 + x_2 + x_3) = \text{average}\{x_1, x_2, x_3\}$.

Now let $p = \infty$. We want to minimize

$$\begin{aligned} H(c) &= \|x - c\|_\infty \\ &= \max(|x_1 - c|, |x_2 - c|, |x_3 - c|) \end{aligned}$$

The shape is shown in Figure 1.2(d). The minimum of $H(c)$ is at

$$c = \frac{1}{2}\big(\max\{x_1, x_2, x_3\} + \min\{x_1, x_2, x_3\}\big).$$

The 2-norm gives rise to an inner product space which has the usual linear theory available. If one of the values contains a large error, the minimum value of G, which is the average of the values of x, will be greatly affected.

The ∞-norm might be most useful when working with periodic errors. Such a system is shown in Figure 1.3 in which the quantity that is measured is essentially constant. If the error is due to such a vibration, the chances of measuring the system near its minimum or maximum is greater than near the middle, so taking measurements and calculating $\frac{1}{2}\big(\max(x) + \min(x)\big)$ would give a good estimate of the desired quantity. When a large error is measured, the maximum or minimum value of H will be affected and therefore also the estimate of the desired quantity.

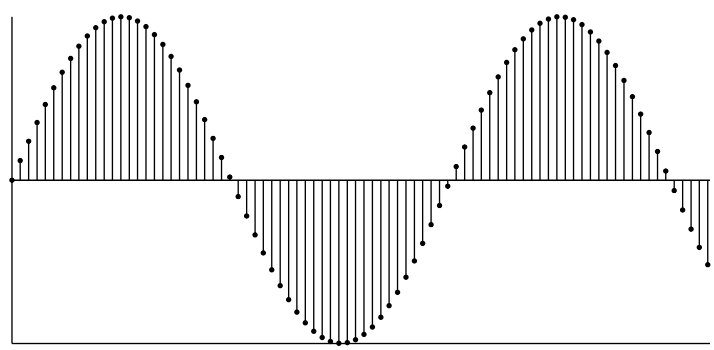


Figure 1.3: Measuring a periodic system

Because the minimum in the 1-norm is given by the median value, it is the most natural norm when one of the three values contains a large error, since the size of this does not affect the median very much, provided the other errors are small. We say the estimate is robust with respect to errors.

This illustration, apart from demonstrating the appropriateness of choice of norm, also provides a clue as to why the median and related smoothers work well when dealing with impulsive noise.

1.2 Linear filters

The theory of linear digital filtering is well established. A good introduction to the field is Hamming[8].

For all sequences x, y and $c \in \mathbb{R}$, an operator A is linear if

- $A(x + y) = A(x) + A(y)$
- $A(c\,x) = c\,A(x)$

We will consider the running average filter, the discrete Fourier transform [8] and the wavelet transform [3]. When looking at each one we will see problems that can arise with that class of filters when dealing with impulsive noise.

1.2.1 The running average filter

The running average filter R_n is defined by

$$(R_n\,x)_i = \frac{1}{2n+1} \sum_{j=i-n}^{i+n} x_j$$

where we use a window of size $2n + 1$. In Figure 1.4 we took a signal which represents measurements of an essentially constant quantity in the presence of impulsive noise (the spikes) and applied the 3-point running average filter, R_1, to it.

We note that the filtered signal looks smoother than the original. However the pulse is not simply removed but is spread out around the point it occurred. The average filter does not remove the pulses cleanly but actually contaminates the data around them.

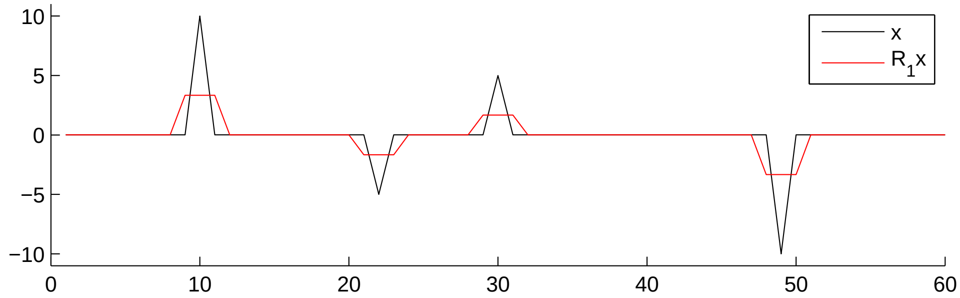


Figure 1.4: Filtering a signal with R_1

1.2.2 The discrete Fourier transform

The discrete Fourier transform involves the following transformation of the discrete data x.

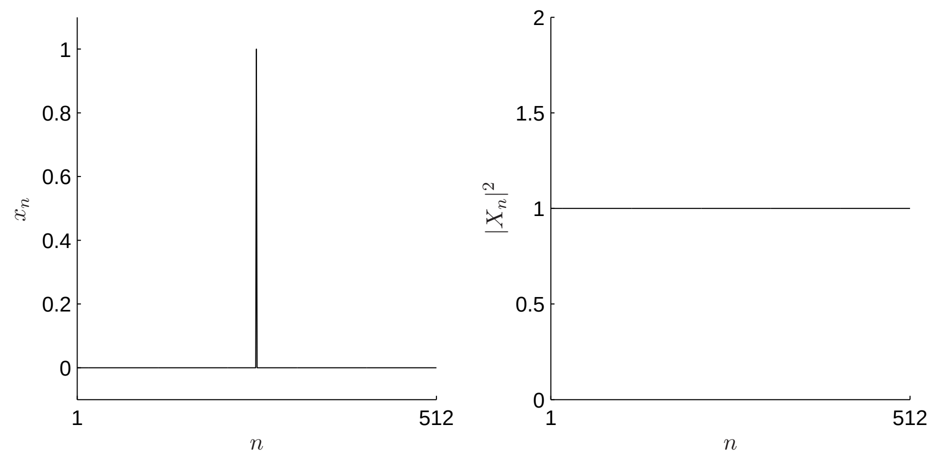
$$X_k = \sum_{n=0}^{N-1} x_n \exp\left(-\frac{2\pi i k n}{N}\right), \quad k = 0, 1, \ldots, N-1$$

where x and X both have length N. The complex numbers X_k represent the amplitude and phase of the various sinusoidal components of x. Parseval's identity states that

$$\sum_{n=0}^{N-1} |x_n|^2 = \frac{1}{N}\sum_{n=0}^{N-1} |X_n|^2$$

This quantity can be interpreted as a measure of the energy in the signal x. The sequence $|X_n|^2$ can be called the power spectrum of x. The power spectrum gives the energy in the frequencies present in the data. One could smooth a signal by identifying those frequencies that represent the noisy component and discarding them. Typically the higher frequencies are discarded to create a low-pass filter. Figure 1.5 shows a zero sequence with a pulse at $x_i = 1$ and its power spectrum.

Notice that the power spectrum is a constant function. This means that all N sinusoidal components are needed in the Fourier series representation of the pulse. Only one of the components of x is nonzero but there are no zero elements in X. Due to the linearity of the discrete Fourier transform, the appearance of a noisy pulse in a signal is therefore difficult, if not impossible



(a) A pulse (b) Its power spectrum

Figure 1.5: The power spectrum of a pulse

to eliminate with this technique because the energy in the pulse is spread out when smoothing the signal.

1.2.3 The Wavelet transform

A wavelet is a function $\Psi(t) \in L_2(\mathbb{R})$ such that the family of functions

$$\Psi_{j,k} = 2^{j/2}\Psi(2^j t - k)$$

where j and k are arbitrary integers, is an orthonormal basis in the Hilbert space $L_2(\mathbb{R})$. We now look at the most elementary wavelet as an example, which is the Haar wavelet. It is defined by

$$H(t) = \begin{cases} 1 & t \in \left[0, \frac{1}{2}\right) \\ -1 & t \in \left[\frac{1}{2}, 1\right] \\ 0 & \text{otherwise.} \end{cases}$$

Figure 1.6 shows a graph of the Haar wavelet.

Wavelets are used to create a multiresolution analysis of a signal. These resolution levels are orthogonal to each other. Smoothing could be done by discarding resolution levels that may mainly contain the noisy component. Figure 1.7 shows the decomposition of the pulse shown in 1.5(a).

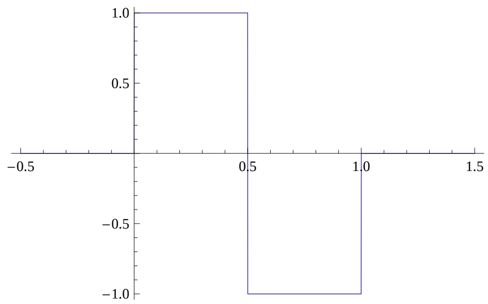


Figure 1.6: The Haar wavelet

Notice that on each level of the decomposition a wavelet can be seen. It means that the energy in the pulse is once again spread out on all levels. We used a signal of length $n = 512$, resulting in $\log_2(n) - 1 = 9$ resolution levels. It can be shown in general that there will appear a wavelet on every level. Due to the linearity of wavelet decomposition this will be a problem when a noise pulse is added to a signal.

1.2.4 Linear filters and impulsive noise

We have seen that with all the linear filters we considered, impulsive noise is a problem. We can understand this by seeing that all linear filters can be written as a convolution of a mask with a sequence [8]. A convolution is defined by

$$(m * x)_i = \sum_{k=-n}^{n} m_k x_{i-k}$$

where the mask m has length $2n+1$, in other words $m_k = 0$ when $|k| > n$. Define

$$\delta_{i,a} = \begin{cases} 1, & i = a \\ 0, & \text{otherwise} \end{cases}$$

Now the convolution of a mask m with a single pulse of length 1 and size α is

$$\begin{aligned} (m * \alpha\delta_{i,a})_i &= \sum_{k=-n}^{n} m_k \alpha \delta_{i-k,a} \\ &= \alpha m_{i-a} \end{aligned}$$

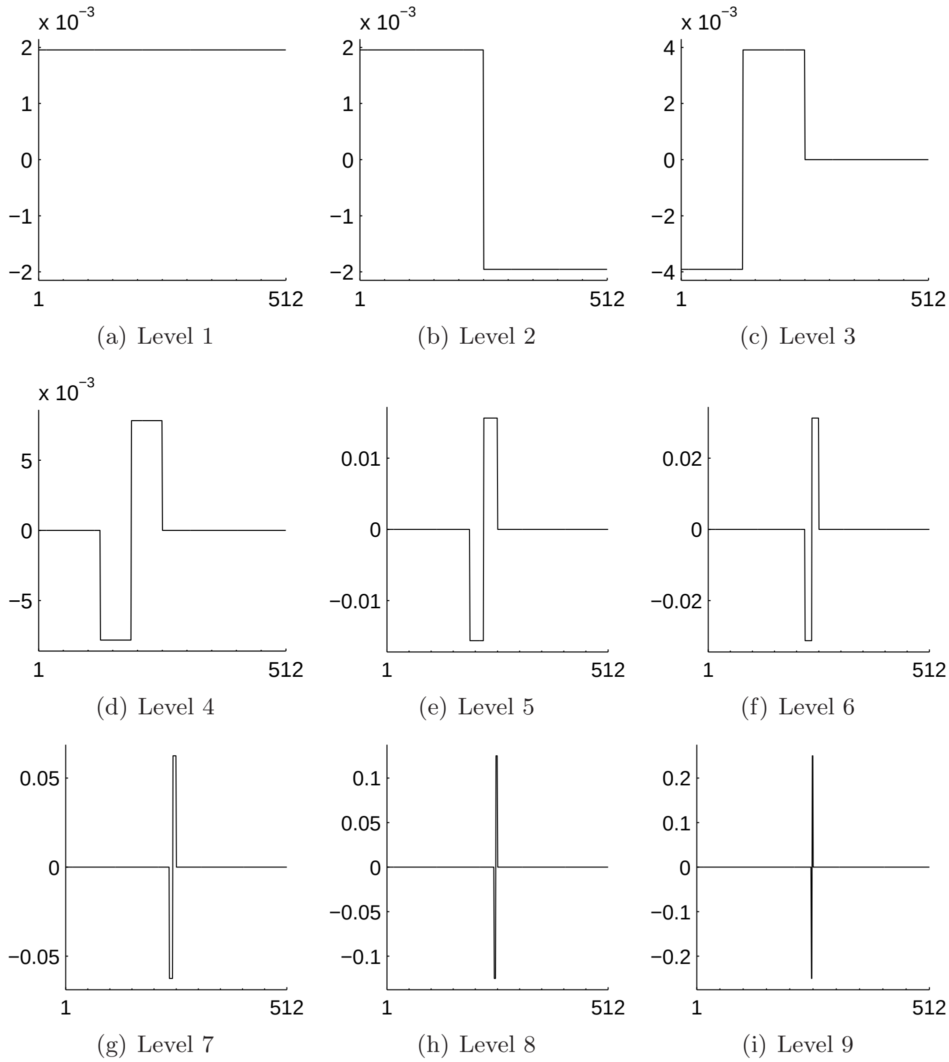


Figure 1.7: The Haar Wavelet decomposition of a pulse

This means that all points for which $|k| = |i - a| \leq n$ will be affected. Adding an impulse to a signal x would convolute in the following way

$$(m * (x + \alpha\delta_i, a))_i = (m * x)_i + \alpha m_{i-a}$$

so that the problem exists for any linear filter. Thus we have to look further than linear filters to find a practical solution to the problem of impulsive noise.

1.3 The median smoother

A solution to the problem of impulsive noise is to use a smoother that is not a convolution of a mask and a signal. In other words one must turn to some nonlinear smoother. The median smoother has been a well known tool for removing impulsive noise from data since the 1970's. It was popularised by John W. Tukey who also proposed a few variations of it.

We shall define the median smoother and variations of it and discuss some of their properties.

1.3.1 Definition

When one chooses a smoother it is desirable that it is symmetric with regard to the indices of the data series. In other words, that the output is independent of the direction in which the sequence is traversed. Therefore we define the median smoother for an uneven window size. Let x be a sequence of real numbers, that is $x \in \mathcal{M}_0$.

Define the median smoother M_n by

$$(M_n x)_k = \text{median}\{x_{k-n}, \cdots, x_{k+n}\}$$

where the median value is the middle ranked value of the subsequence of length $2n + 1$.

Figure 1.8 shows data that was filtered with the median smoother M_1. We use squares to indicate the input data, which is discrete. The dotted lines are added to make it more visible. The solid line shows the output.

The reason the 3-point median smoother is useful is because it removes most of the impulses. One would prefer that the output after smoothing is in $\mathcal{M}_2$, in other words no occurrence of 1-blockpulses. This would mean a "smoother" signal in that sense. However, as can be seen in the figure, there are still some impulses of width 1 left in the data.

To solve this problem one could just filter the output again, repeating un-

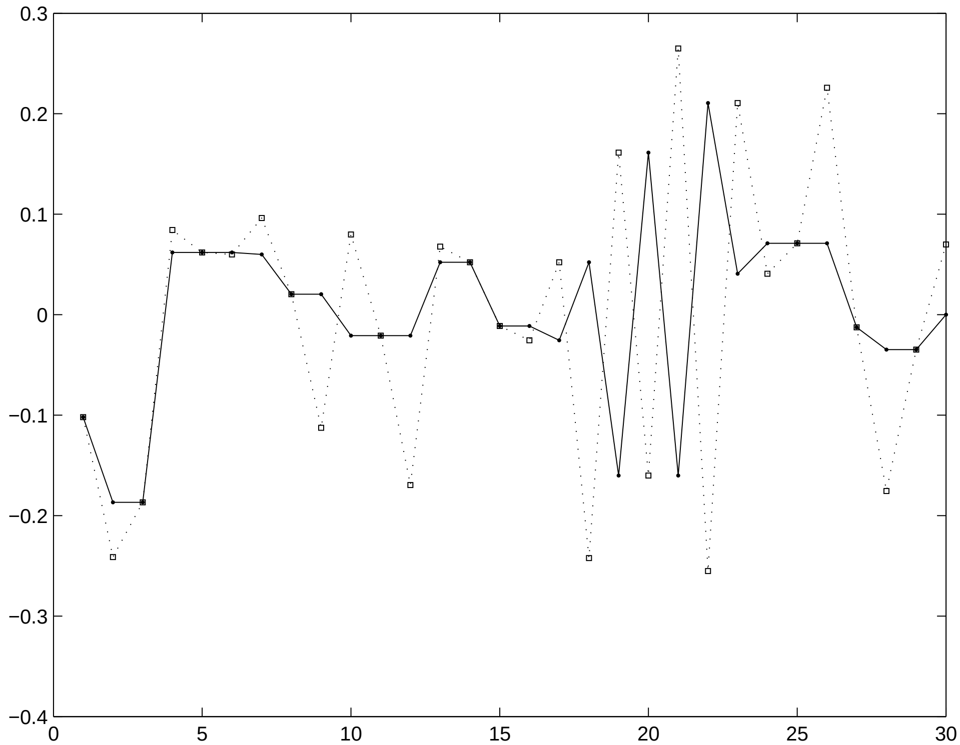


Figure 1.8: Output from the median smoother

til all the impulses are removed. Tukey proposed this and called it "3R" (meaning 3-point median, repeated). Some of its properties were discussed in Mallows[10]. We will denote it by M_1^∞, meaning that the M_1 smoother is applied infinitely many times, but in practice one would terminate the process when the output does not change anymore.

There is also a recursive (as opposed to repeated) version of the median smoother, M_n^*, which can be defined as follows. Let $y = M_n^* x$, then

$$y_k = \text{median}\{y_{k-n}, \cdots, y_{k-1}, x_k, \cdots, x_{k+n}\}.$$

The previously calculated median values are used to calculate the new value at the current index. The output of M_1^* was shown to stop changing after only one iteration. Its output, however, differs from that of the repeated median smoother and so their properties is also different.

1.3.2 Roots

Any signal that is left unchanged by a smoother is called a root of that smoother. Those signals that, through successive application of a smoother, converge to a given root, are called ancestors of that root. The structure of the roots are important to study because they are similar to the pass bands of a digital (linear) filter.

One would like the signal component of the data to be a root so that after smoothing one would be left with only signal and no noise. Noise can be filtered as long as the noisy data (signal + noise) is an ancestor of the same root as the signal.

In Gallagher et al[7] it was proved that for a sequence with length L, at most $\frac{1}{2}(L-2)$ iterations of the median smoother are necessary to reduce a signal to a root.

Defining the set of roots of M_n as R_n, they showed that these sets are nested such that

$$\cdots R_{n+1} \subset R_n \subset R_0 = \mathcal{M}_0$$

In Zhou et al[23] a study was made of the properties of infinite-length roots of the median smoother, which are problems in the study of this smoother. They showed that the roots of M_n are precisely those in $\mathcal{M}_n$ and its infinite roots. The infinite roots may be associated with the enigmatic behaviour of the median smoother. We will not concern ourselves with these.

1.3.3 Properties of the median

The median is a classical robust estimator because it is relatively straightforward to implement and it works well. It is able to preserve signal edges while it filters out impulses.

The classical algorithms for calculating the median is computationally expensive. For a sequence with length N, the calculations can be done in $O(N \log N)$ time using a comparison sort algorithm. Recently, a new algorithm analogous to noncomparison sorting appeared that can perform median smoothing in $O(N)$ time [13].

The filter became so popular because it works well in removing impulsive noise from data, but why it works is unknown. There is very little support from statistics.

The median smoother has only the window size n as a parameter. Therefore it cannot be designed to accommodate special signal or noise characteristics.

Locally monotonic regression was considered in Restrepo et al[14]. They studied the approximation of sequences with locally monotone sequences. They prove that optimal solutions exist and give algorithms for their computation. They also show that the output of the median smoother is not necessarily an optimal solution in the $\|\cdot\|_1$ or $\|\cdot\|_2$ sense. In Mallows[10] it is shown that the repeated median is also not optimal.

The one-dimensional median smoother can be extended to two dimensions or more. This was studied in Nodes et al[12], but we only consider the 1-dimensional case.

1.4 Calculating output distributions

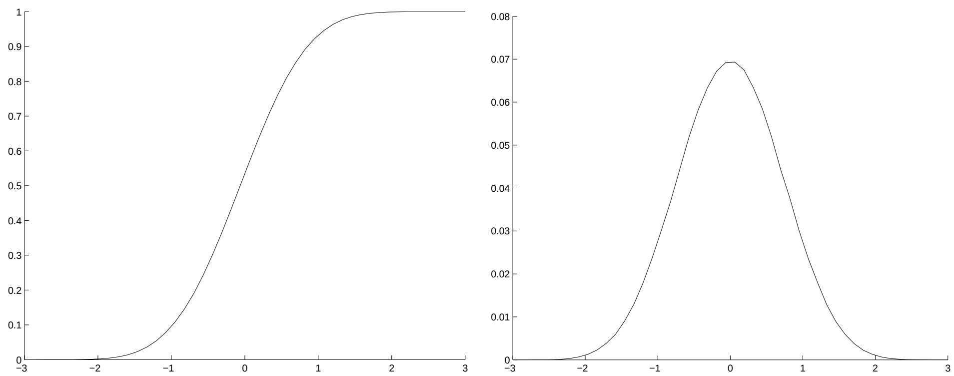
Given a random variable x the function $F(t)$ defined by

$$F(t) = P(x \leq t), \quad -\infty < t < \infty$$

is called the (cumulative) distribution function of x. It is of interest because all probabilities concerning x can be stated in terms of $F(t)$. It has the following properties.

- F is a nondecreasing function, i.e. $a < b \implies F(a) \leq F(b)$.
- $\lim_{b \to -\infty} F(b) = 0$.
- $\lim_{b \to \infty} F(b) = 1$.

Also of interest is $f(t)$, the derivative of the distribution function, called the (probability) density function. An example of a distribution function, the normal distribution, is shown in Figure 1.9(a). The corresponding density function is shown in Figure 1.9(b).



(a) A distribution function (b) Its density function

Figure 1.9: A distribution function and its derivative

We consider operators $P : \mathbb{R}^{\mathbb{Z}} \to \mathbb{R}^{\mathbb{Z}}$. An operator does not have a distribution, as it is not a random variable. However, if the input is a sequence of random variables, then the output is also a sequence of random variables. We shall consider random sequences as having random elements drawn identically, independently from a distribution F, and call them i.i.d. sequences.

The elements of the output sequence will again be identically distributed because the indexing into the sequence is arbitrary. They are not independent because the description of the operator provides a relationship between elements.

Suppose we know the operators A and B and we want to calculate $ABx = A(Bx)$ where we first calculate Bx and then apply A to the result.

Denote the distribution of x by F_x and the distribution of $A(x)$ by F_A. If we know F_A in terms of F_x as $F_A(F_x)$ and similarly $F_B = F_B(F_x)$ then one might naively think that we could find F_{AB} by substitution as $F_{AB} = F_A(F_B)$. This is clearly incorrect, as the sequence Bx is, in general, not i.i.d. any more.

In Section 2.2 we define the LULU smoothers as compositions of maximum and minimum operators. The situation sketched in the previous paragraph therefore applies to them. In this thesis we set out to find a general formula for the output distribution of C_n, a composition of successive LULU smoothers. C_n is useful as it is closely related to the discrete pulse transform (discussed later), which creates a multiresolution decomposition of a sequence using pulses.

With the assumption of a constant input sequence corrupted by random additive noise taken from an i.i.d. distribution, we will obtain the output distribution of some smoothers in terms of the input distribution. The usefulness of doing this is that the smoother's handling of noise can be studied.

1.5 Objectives

The purpose of this study is

- to investigate some statistical properties of the LULU smoothers
- to compare the LULU smoothers with the median smoothers in terms of these statistical properties
- to derive the output distributions of the LULU smoothers, in terms of i.i.d. input distributions, and more specifically a recursive formula for the output distribution of C_n

1.6 Thesis Outline

We have already discussed the weakness of linear filters when dealing with impulsive noise and given an overview of the very popular median smoother.

The median smoother turns out to be linked to LULU smoothers. In the next chapter we will look at the LULU smoothers and give an overview of some of the properties that make them attractive.

In Chapter 3 we will look at the output distributions of median and LULU smoothers and then compare them based on their distributions.

The output distributions of the LULU smoothers, in terms of i.i.d. input distributions, which are only stated in Chapter 3, are derived in chapter 4.

Chapter 2

Overview of the LULU smoothers

2.1 Introduction

In this chapter we will give an overview of the LULU smoother. First we will define the basic LULU smoothers and describe the discrete pulse transform. We then discuss some of the properties of LULU smoothers, namely idempotence, co-idempotence, consistency, variation preservation and shape preservation. We also look at related concepts, such as the LULU semigroup, LULU intervals. We will illustrate how LULU smoothers are linked to the median smoothers. The LULU smoothers have been studied for the last twenty years. For a more detailed overview and proofs of theorems, Rohwer[16] can be consulted.

2.2 Definitions

The basic building blocks for LULU smoothers are the maximum and minimum operators. Let x be a doubly infinite sequence that is absolutely summable, i.e.

$$\sum_{i=-\infty}^{\infty} |x_i| < \infty$$

The maximum and minimum operators $\bigvee_n$ and $\bigwedge_n$ for window size $n+1$ are defined by

$$y = \bigvee_n x, \text{ when } y_i = \max\{x_i, \dots, x_{i+n}\}$$

$$y = \bigwedge_n x, \text{ when } y_i = \min\{x_{i-n}, \dots, x_i\}$$

These operators are then combined to give the L_n and U_n operators defined by

$$L_n = \bigvee_n \bigwedge_n$$

$$U_n = \bigwedge_n \bigvee_n$$

In terms of x these can be written as

$$(L_n x)_i = \max\{\min\{x_{i-n}, \dots, x_i\}, \dots, \min\{x_i, \dots, x_{i+n}\}\}$$

$$(U_n x)_i = \min\{\max\{x_{i-n}, \dots, x_i\}, \dots, \max\{x_i, \dots, x_{i+n}\}\}$$

Notice that L_n and U_n have an effective window size of $2n+1$. We define a LULU operator as any finite composition of L_m and U_n. The LULU operators have been shown to be smoothers.

L_n and U_n can in turn be combined by composition to yield the L_nU_n and U_nL_n operators, so that

$$L_nU_n = \bigvee_n \bigwedge_n \bigwedge_n \bigvee_n$$

$$U_nL_n = \bigwedge_n \bigvee_n \bigvee_n \bigwedge_n$$

As we shall often apply these successively, we define the ceiling operator C_n recursively as

$$\begin{aligned} C_n &= L_nU_nC_{n-1} \\ &= (L_nU_n)(L_{n-1}U_{n-1})\cdots(L_1U_1), \end{aligned}$$

and the flooring operator F_n as

$$\begin{aligned} F_n &= U_nL_nF_{n-1} \\ &= (U_nL_n)(U_{n-1}L_{n-1})\cdots(U_1L_1). \end{aligned}$$

2.3 The discrete pulse transform

By analogy to the discrete Fourier transform (DFT) and the discrete wavelet transform (DWT) we can use LULU smoothers to create a multiresolution analysis of a signal with pulses. We will show how to construct such a decomposition using L_nU_n. A similar decomposition is possible using U_nL_n.

We define the residual operator

$$D_n = C_{n-1} - C_n = (I - L_nU_n)C_{n-1}$$

When applied to a sequence x, it yields the resolution sequence $r_n = D_nx$, so that we have the discrete pulse transform (DPT)

$$DPT(x) = D(x) = [D_1(x), D_2(x), \ldots, D_N(x)]$$

where N is the length of x. Figure 2.1 shows a schematic diagram for obtaining the different resolution levels.

Denote by $z = C_{n-1}x$ the output on level $n-1$. L_nU_n, which is the smoothing operator at the n-th level, can be seen as a separation in two stages. First z is separated into U_nz and $(I - U_n)z$, then U_nz is separated into L_nU_nz and $(I - L_n)U_nz$. So we have

$$\begin{aligned} r_n &= I - L_nU_nz \\ &= (I - U_n)z + (I - L_n)U_nz \end{aligned}$$

It is easy to show [16] that the residual of U_n, i.e. $I - U_n$, consists only of downward n-blockpulses. Similarly $I - L_n$ consists only of upward n-blockpulses. So we have

$$r_n = r_n^- + r_n^+$$

where r_n^- contains only downward blockpulses and r_n^+ only upward blockpulses.

Abbusing the notation somewhat, let b_{nj} be a sequence consisting of only one n-blockpulse starting at index j with amplitude b_{nj}. We can write

$$r_n = \sum_{j=\Omega_n} b_{nj}$$

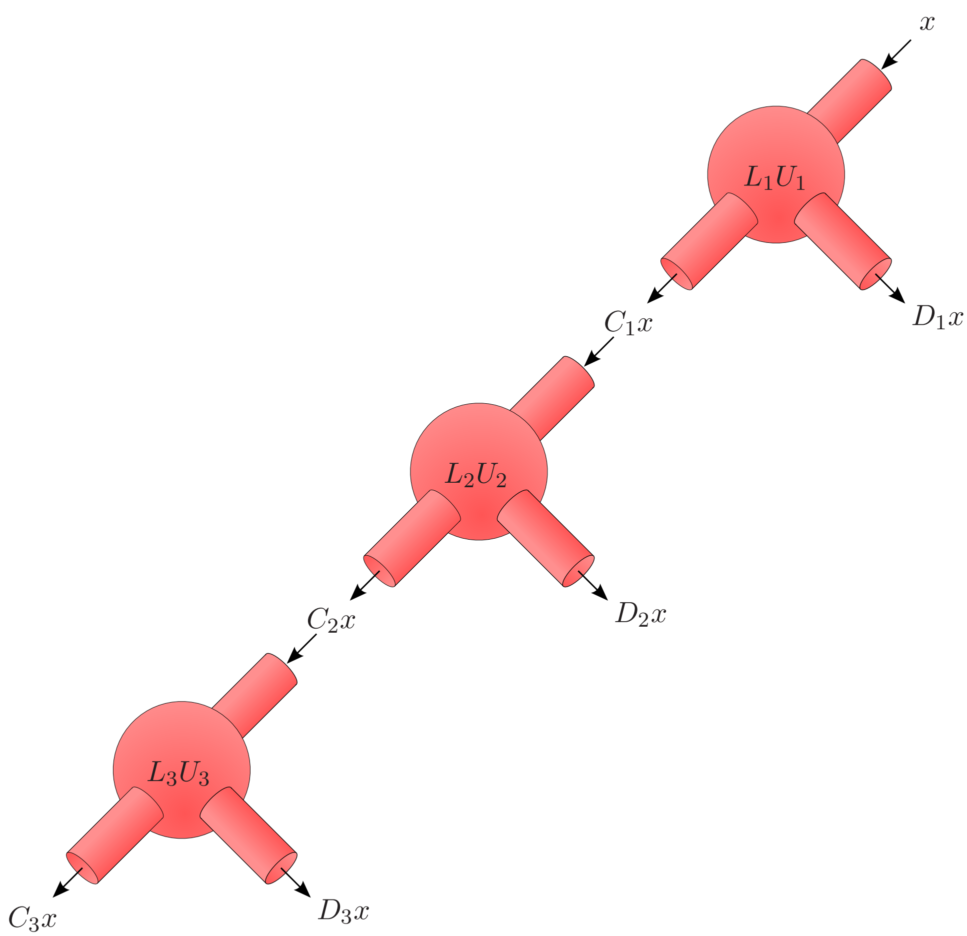


Figure 2.1: The discrete pulse transform

where Ω_n is an index set. Then we can represent x as

$$x = \sum_{n=1}^{N} r_n = \sum_{n=1}^{N} \sum_{j=\Omega_n} b_{nj}$$

The sequence x can now be written as a matrix of N rows (the resolution levels) with the n-th row containing the starting indices of all n-blockpulses.

Figure 2.2 shows the DPT with L_nU_n of a sequence representing the side profile of a fort with impulsive noise added. Note that only blockpulses of width n appear on the n-th resolution level. The important structures of the original sequence is clearly visible. The impulsive noise, consisting solely of 1-impulses in this case, is removed in the first resolution level.

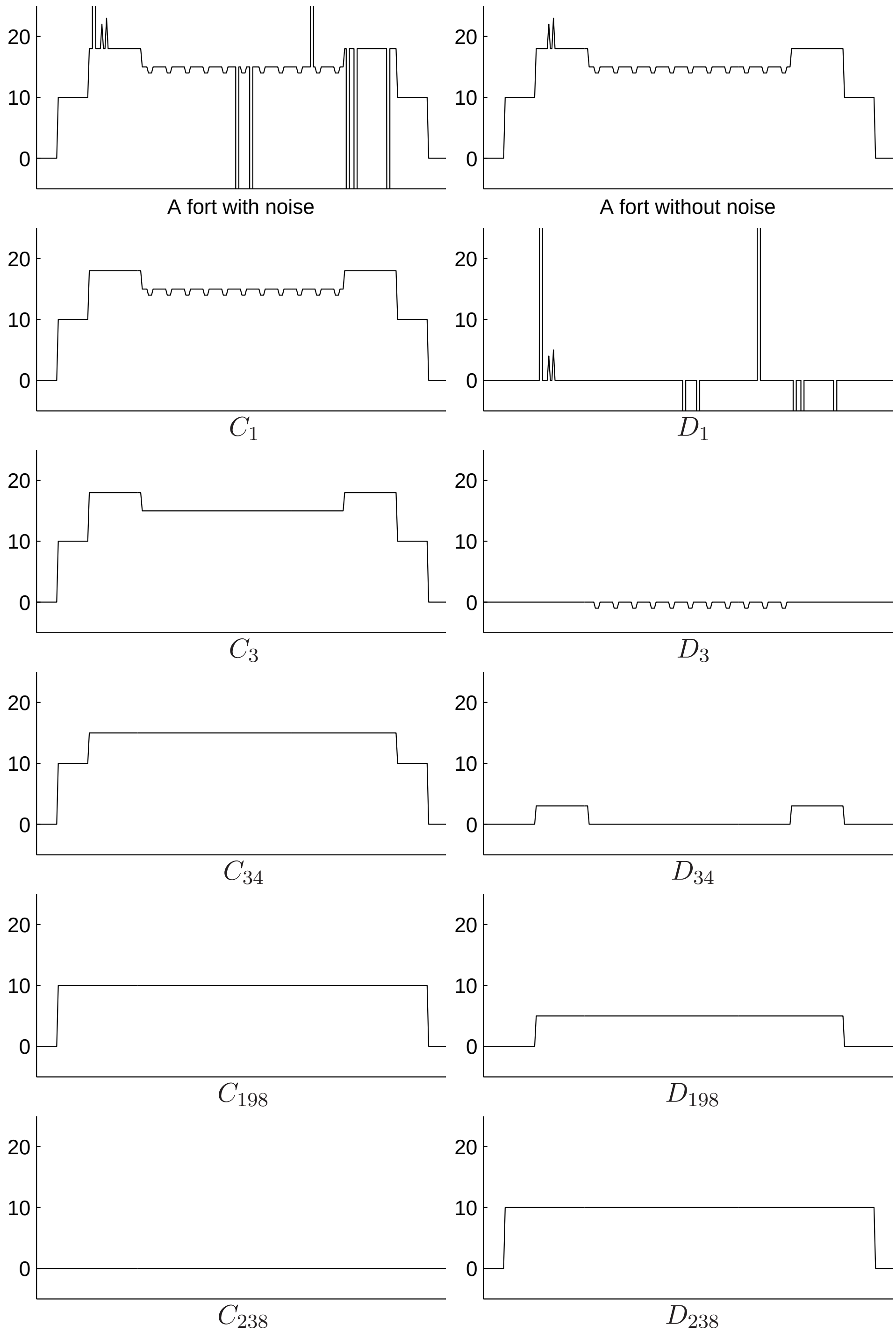


Figure 2.2: The DPT of a fort

2.4 The LULU smoothers are separators

Definition. A separator S is a smoother that also has the following properties

- Idempotence: $S^2 = S$
- Co-idempotence: $(I - S)^2 = I - S$

In Figure 2.3 is shown a two stage cascade diagram of a general operator P.

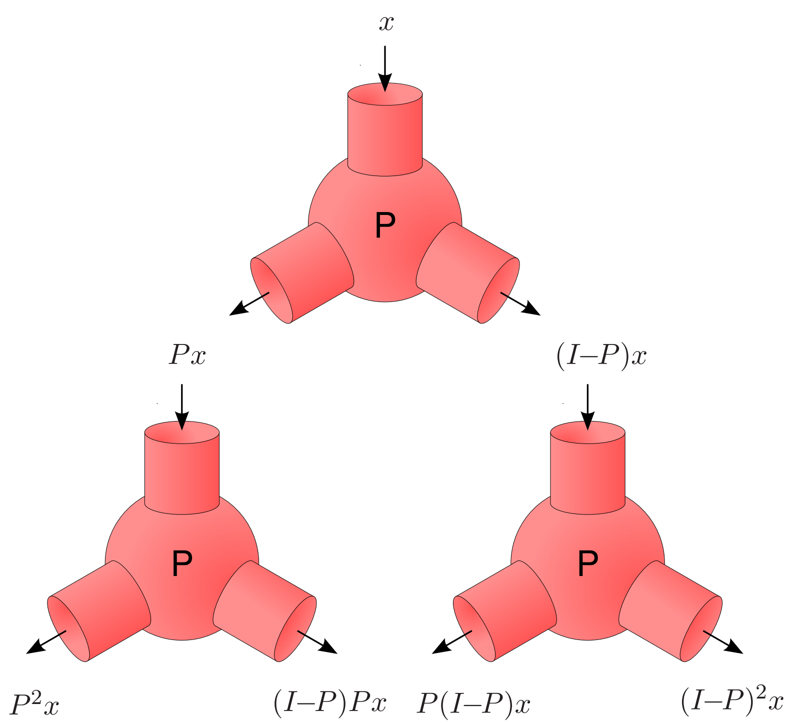


Figure 2.3: A general two-stage operator cascade

Idempotence means that the smoothed sequence Px is unchanged when smoothed again by the same operator P. Px is therefore a root of P. Co-idempotence means that the residual is mapped onto the zero sequence. Note that for a linear operator, idempotence and co-idempotence are equivalent.

A two stage cascade diagram is shown for a separator in Figure 2.4. Note that $S(I - S) = 0$ iff. $(I - S)^2 = I - S$ and $(I - S)S = 0$ is equivalent to the idempotence property. The LULU smoothers have been shown to be both idempotent and co-idempotent.

The LULU smoothers L_nU_n and U_nL_n have been shown to map a sequence onto $\mathcal{M}_n$. The residual is then in $\mathcal{M}_{n-1}$. We can use these facts to define signal and noise with respect to the LULU smoothers. We can consider sequences in $\mathcal{M}_n$ as signal and those in $\mathcal{M}_{n-1}$ as noise for some choice of n.

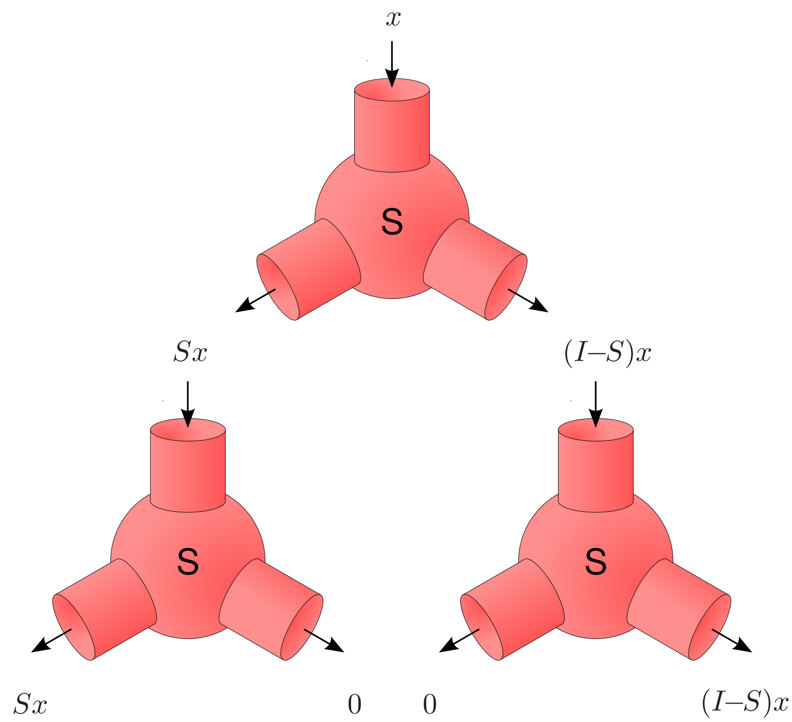


Figure 2.4: A two-stage separator cascade

2.5 The LULU semi-group

The idempotence properties can be generalized in that, for $n \leq m$ we have that

$$L_nL_m = L_m$$
$$U_nU_m = U_m$$

For $m = n$ this corresponds to idempotence for L_n and U_n. For each n the operators yield a semi-group generated by the pair L_n and U_n which, once it has been proven that $U_nL_nU_n = L_nU_n$ and $L_nU_nL_n = U_nL_n$, is easily seen to contain only four elements. Write $L = L_n$ and $U = U_n$. These four equalities give the following multiplication table.

	L	U	UL	LU
L	L	LU	UL	LU
U	UL	U	UL	LU
UL	UL	LU	UL	LU
LU	UL	LU	UL	LU

Table 2.1: The LULU semi-group

2.6 The LULU intervals

The elements of the LULU semigroup can be shown to satisfy

$$L_n \leq U_nL_n \leq F_n \leq C_n \leq L_nU_n \leq U_n$$

We also know how some LULU operators relate for different values of n.

$$L_{n+1} \leq L_n \leq I \leq U_n \leq U_{n+1}$$

Writing these results in terms of intervals, we have the the LULU intervals

$$[F_n, C_n] \subseteq [U_nL_n, L_nU_n] \subseteq [L_n, U_n] \subseteq [L_{n+1}, U_{n+1}].$$

2.7 A link between LULU smoothers and the median

One of the main results of the LULU theory is the following inequality linking the median operator to LULU smoothers:

$$U_nL_n \leq M_n \leq L_nU_n$$

Figure 2.5 shows L_nU_n, U_nL_n and M_n for a signal with components from a Gaussian distribution. Notice that M_n always lies in between U_nL_n and L_nU_n and any particular component of $(M_nx)_i$, always equals the corresponding component of either L_nU_nx or U_nL_nx.

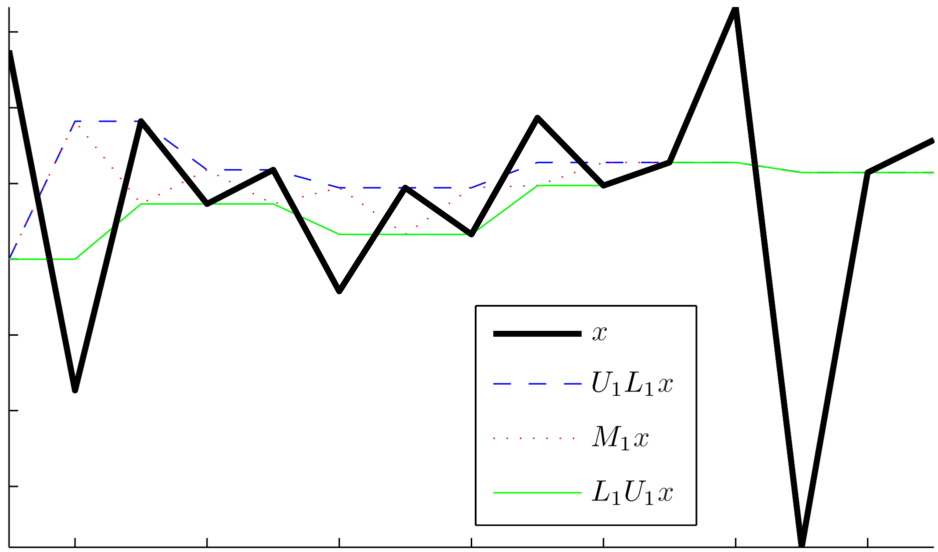


Figure 2.5: A link between U_nL_n, L_nU_n and M_n

2.8 Consistency

A first type of consistency is clear, namely that reconstruction of a sequence is achieved by simple addition of the resolution levels. Further consistencies are however required. Ultimately a very strong consistency result has been proved. This was originally stated as the Highlight Conjecture (see [18] and [19]).

Highlight Conjecture. *If*

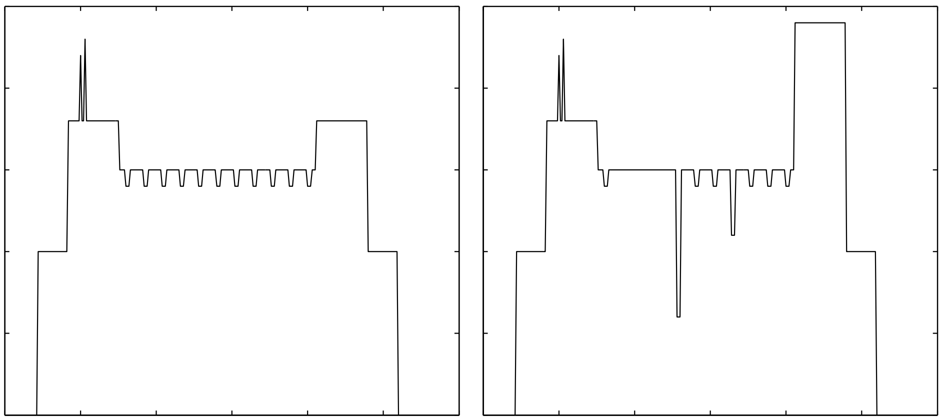
$$x = \sum_{n=1}^{N} \sum_{j=\Omega_n} b_{nj}(x) \quad \textit{and} \quad z = \sum_{n=1}^{N} \sum_{j=\Omega_n} \alpha_{nj} b_{nj}(x)$$

with $\alpha_{nj} \geq 0$ *arbitrary weights, then, when we decompose* z*, we have*

$$b_{nj}(z) = \alpha_{nj} b_{nj}(x).$$

This result is quite surprising, since the LULU smoothers are nonlinear operators. Historically, it was first shown to hold for the resolution levels r_n, then for the positive and negative resolution levels r_n^+ and r_n^-, and finally it was proven for individual blockpulses.

Figure 2.6 shows the decomposition of a sequence representing a fort using the DPT with $L_n U_n$. Some of the blockpulses were multiplied by a positive number. A few of the downward 3-blockpulses (refer to Figure 2.2) were discarded. The resulting sequence, when decomposed, yields precisely the modified resolution levels.



(a) (b)

Figure 2.6: Pulse highlighting using the DPT

The condition $\alpha_{nj} \geq 0$ means that in the cone generated by the individual pulses (i.e. an arbitrary nonnegative combination of blockpulses), a sequence is decomposed consistently. When working with images, this condition is not a limitation, because images cannot have negative luminosities.

2.9 Variation preservation

The total variation of a sequence is a measure of smoothness. The total variation of a sequence x is defined by:

$$T(x) = \sum_{i=-\infty}^{\infty} |x_{i+1} - x_i|$$

It has been proven that the LULU smoothers L_n and U_n preserve the total variation in a signal [15]. Let P denote a LULU smoother. It was shown that

$$T(x) = T(Px) + T(x - Px)$$

In other words, the variation of the smoothed data added to that of the residual equals the variation of the original data.

Since this is true for L_n and U_n , it is also true for L_nU_n and U_nL_n, and therefore it also holds for C_n and F_n. We demonstrate this property with a random sequence taken from a Gaussian distribution. The decomposition of the data using C_n is shown in Figure 2.7. The variation of each resolution level is calculated and plotted as a fraction of the total.

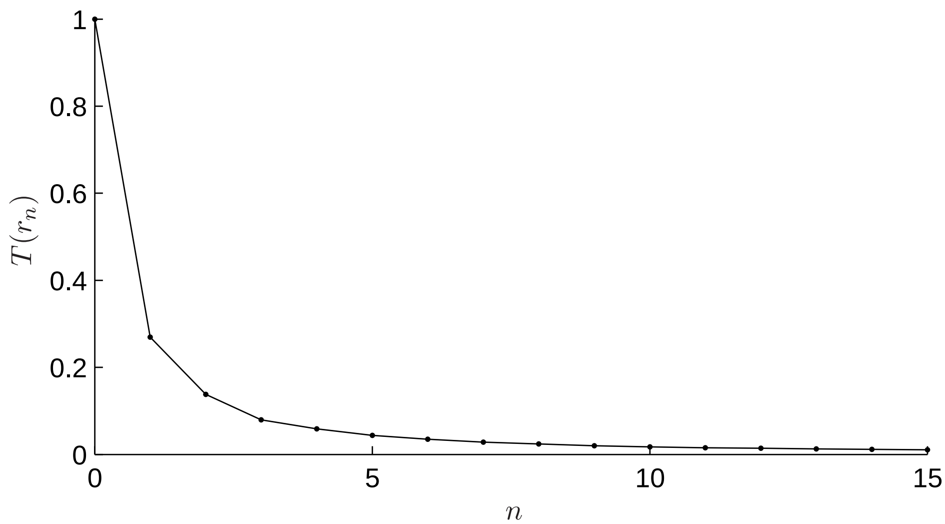


Figure 2.7: Total variation

Similar to the power spectrum associated with the DFT, which is useful for decision making, we can define the *variation spectrum* associated with the DPT.

Definition. *The variation spectrum of a sequence x is defined as*

$$t(x) = \{t_n = T(r_n(x));\ n \in \mathbb{Z}\}$$

The variation spectrum can be used to define a stopping criterion for the smoothing process. One criterion could be that whenever the ratio of the variation of the smoothed sequence to that of the original sequence falls beneath a certain threshold, the process is terminated. In an effort to separate perceived noise from signal, resolution levels in the DPT can thus be discarded based on their variation.

2.10 Shape preservation

The LULU smoothers possess some strong shape preservation properties in which the order of neighbours in a sequence is preserved by the smoothers.

Definition. An operator S is neighbour trend preserving (NTP) if for any sequence x, $x_i \leq x_{i+1} \implies (Sx)_i \leq (Sx)_{i+1}$ and $x_i \geq x_{i+1} \implies (Sx)_i \geq (Sx)_{i+1}$.

L_n and U_n are easily shown to be NTP. All compositions of these inherit the property. We can define an even stronger property.

Definition. A separator S is fully trend preserving (FTP) if both S and $I - S$ are NTP.

All LULU compositions turn out to be FTP. One consequence of this property is that, for a general sequence x, every vector in the cone generated by individual blockpulses in the DPT of x has the same order relations as x. This implies edge preservation, which is useful in image analysis.

2.11 Summary

The discrete pulse transform can be used to create a multiresolution analysis of a sequence with pulses. The DPT is consistent, and the Highlight Conjecture is a remarkable result since these are nonlinear operators.

LULU smoothers have other very attractive properties. They are idempotent as well as co-idempotent, making them separators. The simplicity of the theory is seen in the LULU semigroup, for which a natural order relation arises.

The median smoother is closely related to the LULU smoothers and we can now understand why the median smoother performs well when it does, based on our understanding of the performance of the LULU smoothers.

The LULU smoothers preserve total variation, which is a natural measure of smoothness. They are also fully trend preserving, with the consequence that edges are preserved, which is useful in image processing.

We can calculate L_nU_n, applied to a sequence of length N, in $O(N)$ time [5], which means that it outperforms the classical implementation of the median smoother. Laurie[18] introduced a new technique called the Roadmaker's Algorithm. Using this technique one can calculate C_n, and hence also the DPT, in $O(N)$ time.

Chapter 3

The output distributions

3.1 Introduction

Much information can be gained from studying the output distributions of smoothers. The moments are useful aids in the analysis of the smoothers for bias and other characteristics.

We will look at the output distributions of the median smoother and its variants, before we turn our attention to the LULU smoothers. In section 3.3 we state the distributions of $\bigvee_n$ and $\bigwedge_n$, L_n and U_n, L_nU_n and U_nL_n, and C_n. In Section 3.4 we look at the first and second moments of these distributions.

In Section 3.5 we consider an alternative composite LULU smoother and compare its performance to C_n. In Section 3.6 we discuss how the first resolution level in the discrete pulse transform can be used to estimate the moments of an unknown error distribution under certain conditions. Section 3.7 reviews the results known for asymptotic distributions of the LULU smoothers.

Note that the proofs for the distributions given here are derived in the next chapter.

Here we consider input sequences with elements coming from the following four distributions: the uniform, linear spline, normal and Poisson distributions. An overview of these distributions can be found in Appendix A. We show the output of smoothing these sequences up to window size $n = 6$.

3.2 The median smoothers

The output distributions of M_n

The distribution of the median smoother can be obtained by considering a generalization of the median operator, namely the ranked order operator $R_{n,k}$ defined by

$$(R_{n,k}x)_j = \text{ the } k\text{-th smallest value of } \{x_{j-n}, \cdots, x_{j+n}\}$$

In Nodes et al[12] it is shown that the probability that the k-th ranked value $y = R_{n,k}(x)$ is not larger than t is given by

$$F_y(t) = \sum_{j=k}^{2n+1} \binom{2n+1}{j} F^j (1-F)^{2n+1-j}$$

where $F = F_x(t)$. Its density function is given by

$$f_y(t) = (2n+1)\binom{2n}{k-1} F^{k-1}(1-F)^{2n+1-k} f_x(t)$$

where $f_x(t)$ is the derivative of $F_x(t)$. The median is then the $(n+1)$-th ranked value and its distribution is given by

$$\begin{aligned} F_{M_n}(t) &= F_{R_{n,n+1}}(t) \\ &= \sum_{j=n+1}^{2n+1} \binom{2n+1}{j} F^j (1-F)^{2n+1-j} \end{aligned}$$

and its density function by

$$f_{M_n}(t) = (2n+1)\binom{2n}{n} F^n (1-F)^n f_x(t)$$

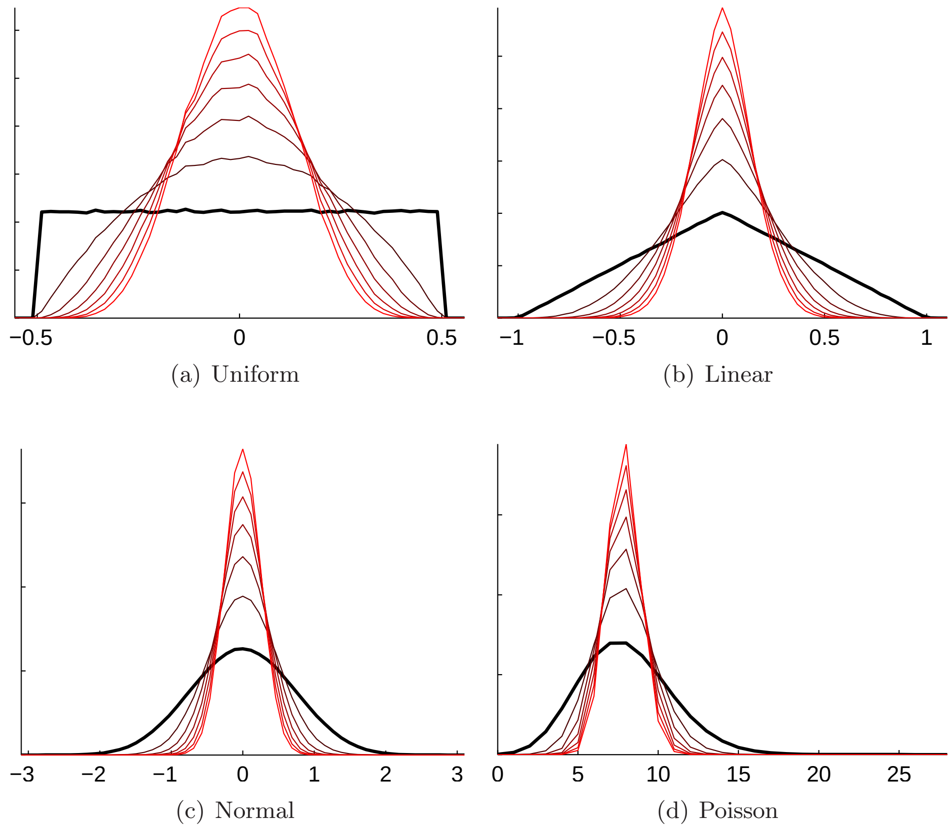


Figure 3.1: Density functions of M_n

The distributions of M_n are shown in Figure 3.1. In all cases, the wider the window, the more peaked the output.

If the input data is symmetrically distributed then the distribution of the median smoother's output is also symmetric [12]. In addition the statistical median as well as the statistical mean are preserved under median smoothing. The symmetry can be seen in the formula for $F_{M_n}(t)$, because swapping F and $1-F$ leaves the formula unchanged.

The output distributions of M_n^∞

Mallows[10] gives the distribution of the 3-point repeated median $y = M_1^\infty x$,

$$F_y(t) \;=\; \frac{2F^2 - F^3}{1 - F + F^2}$$

where $F = F_x(t)$. The corresponding density function is

$$f_y(t) \;=\; \frac{F(1-F)(4-F+F^2)}{(1-F+F^2)^2} f_x(t)$$

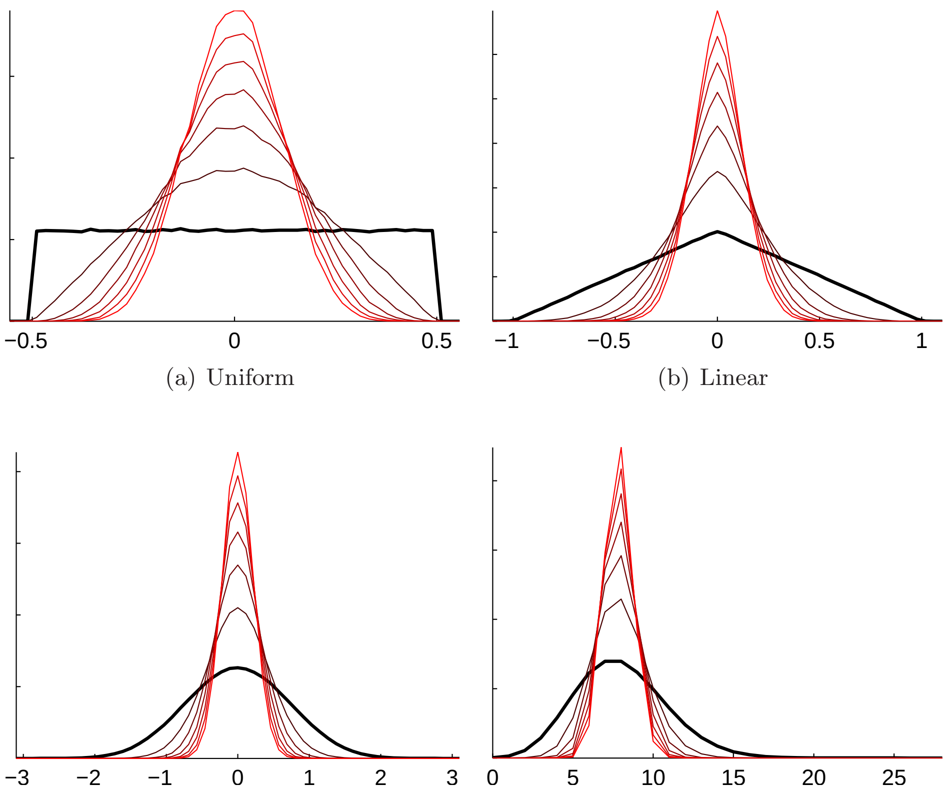


Figure 3.2: Density functions of M_n^∞

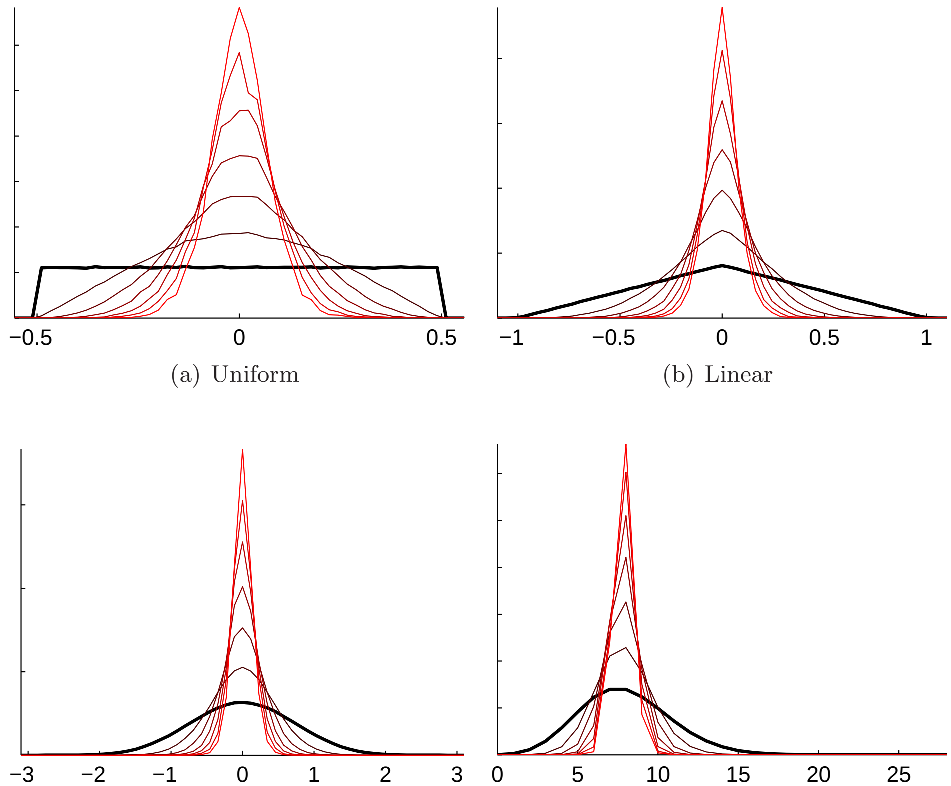
The output distributions of M_n^*

In Shmulevich et al[20] the distribution of the recursive median $y = M_n^* x$ is shown to be

$$F_y(t) \;=\; \left(1 + \left(\frac{G}{F}\right)^{n-1}\left(\frac{1-F^n}{1-G^n}\right)\right)^{-1}$$

where $F = F_x(t)$ and $G = 1 - F$. The corresponding density function is

$$f_y(t) \;=\; \frac{n(1-F^{n+1}-G^{n+1})-(1-F^n)(1-G^n)}{(FG^n+F^nG-F^nG^n)^2} F^n\, G^n\, f_x(t)$$



(a) Uniform (b) Linear

(c) Normal (d) Poisson

Figure 3.3: Density functions of M_n^*

Note that the repeated median is symmetrical, while the recursive median is not. This can be seen in the expression above, as well as in Figure 3.3.

3.3 The LULU smoothers

An operator Q is said to be a dual of an operator R if $Q(x) = -R(-x)$. It is easy to show that $\bigvee_n$ and $\bigwedge_n$, L_n and U_n, L_nU_n and U_nL_n, and C_n and F_n are duals of each other. Their output distributions are then also related (see section 4.5.1). When considering a sequence with elements from a symmetric distribution, the output distributions of duals applied to this sequence are mirrored images of each other around the mean of the input sequence. We will therefore only show graphs of the output density functions of $\bigvee_n$, U_n, L_nU_n and C_n. In what follows, we shall write F for F_x.

The output distributions of $\bigvee_n$ and $\bigwedge_n$

The output distributions of $\bigvee_n$ and $\bigwedge_n$ are given by

$$F_{\bigvee_n} = F^{n+1}$$
$$F_{\bigwedge_n} = 1 - (1 - F)^{n+1}$$

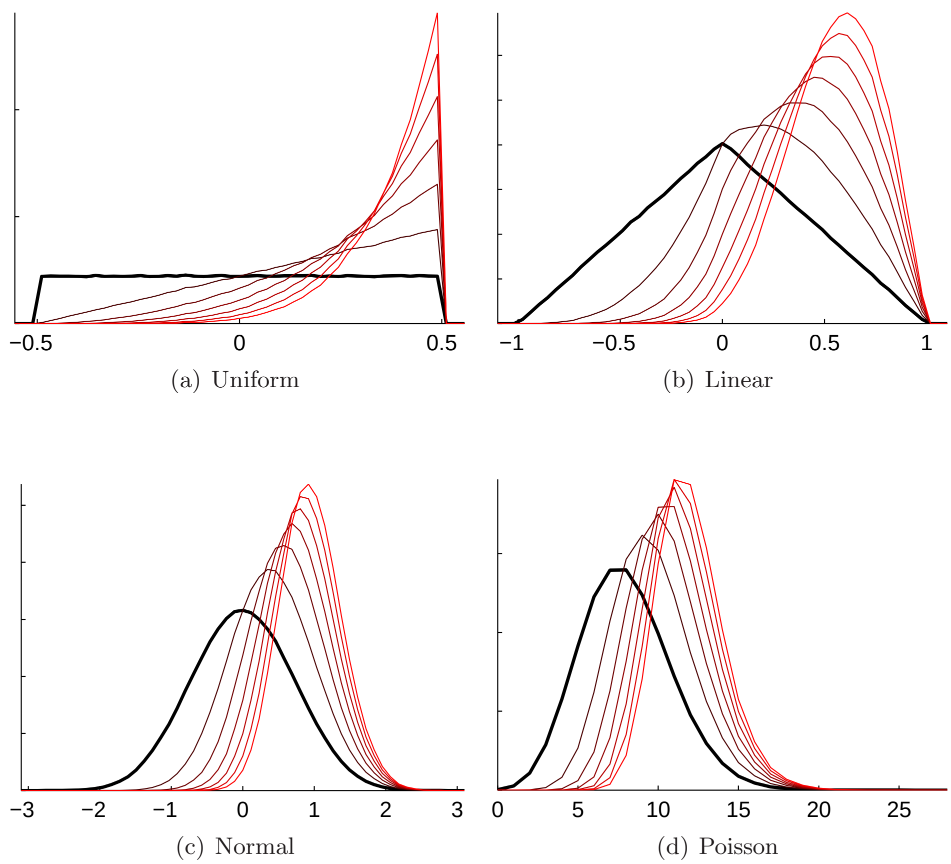


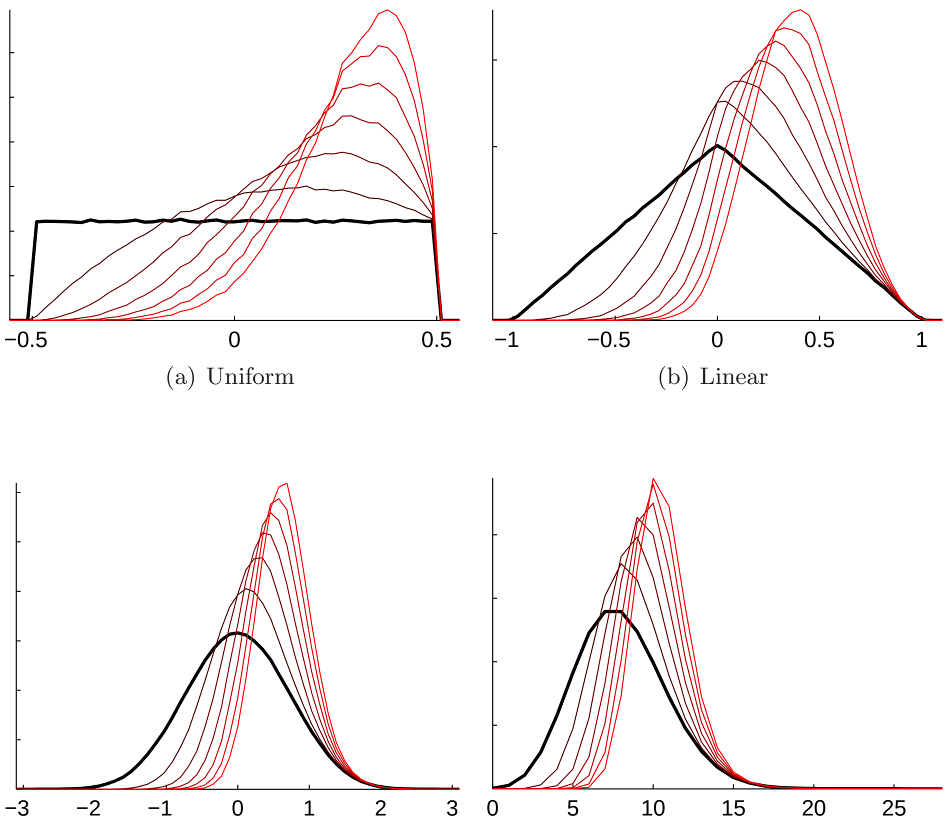
Figure 3.4: Density functions of $\bigvee_n$

The output distributions of U_n and L_n

The output distributions of U_n and L_n are given by

$$F_{U_n} = (n + 1)F^{n+1} - nF^{n+2}$$

$$F_{L_n} = 1 - (n+1)(1-F)^{n+1} + n(1-F)^{n+2}$$



(a) Uniform (b) Linear

(c) Normal (d) Poisson

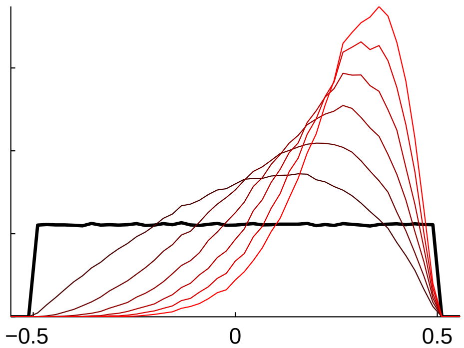
Figure 3.5: Density functions of U_n

The output distributions of L_nU_n and U_nL_n

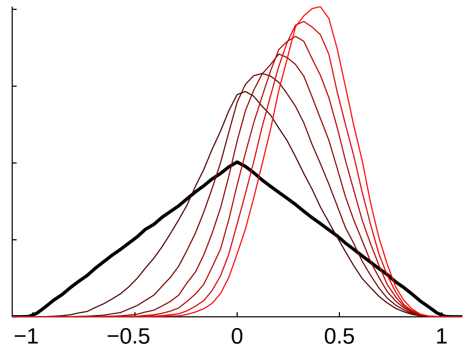
The output distributions of L_nU_n and U_nL_n are given by

$$\begin{aligned}
F_{L_nU_n} &= F^{n+1} - nF^{n+1}(1-F) + F^{2n+2}(1-F) \\
&\quad + \frac{1}{2}(n-1)(n+2)F^{2n+2}(1-F)^2 \\
F_{U_nL_n} &= 1 - (1-F)^{n+1} + nF(1-F)^{n+1} - F(1-F)^{2n+2} \\
&\quad - \frac{1}{2}(n-1)(n+2)F^2(1-F)^{2n+2}
\end{aligned}$$

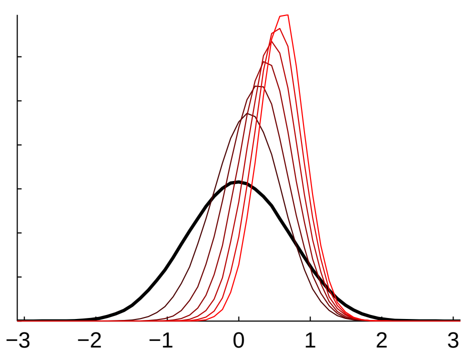
This result agrees with that found in [4]. The density functions of $L_n U_n$ are shown in Figure 3.6.



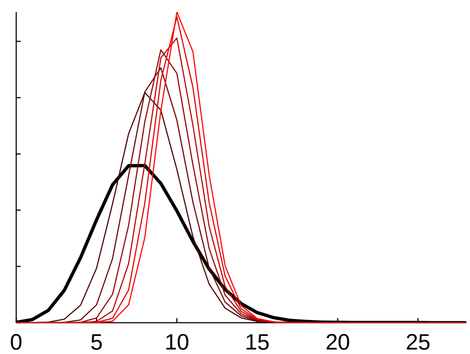
(a) Uniform



(b) Linear



(c) Normal



(d) Poisson

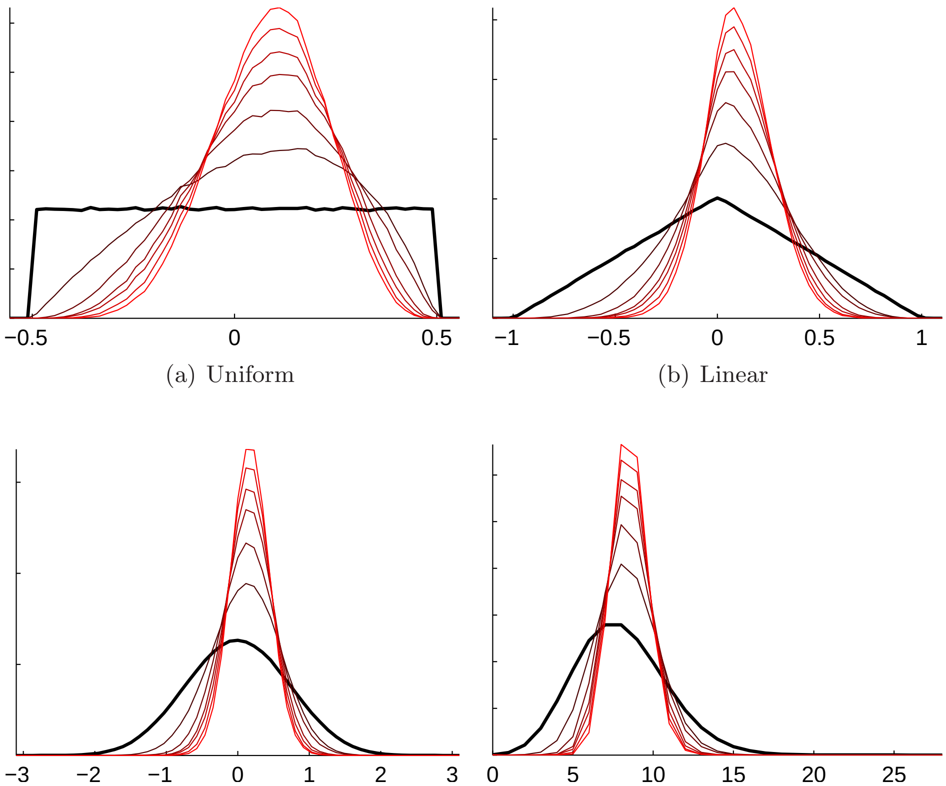
Figure 3.6: Density functions of $L_n U_n$

The output distributions of C_n

The output distributions of C_n are given in the following two recursive forms:

$$(a) \quad F_{C_n} = F_{C_{n-1}} + n(G_{2n} - G_{2n-1})$$

$$(b) \quad F_{C_n} = F + \sum_{k=1}^{n} k(G_{2k} - G_{2k-1})$$



(a) Uniform (b) Linear

(c) Normal (d) Poisson

Figure 3.7: Density functions of C_n

where $G_n(F)$ is a probability defined in Section 4.5.5. The following set of recursive formulae can be used to calculate $G_n(F)$ for $n = 1 \dots 12$:

$$
\begin{aligned}
G_n(F) &= I_n^2(F)J_n(F) \\
I_n(F) &= F^2(1-F)\Big[H_n(F) - n \bmod 2\Big] \\
H_n(F) &= H_{n-2}(F) + I_{n-1}(F)J_{n-1}(F) \\
H_1 &= 0 \\
H_2 &= 1
\end{aligned}
$$

where $J_n(F)$ for $n = 1 \dots 12$ is listed in table 3.1.

The formulae for the output distributions of C_n hold for all values of n. The formulae to calculate G_n are only known for $n = 1 \dots 12$. We can therefore calculate the exact form of F_{C_n} for $n = 1 \dots 6$.

n	$J_n(F)$	n	$J_n(F)$
1	F^{-3}	2	$(1-F)^{-1}$
3	F^{-2}	4	1
5	F^{-1}	6	1
7	1	8	$1-F^2$
9	1	10	$1-2F^2+F^3$
11	$2F-F^2$	12	$1-2F^2+F^4$

Table 3.1: Values of $J_n(F)$

Although we did not endeavour to derive the output distribution of F_n, we know that it would have a similar structure to that of C_n, because they are duals and their output distributions are therefore related by the equation

$$F_{F_n}(F_x) = 1 - F_{C_n}(1 - F(x)).$$

Until recently we only had results for the distribution transfer of U_n, L_nU_n and their duals [4]. The output distribution of C_n, for $n = 1, 2, 3$, were known. C_n occurs naturally in the DPT and the new results for its distribution transfer means that we are now in a much better position to study the handling of signal in the DPT.

3.4 The mean and variance

Given a continuous distribution f, with the associated cumulative distribution $F(x) = \int_{-\infty}^{x} f(t)\,dt$ and an independently, identically generated sequence x of numbers from this distribution, it is of interest how a given smoother S handles this "noise". We can gain insight from the moments of $S(x)$, which are defined by

$$\mu_1 = \int_{-\infty}^{\infty} t f(t)\,dt \text{ and } \mu_2 = \int_{-\infty}^{\infty} (t-\mu_1)^2 f(t)\,dt.$$

where μ_1 is the first non-central mean and μ_2 is the first central variance.

Comparing the mean of $S(x)$ with the corresponding moment of x gives us an indication of the bias of S. The variance is a measure of smoothness and we would compare the variance of $S(x)$ with the second moment of x to see if it is in fact reduced by S, as expected.

Next we will use U_n and L_nU_n, of which the distribution functions were derived in the previous sections, to calculate the moments of those smoothers. We will take the input to be from the uniform, the linear spline and the normal distributions. We will show analytical results for the moments in the case of the uniform distribution.

The uniform distribution

First we apply U_n to a sequence from the uniform distribution. Differentiating the output distribution of U_n gives us its density function

$$f_{U_n} = (n+1)^2 F^n f_x - n(n+2)F^{n+1} f_x$$

Substitution gives (with $t \in [-\frac{1}{2}, \frac{1}{2}]$)

$$f_{U_n} = (n+1)^2 (t+\frac{1}{2})^n - n(n+2)(t+\frac{1}{2})^{n+1}$$

The first two moments in terms of n are

$$\begin{aligned}
\mu_1 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} t f_{U_n}(t)dt = \frac{n(n+1)}{2(n+2)(n+3)} \\
\mu_2 &= \int_{-\frac{1}{2}}^{\frac{1}{2}} (t-\mu_1)^2 f_{U_n}(t)dt = \frac{2n^3+10n^2+9n+12}{(n+2)^2(n+3)^2(n+4)}
\end{aligned}$$

The result for L_n is similar with

$$\begin{aligned}
\mu_1 &= -\frac{n(n+1)}{2(n+2)(n+3)} \\
\mu_2 &= \frac{2n^3+10n^2+9n+12}{(n+2)^2(n+3)^2(n+4)}
\end{aligned}$$

U_n and L_n are symmetric about the mean of a uniform distribution (i.e. 0).

Now we apply L_nU_n. The density function of L_nU_n is given by

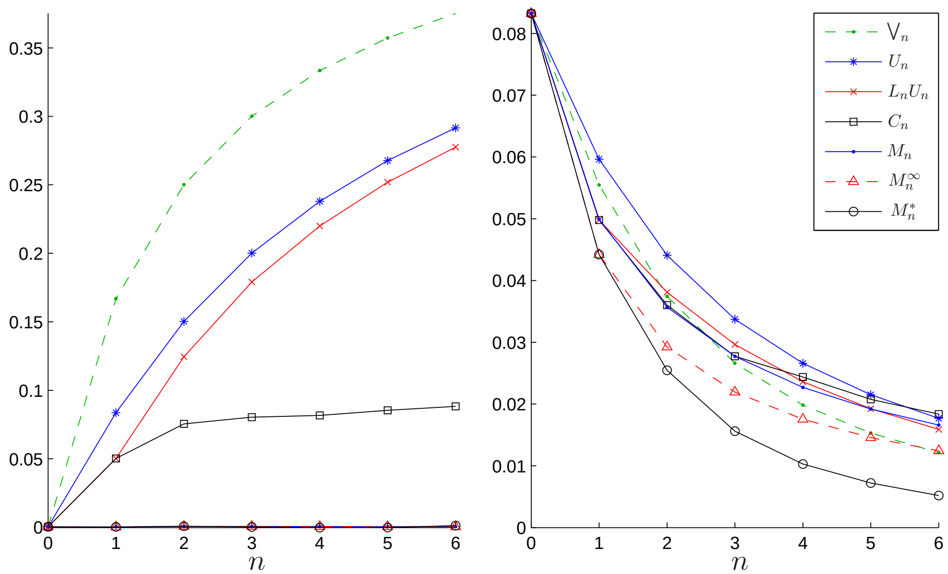
$$\begin{aligned}
f_{L_nU_n} &= f_x F^n \left[(1-n^2) + n(n+2)F + n(n+1)^2 F^{n+1} \right. \\
&\quad \left. -(2n^3+5n^2+n-3)F^{n+2} + (n-1)(n+2)^2 F^{n+3}\right]
\end{aligned}$$

Substituting for the uniform distribution and integrating gives for the first two moments of L_nU_n

$$\mu_1 \;=\; \frac{-9+3\,n+25\,n^2+19\,n^3+4\,n^4}{2\,(2+n)\,(3+n)\,(3+2\,n)\,(5+2\,n)}$$

$$\mu_2 = \frac{8316+32031\,n+52464\,n^2+47462\,n^3+25702\,n^4+8366\,n^5+1520\,n^6+119\,n^7}{4\,(2+n)^2\,(3+n)^2\,(4+n)\,(3+2\,n)^2\,(5+2\,n)^2}$$

Again, L_nU_n and U_nL_n are symmetric about the mean of a uniform distribution, since they are duals.



(a) Mean (b) Variance

Figure 3.8: Moments for the uniform distribution

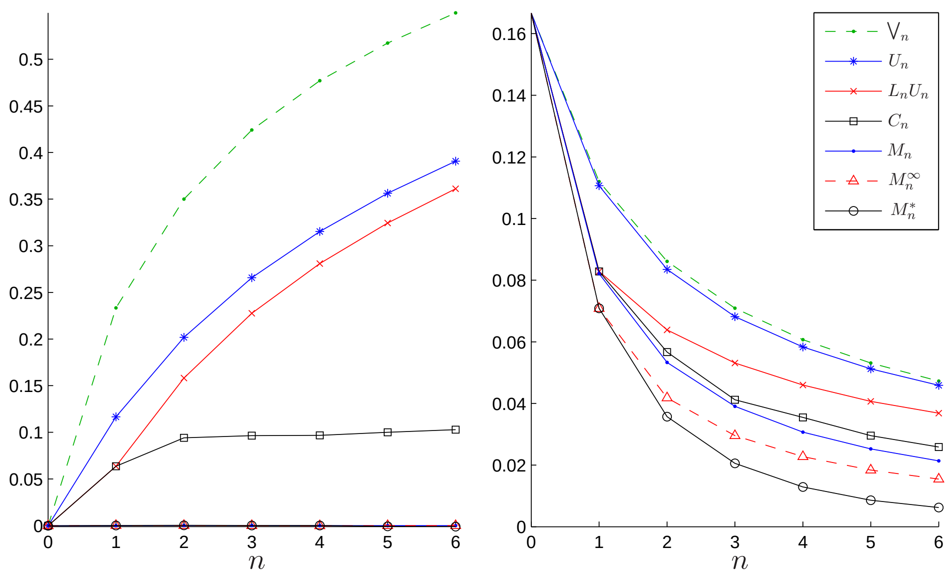
The linear spline distribution

When applying U_n to a sequence from the linear spline distribution, the density function is

$$f_{U_n} = \begin{cases} \frac{1}{2^{n+1}}(1+t)q_+^n\left[2(n+1)^2 - n(n+2)q_+\right], & t\in[-1,0] \\ \frac{1}{2^{n+1}}(1-t)q_-^n\left[2(n+1)^2 - n(n+2)q_-\right], & t\in[0,1] \\ 0, & elsewhere \end{cases}$$

where $q_{\pm} = (1 + 2t \pm t^2)$.

If we substitute for the linear spline distribution function in the expression for $f_{L_n U_n}$ and calculate the first two moments we get the plots shown in Figure 3.9.



(a) Mean (b) Variance

Figure 3.9: Moments for the linear distribution

The normal and Poisson distributions

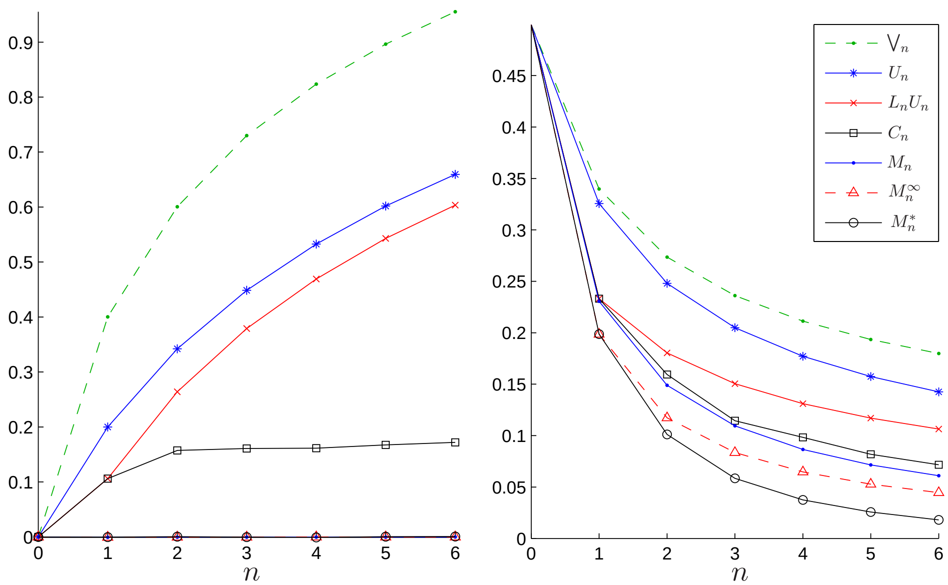
We are considering the limit of the B-spline distributions (described in Appendix A), which is the normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. The normal distribution involves the error function $\text{erf}(x)$, so we only show numerical results in Figure 3.10.

For the Poisson distribution we only show numerical results in Figure 3.11.

Discussion of the moments

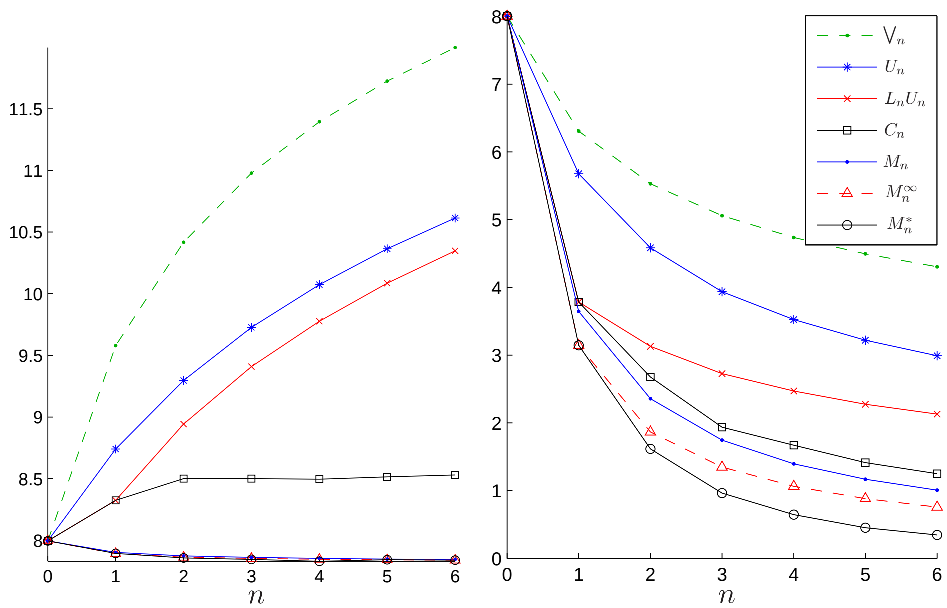
From the graphs of the moments, Figures 3.8 to 3.11, it can be seen that

$$\mu_1(C_n) \leq \mu_1(L_n U_n) \leq \mu_1(U_n) \leq \mu_1(\bigvee_n)$$



(a) Mean (b) Variance

Figure 3.10: Moments for the normal distribution



(a) Mean (b) Variance

Figure 3.11: Moments for the poisson distribution

which can be expected from the LULU intervals. The mean of C_n grows much slower than the mean of U_n and L_nU_n, as n increases. This means that for a sequence containing noise, C_n would be a better estimator of the underlying signal.

For the median and its variants it can be seen that, in the case of the B-spline distributions which are symmetric,

$$\mu_1(M_n) \approx \mu_1(M_n^\infty) \approx \mu_1(M_n^*) \approx 0$$

since M_n, and M_n^∞ by implication are symmetric operators. M_n^* is not a symmetric operator but its mean is still approximately zero. For the Poisson distribution, which is unsymmetric, we have that

$$\mu_1(M_n) \approx \mu_1(M_n^\infty) \approx \mu_1(M_n^*) < 0$$

which is to be expected since the Poisson distributed is slightly skewed to the left.

We can see clearly that the variance of the original sequence is reduced by all smoothers studied.

3.5 Alternating LULU smoothers and their distributions

Composite LULU smoothers other than C_n and F_n can be constructed. We can construct a smoother K_n in the following manner

$$K_n = \begin{cases} L_nU_nU_{n-1}L_{n-1}\ldots U_2L_2L_1U_1, & n \text{ uneven} \\ U_nL_nL_{n-1}U_{n-1}\ldots U_2L_2L_1U_1, & n \text{ even} \end{cases}$$

Note that K_n is similar to C_n with the order of every second pair of operators swapped.

We compare the moments of C_n and K_n to observe the effect of this change. Figure 3.12 shows the first moment of the output for different n when applied to a sequence with elements from a normal distribution and Figure 3.13 shows similar results for the second moment.

From these two graphs we notice that the average of K_n is significantly lower than C_n even already for small values of n. The variances have a negligible

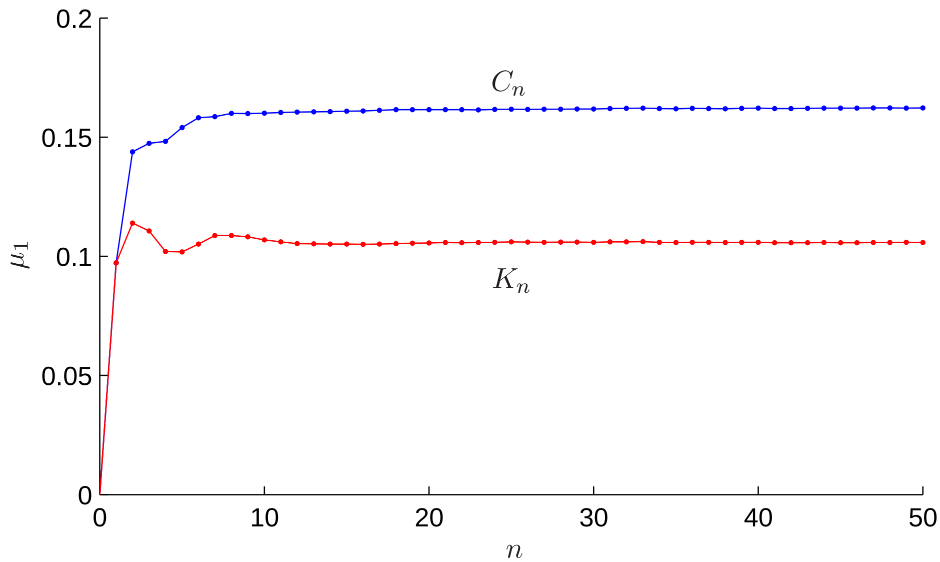


Figure 3.12: Average of C_n and K_n

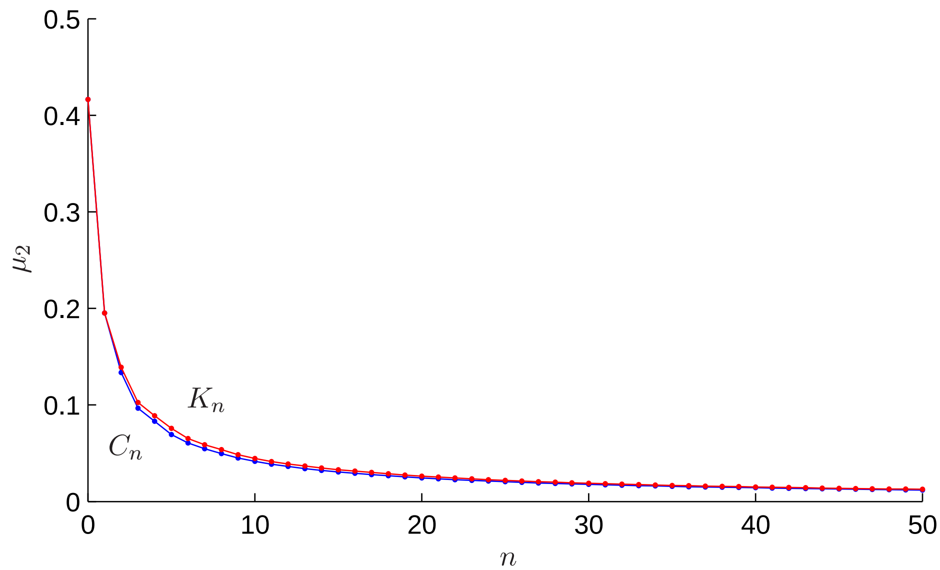


Figure 3.13: Variance of C_n and K_n

difference. This means that K_n is less biased than C_n when applied to normally distributed data.

It is easy to show that $K_n \leq C_n$ in general. This means that K_n applied to a sequence is a closer approximation than C_n.

These observations favour the use of K_n above C_n. However, a more thorough study of the output of C_n and K_n and their residuals would be necessary in order to understand the differences between them and the importance of each one.

3.6 Characterising an unknown error distribution

If one makes a decomposition of a sequence x with independently, identically distributed elements using a LULU smoother, it can be demonstrated that the number of pulses in the decomposition does not exceed the number of nonzero elements in the sequence. In the case of decomposition with $C_n(x)$ it can be shown that exactly a $\frac{1}{3}$ of the downward pulses and exactly a $\frac{1}{5}$ of the upward pulses appear in the first resolution level $D_1(x)$ regardless of the distribution of the sequence [17]. Thus more than half of all the pulses appear in the first resolution level.

The probability of a single pulse appearing in $D_1(x)$ would be small if a sequence is highly correlated. In this case, if the number of downward pulses in $D_1(x)$ is close to $\frac{1}{3}$ of the total number of downward pulses, we can assume that most of these are random noise added to the sequence.

We can therefore ask whether it is possible to estimate the noise in a signal using $D_1(x)$. How much economising is possible by omitting most of these? And what will the level of signal distortion be? This was investigated by Rohwer[17], and we show the results here.

We consider the amplitude of the downward pulses in $D_1(x)$. We assume the random noise to have a B-spline distribution of order n, and then increase n to see if some limit is approached. It is found that the mean of the downward pulses converges to a multiple of the standard deviation σ_x of the sequence x. Denote by y the sequence of negative values in $x - U_1 x$. Then we have

$$\lim_{n\to\infty} \mu_1(y) = k\sigma_x$$

Figure 3.6 shows a plot of the mean of the downward pulses in $D_1(x)$ divided by the standard deviation of the input sequence x with a B-spline

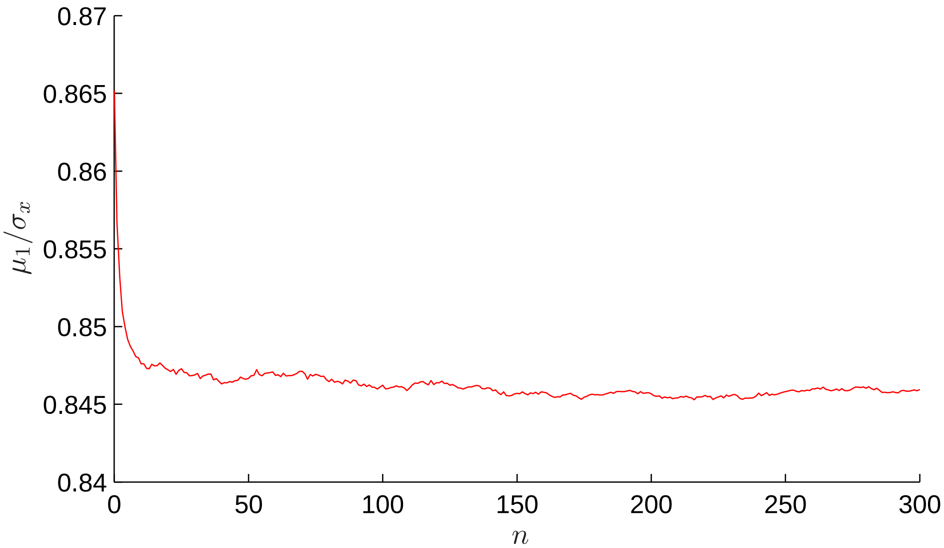


Figure 3.14: Estimating σ_x of a normal error distribution using D_1x

distribution B_n. We confirm the analytical result in [17] that this value tends to a limit of $k = \frac{3}{2\sqrt{\pi}} \approx 0.846$, as can be seen in the graph.

Since the limit of the B-spline distributions is the normal distribution, which is characterised by only two parameters, namely its first two moments, this result means that the second moment of normally distributed noise can be estimated, under certain conditions, from the input data using the first resolution level.

3.7 The asymptotic distributions

An interesting problem is to find the limiting distributions of the LULU smoothers as the window size n tends to infinity. Since they are constructed from the maximum and minimum operators, one has to turn to Extreme Value Theory (EVT) for an answer.

Let x be a sequence of i.i.d. random variables from a distribution F_x and let $x_{n,n} = \bigvee_n \{x_i, \ldots x_{i+n}\}$ be the maximum value. In [4] the Fischer-Tippett theorem [6] is employed to show that, if there exist constant sequences $a_n > 0$

and b_n such that the distribution of $\frac{x_{n,n}-b_n}{a_n}$ converges as $n \to \infty$, then the limiting distribution H can have one of three possible forms . These are the Fréchet, Weibull or Gumbel distributions, given by the following three density functions:

$$\begin{aligned}
\text{Fréchet:} \qquad \Phi_\alpha(t) &= \begin{cases} 0, & t \leq 0 \\ \exp\{-t^{-\alpha}\}, & t > 0 \end{cases} \\
\text{Weibull:} \qquad \Psi_\alpha(t) &= \begin{cases} \exp\{-(-t)^{\alpha}\}, & t \leq 0 \\ 0, & t > 0 \end{cases} \\
\text{Gumbel:} \qquad \Lambda(t) &= \exp\{-e^{-t}\},\ t \in \mathbb{R}
\end{aligned}$$

In this case we say F_x lies in the maximum domain of attraction of H and write $F_x = \mathrm{MDA}(H)$.

If we apply this to the LULU smoothers, we find the following results for U_n and L_nU_n:

$$F_{U_n}(a_n t + b_n) \xrightarrow{D} H(t) - H(t)\log H(t) \equiv J(t)$$

$$F_{L_nU_n}(a_n t + b_n) \xrightarrow{D} H(t) - H(t)\log H(t) + \frac{1}{2}[H(t)\log H(t)]^2 \equiv G(t)$$

where $\xrightarrow{D}$ indicates convergence in distribution. The proof is given in Conradie et al[4].

Now let $x_{n-k+1,n}$ denote the k-th largest order statistic. Then for constant sequences $a_n > 0$ and b_n, it is known that

$$P\Big(\frac{x_{n-k+1,n} - b_n}{a_n} \leq t\Big) \to H_k(t)$$

where

$$H_k(t) \ = \ H(t) \sum_{m=0}^{k-1} \frac{(-\log H(t))^k}{k!}$$

The second and third largest order statistics have the following distributions

$$H_2(t) \ = \ H(t)[1 - \log H(t)]$$

$$H_3(t) \ = \ H(t)\left[1 - \log H(t) + \frac{1}{2}[\log H(t)]^2\right]$$

Thus we have that

$$J(t) = H_2(t) \leq G(t) \leq H_3(t)$$

This means that U_n has the same asymptotic distribution as the second largest order statistic, and L_nU_n is stochastically bounded between the second and third largest order statistics.

The interesting question is what the asymptotic distribution of C_n is, which occur naturally in the DPT. Currently no result is available.

3.8 Summary

In this chapter we discussed the distributions of the median smoothers M_n, M_n^∞, M_n^*, as well as $\bigvee_n$ and the LULU smoothers U_n, L_nU_n, C_n and their duals and considered these distributions and their moments for various input distributions. An alternative LULU smoother to C_n was considered and was found to be less biased to input.

We showed that the first resolution level of the DPT can be used to estimate the variance of an underlying normal error distribution due to the relationship between the mean and standard deviation.

We also investigated the asymptotic distributions of LULU smoothers as the window size tends to infinity and found that L_nU_n is stochastically bounded between the second and third largest order statistics.

Chapter 4

Derivation of the distributions

4.1 Introduction

The main aim of this chapter is to find an expression for the distribution of C_n. The work presented here contains new results and will be submitted in the near future [1].

In the next section we point out the difficulties calculating the distribution. Section 4.3 defines a notation to simplify the algebra and section 4.4 derives formulas used in the following sections.

Sections 4.5.2 through 4.5.5 derive the distributions $F_{\bigvee_n}$ (and $F_{\bigwedge_n}$), F_{L_n} (and F_{U_n}), $F_{L_n U_n}$ (and $F_{U_n L_n}$), F_{C_n} (and F_{F_n}), respectively. Section 4.6 finds an expression for G_n (which is defined in section 4.5.5).

Section 4.7 lists a Mathematica program for calculating the distribution of C_n recursively.

4.2 Problems calculating the distributions of C_n

Before a recursive formula was found for finding the desired expression, the distribution of C_n had to be derived from first principles. First I show the calculation of $F_{C_1} = F_{LU}$ and then discuss the difficulties that arise.

Let x be a doubly infinite sequence of variables coming from an identically independent distribution. Now define

$$
\begin{aligned}
y_i &= \bigvee(x_i, x_{i+1}) \\
z_i &= \bigwedge(y_i, y_{i-1}) \\
a_i &= \bigwedge(z_i, z_{i-1}) \\
b_i &= \bigvee(a_i, a_{i+1})
\end{aligned}
$$

Then $b = L_1U_1(x)$.

$$
\begin{aligned}
&F_{LU}(t) \\
&= P(b_0 \le t) \\
&= P(a_0, a_1 \le t) \\
&= P(a_0 \le t) - P(a_0 \le t, a_1 > t) \\
&= 1 - P(a_0 > t) - P(a_1 > t) + P(a_0, a_1 > t) \\
&= 1 - 2P(a_0 > t) + P(a_0, a_1 > t) \\
&= 1 - 2P(z_0, z_{-1} > t) + P(z_1, z_0, z_{-1} > t) \\
&= 1 - 2P(y_0, y_{-1}, y_{-2} > t) + P(y_1, y_0, y_{-1}, y_{-2} > t) \\
&= 1 - 2P(y_0, y_1, y_2 > t) + P(y_0, y_1, y_2, y_3 > t) \\
&= 1 - 2P(y_0, y_1, y_2 > t) + P(y_0, y_1, y_2 > t) \\
&\quad - P(y_0, y_1, y_2 > t, y_3 \le t) \\
&= 1 - P(y_0, y_1, y_2 > t) - P(y_0, y_1, y_2 > t, y_3 \le t)
\end{aligned}
$$

$$
\begin{aligned}
&P(y_0, y_1, y_2 > t) \\
&= P(y_0, y_1 > t) - P(y_0, y_1 > t, y_2 \le t) \\
&= P(y_0 > t) - P(y_0 > t, y_1 \le t) - P(y_0 > t, y_2 \le t) \\
&\quad + P(y_0 > t, y_1, y_2 \le t) \\
&= 1 - P(y_0 \le t) - P(y_1 \le t) + P(y_0, y_1 \le t) - P(y_2 \le t) \\
&\quad + P(y_0, y_2 \le t) + P(y_1, y_2 \le t) - P(y_0, y_1, y_2 \le t) \\
&= 1 - 3P(y_0 \le t) + 2P(y_0, y_1 \le t) + P(y_0, y_2 \le t)
\end{aligned}
$$

$$- P(y_0, y_1, y_2 \leq t)$$

Similarly

$$\begin{aligned}
&P(y_0, y_1, y_2 > t, y_3 \leq t) \\
&= P(y_3 \leq t) - P(y_0, y_3 \leq t) - P(y_1, y_3 \leq t) + P(y_0, y_1, y_3 \leq t) \\
&\quad - P(y_2, y_3 \leq t) + P(y_0, y_2, y_3 \leq t) + P(y_1, y_2, y_3 \leq t) \\
&\quad - P(y_0, y_1, y_2, y_3 \leq t) \\
&= P(y_0 \leq t) - P(y_0, y_3 \leq t) - P(y_0, y_2 \leq t) + P(y_0, y_1, y_3 \leq t) \\
&\quad - P(y_0, y_1 \leq t) + P(y_0, y_2, y_3 \leq t) + P(y_0, y_1, y_2 \leq t) \\
&\quad - P(y_0, y_1, y_2, y_3 \leq t)
\end{aligned}$$

and so we have

$$\begin{aligned}
&F_{LU}(t) \\
&= 2P(y_0 \leq t) - P(y_0, y_1 \leq t) + P(y_0, y_3 \leq t) - P(y_0, y_1, y_3 \leq t) \\
&\quad - P(y_0, y_2, y_3 \leq t) + P(y_0, y_1, y_2, y_3 \leq t) \\
&= 2P(x_0, x_1 \leq t) - P(x_0, x_1, x_2 \leq t) + P(x_0, x_1, x_3, x_4 \leq t) \\
&\quad - P(x_0, x_1, x_2, x_3, x_4 \leq t) - P(x_0, x_1, x_2, x_3, x_4 \leq t) \\
&\quad + P(x_0, x_1, x_2, x_3, x_4 \leq t) \\
&= 2P(x_0, x_1 \leq t) - P(x_0, x_1, x_2 \leq t) + P(x_0, x_1, x_3, x_4 \leq t) \\
&\quad - P(x_0, x_1, x_2, x_3, x_4 \leq t) \\
&= 2F_x^2 - F_x^3 + F_x^4 - F_x^5
\end{aligned}$$

Clearly, the more $\bigvee$ and $\bigwedge$ operators in the LULU operator under consideration, the more complex will be the derivation of its output distribution.

The distributions are polynomials in the input distribution and the degree of the polinomials grows approximately quadratically. For $n = 1, 2$ we could still calculate it by hand, but it was calculated for $n = 3$ on a computer by Wild using a more general method [1]. The next table shows the degree of the polynomials (derived as above) for different values of n.

n	Degree
1	5
2	12
3	22

Table 4.1: Degree of the distribution polinomials

4.3 Definitions

For the derivation of the distributions a notation was developed to simplify the expression for the probabilities. The result was that a few patterns could be identified that would have been difficult otherwise.

Our approach differs from that followed in [4] and [9]. Whereas one of their expressions for the output distribution might be in terms of the elements of two or more related sequences, we are careful to refer to only one sequence in any given expression.

We start by defining the notation.

Let y be a doubly infinite sequence of identically distributed variables. Let i and j be subsequences of $\{0, 1, \ldots, n\}$.
Then the notation $\{i_1, i_2, \ldots, i_k, \overline{j_1}, \overline{j_2}, \ldots, \overline{j_l}\}_y$ denotes the set/event

$$\begin{aligned}\{y: \quad & y_{i_1} \leq t, y_{i_2} \leq t, \ldots, y_{i_k} \leq t, \\ & y_{j_1} > t, y_{j_2} > t, \ldots, y_{j_l} > t\}\end{aligned}$$

where $|i| = k$ and $|j| = l$.
Also, the parameter $(i_1, i_2, \ldots, i_k, \overline{j_1}, \overline{j_2}, \ldots, \overline{j_l})_y$ denotes the probability

$$P\left(\{i_1, i_2, \ldots, i_k, \overline{j_1}, \overline{j_2}, \ldots, \overline{j_l}\}_y\right)$$

In practice we shall omit the commas, writing for example $(0\,1\,2\,\overline{3}\,\overline{4})_y$ instead of $(0, 1, 2, \overline{3}, \overline{4})_y$.
Furthermore the notation is invariant to permutations of the argument, e.g. $(01\overline{2})_y = (\overline{2}01)_y$.

The following is true

$$(i, i)_b = (i)_b$$

$$(i, \overline{i})_b = 0$$

4.4 Types of operations with the vectors

Now we look at the basic operations that are performed on the vectors. These include shifting of the indexes (the components of the vectors), inverting the vector and expanding the vector (explained in that section).

4.4.1 Shifting

Given a sequence x of identically distributed variables, it is true that

$$P(x_i \leq t) = P(x_0 \leq t), \quad i \in \mathbb{Z}$$

Written in the notation presented in the previous section, the statement becomes

$$(i)_x = (0)_x$$

For a probability involving indexes i, j, k it follows that

$$(ijk)_x = (0, j-i, k-i)_x$$

4.4.2 Inverting

Inverting the vector would be to go from an expression involving ">" to one involving "≤" or vice versa in the original probability notation.

We have already expressed the probability $P(x_i > t)$ as

$$(\bar{i})_x = 1 - (i)_x$$

The idea is to rewrite the probability vector in a form that can be expanded using the formules in the next section. Theorem 1 rewrites the vector without shifting any indexes at any point and Theorem 3 involves shifting.

Theorem 1 (Inverting without shifting). *Let x be a doubly infinite sequence of identically distributed variables. Then*

$$(a) \quad (\overline{0 \ldots n})_x = 1 - (0)_x - \sum_{i=0}^{n-1} (\overline{0 \ldots i}, i+1)_x$$

$$(b) \quad (\overline{0 \ldots n})_x = 1 - (n)_x - \sum_{i=0}^{n-1} (n-i-1, \overline{n-i, \ldots, n})_x$$

Proof.

(a)

$$\begin{aligned}
(\overline{0\ldots n})_x &= (\overline{0\ldots n-1})_x - (\overline{0\ldots n-1}, n)_x \\
&= (\overline{0\ldots n-2})_x - (\overline{0\ldots n-2}, n-1)_x - (\overline{0\ldots n-1}, n)_x \\
&\vdots \\
&= (\overline{0})_x - (\overline{0}1)_x - (\overline{01}2)_x - \ldots - (\overline{0\ldots n-1}, n)_x \\
&= 1 - (0)_x - (\overline{0}1)_x - (\overline{01}2)_x - \ldots - (\overline{0\ldots n-1}, n)_x \\
&= 1 - (0)_x - \sum_{i=0}^{n-1} (\overline{0\ldots i}, i+1)_x
\end{aligned}$$

(b) The derivation is similar as in (a).

$$\begin{aligned}
(\overline{0\ldots n})_x &= (\overline{1\ldots n})_x - (0, \overline{1\ldots n})_x \\
&= (\overline{2\ldots n})_x - (1, \overline{2\ldots n})_x - (0, \overline{1\ldots n})_x \\
&\vdots \\
&= (\overline{n})_x - (n-1, \overline{n})_x - (n-2, \overline{n-1, n})_x - \ldots - (0\overline{1\ldots n})_x \\
&= 1 - (n)_x - (n-1, \overline{n})_x - (n-2, \overline{n-1, n})_x - \ldots \\
&\quad - (0, \overline{1\ldots n})_x \\
&= 1 - (n)_x - \sum_{i=0}^{n-1} (n-i-1, \overline{n-i, \ldots, n})_x
\end{aligned}$$

□

Corollary 2. *Let x be a doubly infinite sequence of identically distributed variables. Then*

$$(a) \quad (0\ldots n)_x = 1 - (\overline{0})_x - \sum_{i=0}^{n-1} (0\ldots i, \overline{i+1})_x$$

$$(b) \quad (0\ldots n)_x = 1 - (\overline{n})_x - \sum_{i=0}^{n-1} (\overline{n-i-1}, n-i, \ldots, n)_x$$

Theorem 3 (Inverting with shifting). *Let x be a doubly infinite sequence of identically distributed variables. Then*

$$(\overline{0\ldots n})_x = 1 - [n+1](0)_x + \sum_{i=0}^{n-1} [n-i](0, \overline{1\ldots i}, i+1)_x$$

Proof.

$$
\begin{aligned}
&(\overline{0\ldots k}, k+1)_x \\
&= (k+1)_x - (k,k+1)_x - \sum_{i=0}^{k-1}(k-i-1, \overline{k-i,\ldots,k}, k+1)_x \\
&= (0)_x - (01)_x - \sum_{i=0}^{k-1}(0, \overline{1\ldots i+1}, i+2)_x
\end{aligned}
$$

$$
\begin{aligned}
&(\overline{0\ldots n})_x \\
&= 1 - (0)_x - \sum_{k=0}^{n-1}(\overline{0\ldots k}, k+1)_x \\
&= 1 - (0)_x - \sum_{k=0}^{n-1}\left[(0)_x - (01)_x - \sum_{i=0}^{k-1}(0, \overline{1\ldots i+1}, i+2)_x\right] \\
&= 1 - [n+1](0)_x + n(01)_x + \sum_{k=0}^{n-1}\sum_{i=0}^{k-1}(0, \overline{1\ldots i+1}, i+2)_x \\
&= 1 - [n+1](0)_x + n(01)_x + \sum_{k=1}^{n-1}\sum_{i=0}^{k-1}(0, \overline{1\ldots i+1}, i+2)_x \\
&= 1 - [n+1](0)_x + n(01)_x \\
&\quad + (0\overline{1}2)_x \\
&\quad + (0\overline{1}2)_x + (0\overline{12}3)_x \\
&\qquad \vdots \\
&\quad + (0\overline{1}2)_x + (0\overline{12}3)_x + \ldots + (0, \overline{1\ldots n-1}, n)_x \\
&= 1 - [n+1](0)_x + n(01)_x + [n-1](0\overline{1}2)_x + [n-2](0\overline{12}3)_x + \ldots \\
&\quad + (0, \overline{1\ldots n-1}, n)_x \\
&= 1 - [n+1](0)_x + \sum_{i=0}^{n-1}[n-i](0, \overline{1\ldots i}, i+1)_x \qquad \square
\end{aligned}
$$

Corollary 4. *Let x be a doubly infinite sequence of identically distributed variables. Then*

$$
(0\ldots n)_x = 1 - [n+1](\overline{0})_x + \sum_{i=0}^{n-1}[n-i](\overline{0}, 1\ldots i, \overline{i+1})_x
$$

4.4.3 Expansion

Let x and y be doubly infinite sequences of identically distributed variables. Expansion would occur when we take a probability that is a function of values of $y = y(x)$ and rewrite it in terms of the values of x. Thus the statement

$$P(y_0 \leq t) = P(x_0 \leq t, x_1 \leq t)$$

in terms of the vectors would be

$$(0)_y = (01)_x$$

We will need the following result.

Theorem 5. $\bigvee_1^n = \bigvee_n$ *and* $\bigwedge_1^n = \bigwedge_n$

Proof.

$$\begin{aligned}(\textstyle\bigvee_1^n x)_i &= \max\{\max\{x_i, x_{i+1}\}, \ldots, \max\{x_{i+n-1}, x_{i+n}\}\}\\ &= \max\{x_i, \ldots, x_{i+n}\}\\ &= (\textstyle\bigvee_n x)_i\end{aligned}$$

Similarly $\bigwedge_1^n = \bigwedge_n$. □

The next two theorems show how the vectors $\{\overline{0}, 1 \ldots n-1, \overline{n}\}_b$ and $\{0, \overline{1 \ldots n-1}, n\}_b$, respectively, look when expanded by $\bigvee_r$.

Theorem 6 (Expansion Form 1). *Let* $b = \bigvee_r(a)$. *Then*

$$(\overline{0}, 1 \ldots n-1, \overline{n})_b = (\overline{0}, 1 \ldots n+r-1, \overline{n+r})_a$$

Proof.

$$\begin{aligned}&(\overline{0}, 1 \ldots n-1, \overline{n})_b\\ &= (\overline{0}, 1 \ldots n-1)_b - (\overline{0}, 1 \ldots n)_b\\ &= (1 \ldots n-1)_b - (0 \ldots n-1)_b - (1 \ldots n)_b + (0 \ldots n)_b\\ &= (1 \ldots n+r-1)_a - (0 \ldots n+r-1)_a - (1 \ldots n+r)_a + (0 \ldots n+r)_a\\ &= (\overline{0}, 1 \ldots n+r-1)_a - (\overline{0}, 1 \ldots n+r)_a\\ &= (\overline{0}, 1 \ldots n+r-1, \overline{n+r})_a\end{aligned}$$

□

Corollary 7. *Let $b = \bigwedge_r(a)$. Then*

$$(0, \overline{1 \ldots n-1}, n)_b = (0, \overline{1 \ldots n+r-1}, n+r)_a$$

Theorem 8 (Expansion Form 2). *Let $b = \bigvee_r(a)$. Then*

$$\begin{aligned}
&\{0, \overline{1 \ldots n-1}, n\}_b \\
&= \{0 \ldots r, \overline{r+1}\}_a \left[\bigcap_{i=r+2}^{n-2-r} \{\Omega - \{i \ldots i+r\}_a\}\right] \{\overline{n-1}, n \ldots n+r\}_a
\end{aligned}$$

Proof.

$$\begin{aligned}
&\{0, \overline{1 \ldots n-1}, n\}_b \\
&= \{0\}_b \left[\bigcap_{i=1}^{n-1} \{\Omega - \{i\}_b\}\right] \{n\}_b \\
&= \{0 \ldots r\}_a \left[\bigcap_{i=1}^{n-1} \{\Omega - \{i \ldots i+r\}_a\}\right] \{n \ldots n+r\}_a \\
&= \{0 \ldots r\}_a \cap \{\Omega - \{1 \ldots r+1\}_a\} \left[\bigcap_{i=2}^{n-2} \{\Omega - \{i \ldots i+r\}_a\}\right] \\
&\quad \cap \{\Omega - \{n-1 \ldots n-1+r\}_a\} \cap \{n \ldots n+r\}_a \\
&= \{0 \ldots r\}_a \cap \{\Omega - \{r+1\}_a\} \left[\bigcap_{i=2}^{n-2} \{\Omega - \{i \ldots i+r\}_a\}\right] \\
&\quad \cap \{\Omega - \{n-1\}_a\} \cap \{n \ldots n+r\}_a \\
&= \{0 \ldots r, \overline{r+1}\}_a \left[\bigcap_{i=2}^{n-2} \{\Omega - \{i \ldots i+r\}_a\}\right] \{\overline{n-1}, n \ldots n+r\}_a \\
&= \{0 \ldots r, \overline{r+1}\}_a \left[\bigcap_{i=2}^{r+1} \{\Omega - \{i \ldots r+1 \ldots i+r\}_a\}\right] \\
&\quad \times \left[\bigcap_{i=r+2}^{n-2-r} \{\Omega - \{i \ldots i+r\}_a\}\right] \\
&\quad \times \left[\bigcap_{i=n-1-r}^{n-2} \{\Omega - \{i \ldots n-1 \ldots i+r\}_a\}\right] \{\overline{n-1}, n \ldots n+r\}_a \\
&= \{0 \ldots r, \overline{r+1}\}_a \left[\bigcap_{i=r+2}^{n-2-r} \{\Omega - \{i \ldots i+r\}_a\}\right] \{\overline{n-1}, n \ldots n+r\}_a \qquad \square
\end{aligned}$$

Corollary 9. *Let* $b = \bigwedge_r(a)$. *Then*

$$\begin{aligned}
&\{\overline{0}, 1 \ldots n-1, \overline{n}\}_b \\
&= \{\overline{0 \ldots r}, r+1\}_a \left[\bigcap_{i=r+2}^{n-2-r} \{\Omega - \{\overline{i \ldots i+r}\}_a\} \right] \{n-1, \overline{n \ldots n+r}\}_a
\end{aligned}$$

Theorem 8 is true for all values of r and n. Theorem 10 gives results for specific cases.

Theorem 10. *Let* $b = \bigvee_r(a)$.

$$(0, \overline{1 \ldots n-1}, n)_b = \begin{cases} 0, & n \leq r \\ (0 \ldots r, \overline{r+1, n-1}, n \ldots n+r)_a, & r < n < 2r+4 \end{cases}$$

Proof.

(a) Let $n \leq r$.

$$\begin{aligned}
&\{0, \overline{1 \ldots n-1}, n\}_b \\
&= \{0 \ldots r, \overline{r+1}\}_a \{\overline{n-1}, n \ldots n+r\}_a \bigcap_{i=r+2}^{n-2-r} \{\Omega - \{i \ldots i+r\}_a\} \\
&= \{0 \ldots n-1 \ldots r, \overline{r+1}\}_a \{\overline{n-1}, n \ldots n+r\}_a \bigcap_{i=r+2}^{n-2-r} \{\Omega - \{i \ldots i+r\}_a\} \\
&= \phi
\end{aligned}$$

The result follows.

(b) In Theorem 8 the expression has factors of the form $\Omega - \{i \ldots i+r\}_a$ for values of i for which

$$r+2 \leq n-2-r$$

$$\therefore 2r+4 \leq n$$

Thus for $n < 2r+4$ there are no such factors. The result follows. □

Corollary 11. *Let* $b = \bigwedge_r(a)$.

$$(\overline{0}, 1 \ldots n-1, \overline{n})_b = \begin{cases} 0, & n \leq r \\ (\overline{0 \ldots r}, r+1, n-1, \overline{n \ldots n+r})_a, & r < n < 2r+4 \end{cases}$$

4.5 Deriving the LULU distributions

4.5.1 The output distributions of dual operators

When two operators Q and R are duals of each other, i.e. when

$$Q(x) = -R(-x)$$

then their output distributions are also related. We have

$$\begin{aligned} F_{Q(x)}(t) &= P(Q(x) \leq t) \\ &= P(-R(-x) \leq t) \\ &= P(R(-x) \geq -t) \\ &= 1 - F_{R(-x)}(-t) \end{aligned}$$

Remembering that $F_{-x}(-t) = 1 - F_x(t)$, it translates, in terms of $F_x(t)$, to

$$F_{Q(x)}(F_x(t)) = 1 - F_{R(x)}(1 - F_x(t))$$

or simply $F_Q(F_x) = 1 - F_R(1 - F_x)$.

4.5.2 The output distributions of $\bigvee_n$ and $\bigwedge_n$

Let $y = \bigvee_n(x)$.

$$\begin{aligned} F_{\bigvee_n} &= (0)_y \\ &= (0 \ldots n)_x \\ &= F_x^{n+1} \end{aligned}$$

$$\begin{aligned} F_{\bigwedge_n} &= 1 - F_{\bigvee_n}(1 - F_x) \\ &= 1 - (1 - F_x)^{n+1} \end{aligned}$$

4.5.3 The output distributions of L_n and U_n

Theorem 12.

$$\begin{aligned} &(a) \quad F_{U_n} = (n+1)F_x^{n+1} - nF_x^{n+2} \\ &(b) \quad F_{L_n} = 1 - (n+1)(1 - F_x)^{n+1} + n(1 - F_x)^{n+2} \end{aligned}$$

Proof.

(a) U_n is given by $U_n = \bigwedge_n \bigvee_n$.

$$\begin{aligned}\text{Let } y &= \bigvee_n (x) \\ z &= \bigwedge_n (y)\end{aligned}$$

$$\begin{aligned}F_{U_n} &= (0)_z \\ &= 1 - (\overline{0})_z \\ &= 1 - (\overline{0 \ldots n})_y \\ &= [n+1](0)_y - \sum_{i=0}^{n-1} [n-i](0, \overline{1 \ldots i}, i+1)_y \\ &= [n+1](0 \ldots n)_x - n(0 \ldots n+1)_x\end{aligned}$$

$$F_{U_n} = (n+1)F_x^{n+1} - nF_x^{n+2}$$

(b)

$$\begin{aligned}F_{L_n} &= 1 - F_{U_n}(1 - F_x) \\ &= 1 - (n+1)(1-F_x)^{n+1} - n(1-F_x)^{n+2}\end{aligned}$$ ∎

4.5.4 The output distributions of L_nU_n and U_nL_n

Theorem 13.

$$\begin{aligned}(a) \quad F_{L_nU_n} &= F_x^{n+1} + nF_x^{n+1}(1-F_x) + F_x^{2n+2}(1-F_x) \\ &\quad + \frac{1}{2}(n-1)(n+2)F_x^{2n+2}(1-F_x)^2 \\ (b) \quad F_{U_nL_n} &= 1 - (1-F_x)^{n+1} - nF_x(1-F_x)^{n+1} - F_x(1-F_x)^{2n+2} \\ &\quad - \frac{1}{2}(n-1)(n+2)F_x^2(1-F_x)^{2n+2}\end{aligned}$$

Proof. (a) $\quad L_nU_n = \bigvee_n \bigwedge_n \bigwedge_n \bigvee_n = \bigvee_n \bigwedge_{2n} \bigvee_n$

$$\begin{aligned}\text{Let } a &= \bigvee_n (x) \\ b &= \bigwedge_{2n} (a) \\ c &= \bigvee_n (b)\end{aligned}$$

$$
\begin{aligned}
&F_{L_nU_n} \\
&= (0)_c \\
&= (0...n)_b \\
&= 1 - [n+1](\overline{0})_b + \sum_{i=0}^{n-1}[n-i](\overline{0}, 1...i, \overline{i+1})_b \\
&= 1 - [n+1](\overline{0...2n})_a + n(\overline{0...2n+1})_a \\
&= 1 - [n+1]\left[1 - [2n+1](0)_a + \sum_{i=0}^{2n-1}[2n-i](0, \overline{1...i}, i+1)_a\right] \\
&\quad + n\left[1 - [2n+2](0)_a + \sum_{i=0}^{2n}[2n+1-i](0, \overline{1...i}, i+1)_a\right] \\
&= [n+1](0)_a + \sum_{i=0}^{2n}[i-n](0, \overline{1...i}, i+1)_a \\
&= [n+1](0)_a - n(01)_a + \sum_{i=1}^{n}[i-n](0, \overline{1...i}, i+1)_a + (0, \overline{1...n+1}, n+2)_a \\
&\quad + \sum_{i=n+2}^{2n}[i-n](0, \overline{1...i}, i+1)_a \\
&= [n+1](0...n)_x - n(0...n+1)_x + 0 + (0...n, \overline{n+1}, n+2...2n+2)_x \\
&\quad + \sum_{i=n+2}^{2n}[i-n](0...n, \overline{n+1, i}, i+1...i+1+n)_x \\
&= (n+1)F_x^{n+1} - nF_x^{n+2} + F_x^{2n+2}(1-F_x) + \sum_{i=n+2}^{2n}(i-n)F_x^{2n+2}(1-F_x)^2 \\
&= F_x^{n+1} + nF_x^{n+1}(1-F_x) + F_x^{2n+2}(1-F_x) \\
&\quad + \frac{1}{2}(n-1)(n+2)F_x^{2n+2}(1-F_x)^2
\end{aligned}
$$

(b)

$$
\begin{aligned}
&F_{U_nL_n} \\
&= 1 - F_{L_nU_n}(1-F_x) \\
&= 1 - (1-F_x)^{n+1} - nF_x(1-F_x)^{n+1} - F_x(1-F_x)^{2n+2} \\
&\quad - \frac{1}{2}(n-1)(n+2)F_x^2(1-F_x)^{2n+2} \qquad \square
\end{aligned}
$$

4.5.5 The output distributions of C_n (recursive form)

Theorem 14.

$$(a) \quad F_{C_n} = F_{C_{n-1}} + n(G_{2n} - G_{2n-1})$$
$$(b) \quad F_{C_n} = F_x + \sum_{k=1}^{n} k(G_{2k} - G_{2k-1})$$

where G_n is defined below.

Proof.
(a)

$$\begin{aligned} C_n &= L_n U_n L_{n-1} U_{n-1} L_{n-2} U_{n-2} \dots L_1 U_1 \\ &= L_n U_n L_{n-1} U_{n-1} C_{n-2} \\ &= \bigvee\nolimits_n \bigwedge\nolimits_n \bigwedge\nolimits_n \bigvee\nolimits_n \bigvee\nolimits_{n-1} \bigwedge\nolimits_{n-1} \bigwedge\nolimits_{n-1} \bigvee\nolimits_{n-1} C_{n-2} \\ &= \bigvee\nolimits_n \bigwedge\nolimits_{2n} \bigvee\nolimits_{2n-1} \bigwedge\nolimits_{2n-2} \bigvee\nolimits_{n-1} C_{n-2} \end{aligned}$$

Similarly

$$C_{n-1} = \bigvee\nolimits_{n-1} \bigwedge\nolimits_{2n-2} \bigvee\nolimits_{n-1} C_{n-2}$$

Let

$$\begin{aligned} a &= \bigvee\nolimits_{n-1} C_{n-2}(x) \\ b &= \bigwedge\nolimits_{2n-2}(a) \\ c' &= \bigvee\nolimits_{n-1}(b) \\ c &= \bigvee\nolimits_{2n-1}(b) \\ d &= \bigwedge\nolimits_{2n}(c) \\ e &= \bigvee\nolimits_{n}(d) \end{aligned}$$

Define G_n by

$$\begin{cases} G_{2n} = (0 \dots 2n-1, \overline{2n}, 2n+1 \dots 4n)_b \\ G_{2n-1} = (\overline{0 \dots 2n-2}, 2n-1, \overline{2n \dots 4n-2})_a \end{cases}$$

$$\begin{aligned} F_{C_{n-1}} &= (0)_{c'} \\ &= (0 \dots n-1)_b \\ &= 1 - n(\overline{0})_b + \sum_{i=0}^{n-2} [n-1-i](\overline{0}, 1 \dots i, \overline{i+1})_b \\ &= 1 - n(\overline{0 \dots 2n-2})_a + [n-1](\overline{0 \dots 2n-1})_a \end{aligned}$$

$$\begin{aligned}
F_{C_n} &= (0)_e \\
&= (0\ldots n)_d \\
&= 1-[n+1](\overline{0})_d + \sum_{i=0}^{n-1}[n-i](\overline{0},1\ldots i,\overline{i+1})_d \\
&= 1-[n+1](\overline{0\ldots 2n})_c + n(\overline{0\ldots 2n+1})_c \\
&= 1-[n+1]\left[1-[2n+1](0)_c + \sum_{i=0}^{2n-1}[2n-i](0,\overline{1\ldots i},i+1)_c\right] \\
&\quad + n\left[1-[2n+2](0)_c + \sum_{i=0}^{2n}[2n+1-i](0,\overline{1\ldots i},i+1)_c\right] \\
&= [n+1](0)_c - \sum_{i=0}^{2n}[n-i](0,\overline{1\ldots i},i+1)_c \\
&= [n+1](0\ldots 2n-1)_b - n(0\ldots 2n)_b \\
&\quad + n(0\ldots 2n-1,\overline{2n},2n+1\ldots 4n)_b \\
&= [n+1](0\ldots 2n-1)_b - n(0\ldots 2n)_b + nG_{2n}
\end{aligned}$$

$$\begin{aligned}
&F_{C_n} - nG_{2n} \\
&= [n+1](0\ldots 2n-1)_b - n(0\ldots 2n)_b \\
&= [n+1]\left[1-2n(\overline{0})_b + \sum_{i=0}^{2n-2}[2n-1-i](\overline{0},1\ldots i,\overline{i+1})_b\right] \\
&\quad - n\left[1-[2n+1](\overline{0})_b + \sum_{i=0}^{2n-1}[2n-i](\overline{0},1\ldots i,\overline{i+1})_b\right] \\
&= 1-n(\overline{0})_b + \sum_{i=0}^{2n-1}[n-1-i](\overline{0},1\ldots i,\overline{i+1})_b \\
&= 1-n(\overline{0\ldots 2n-2})_a + [n-1](\overline{0\ldots 2n-1})_a \\
&\quad - n(\overline{0\ldots 2n-2},2n-1,\overline{2n\ldots 4n-2})_a \\
&= F_{C_{n-1}} - nG_{2n-1}
\end{aligned}$$

(b)

$$\begin{aligned}
F_{C_n} &= F_{C_{n-1}} + n(G_{2n}-G_{2n-1}) \\
&= F_{C_{n-2}} + (n-1)(G_{2n-2}-G_{2n-3}) + n(G_{2n}-G_{2n-1}) \\
&\;\;\vdots \\
&= F_{C_0} + \sum_{k=1}^{n} k(G_{2k}-G_{2k-1})
\end{aligned}$$

$$= F_x + \sum_{k=1}^{n} k(G_{2k} - G_{2k-1}) \qquad \square$$

4.5.6 Deriving G_n

Next we derive G_n for $n = 1 \dots 12$. In what follows we shall write $F = F_x$. We shall write

$$\Big[\dots\Big]\Big[\dots\Big] \text{ for } \Big\{\dots\Big\} \cap \Big\{\dots\Big\}.$$

You will notice that all the expressions have pairs of factors which are symmetrical relative to the middle factor. When two factors are symmetrical we will denote the second one with $\Big[\dots\Big|$. When expressing the probability in terms of F we shall also write

$$\begin{pmatrix} m_1 & m_2 & \dots & m_k \\ n_1 & n_2 & & n_k \end{pmatrix} \text{ for } F^{m_1}(1-F)^{n_1} + F^{m_2}(1-F)^{n_2} + \dots + F^{m_k}(1-F)^{n_k}$$

In deriving G_n, we shall repeatedly use Theorems 1, 6, 8 and 10 to alternately invert and expand the terms in its expression.

Note that

$$\bigvee\nolimits_0 = \bigwedge\nolimits_0 = 1$$

$$\therefore L_0U_0 = \bigvee\nolimits_0 \bigwedge\nolimits_0 \bigwedge\nolimits_0 \bigvee\nolimits_0 = 1$$

$$\begin{aligned} C_3(x) &= L_3U_3L_2U_2L_1U_1L_0U_0(x) \\ &= \bigvee\nolimits_3 \bigwedge\nolimits_3 \bigwedge\nolimits_3 \bigvee\nolimits_3 \bigvee\nolimits_2 \bigwedge\nolimits_2 \bigwedge\nolimits_2 \bigvee\nolimits_2 \bigvee\nolimits_1 \bigwedge\nolimits_1 \bigwedge\nolimits_1 \bigvee\nolimits_1 \bigvee\nolimits_0 \bigwedge\nolimits_0 \bigwedge\nolimits_0 \bigvee\nolimits_0(x) \\ &= \bigvee\nolimits_3 \bigwedge\nolimits_6 \bigvee\nolimits_5 \bigwedge\nolimits_4 \bigvee\nolimits_3 \bigwedge\nolimits_2 \bigvee\nolimits_1 \bigwedge\nolimits_0 \bigvee\nolimits_0(x) \end{aligned}$$

$$\begin{aligned} \text{Let } a &= \bigvee\nolimits_0(x) = x \\ b &= \bigwedge\nolimits_0(a) = x \\ c &= \bigvee\nolimits_1(b) = \bigvee\nolimits_1(x) \\ d &= \bigwedge\nolimits_2(c) \\ e &= \bigvee\nolimits_3(d) \\ f &= \bigwedge\nolimits_4(e) \end{aligned}$$

$$
\begin{aligned}
&\{\bar{0}1\bar{2}\}_a \\
&= \{\bar{0}1\bar{2}\}_x \\
G_1 &= F(1-F)^2 \\
&= \begin{pmatrix} 1 \\ 2 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&\{01\bar{2}34\}_b \\
&= \{01\bar{2}34\}_x \\
G_2 &= F^4(1-F) \\
&= \begin{pmatrix} 4 \\ 1 \end{pmatrix}
\end{aligned}
$$

$$
\begin{aligned}
&\{\overline{012}3\overline{456}\}_c \\
&= \Big[\Omega - \{0\}_c\Big]\Big[\Omega - \{1\}_c\Big]\{\bar{2}3\bar{4}\}_c\Big[\Omega - \{5\}_c\Big]\Big[\Omega - \{6\}_c\Big] \\
&= \Big[\Omega - \{01\}_x\Big]\Big[\Omega - \{12\}_x\Big]\{\bar{2}34\bar{5}\}_x\Big[...\Big]\Big[...\Big] \\
&= \Big[\Omega - \{01\}_x\Big]\{\bar{2}34\bar{5}\}_x\Big[...\Big] \\
G_3 &= F^2(1-F)^2\Big[1 - F^2\Big]^2 \\
&= \begin{pmatrix} 2 \\ 2 \end{pmatrix}\left[1 - \begin{pmatrix} 2 \\ 0 \end{pmatrix}\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&\{0123\bar{4}5678\}_d \\
&= \{3\bar{4}5\}_d\Big[\Omega - \{\bar{0}\}_d\Big]\Big[\Omega - \{\bar{1}\}_d\Big]\Big[\Omega - \{\bar{2}\}_d\Big]\Big[...\Big]\Big[...\Big]\Big[...\Big] \\
&= \{3\overline{456}7\}_c\Big[\Omega - \{\overline{012}\}_c\Big]\Big[\Omega - \{\overline{123}\}_c\Big]\Big[\Omega - \{\overline{234}\}_c\Big]\Big[...\Big]\Big[...\Big]\Big[...\Big] \\
&= \{3\overline{456}7\}_c\Big[\Omega - \{\overline{012}\}_c\Big]\Big[...\Big] \\
&= \{3\overline{456}7\}_c\Big[\{2\}_c + \{1\bar{2}\}_c + \{0\overline{12}\}_c\Big]\Big[...\Big] \\
&= \{34\overline{56}78\}_x\Big[\{2\}_x + \{01\bar{2}\}_x\Big]\Big[...\Big] \\
G_4 &= F^4(1-F)^2\Big[F + F^2(1-F)\Big]^2 \\
&= \begin{pmatrix} 4 \\ 2 \end{pmatrix}\left[\begin{pmatrix} 1\,2 \\ 0\,1 \end{pmatrix}\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&\{\overline{01234}5\overline{6789\,10}\}_e \\
&= \{\overline{4}5678\overline{9}\}_d\Big[\mathit{\Omega} - \{0123\}_d\Big]\Big[...\Big] \\
&= \{\overline{4}5678\overline{9}\}_d\Big[\{\overline{3}\}_d + \{\overline{2}3\}_d + \{\overline{1}23\}_d + \{\overline{0}123\}_d\Big]\Big[...\Big] \\
&= \{\overline{456}78\overline{9\,10\,11}\}_c\Big[\{\overline{3}\}_c + \{\overline{012}3\}_c\Big]\Big[...\Big] \\
&= \{\overline{456}78\overline{9\,10\,11}\}_c\Big[\mathit{\Omega} - \{23\}_c - \{1\overline{2}3\}_c - \{0\overline{12}3\}_c\Big]\Big[...\Big] \\
&= \{\overline{6}789\overline{10}\}_x\Big[\mathit{\Omega} - \{45\}_x\Big]\Big[\mathit{\Omega} - \{234\}_x - \{01\overline{2}34\}_x\Big]\Big[...\Big]\Big[...\Big] \\
&= \{\overline{6}789\overline{10}\}_x\Big[\mathit{\Omega} - \{45\}_x - \{234\overline{5}\}_x - \{01\overline{2}34\overline{5}\}_x\Big]\Big[...\Big]
\end{aligned}
$$

$$
\begin{aligned}
G_5 &= F^3(1-F)^2\Big[1 - F^2 - F^3(1-F) - F^4(1-F)^2\Big]^2 \\
&= \binom{3}{2}\left[1 - \begin{pmatrix} 2\;3\;4 \\ 0\;1\;2 \end{pmatrix}\right]^2
\end{aligned}
$$

$$
\begin{aligned}
&\{012345\overline{6}789\,10\,11\,12\}_f \\
&= \{5\overline{6789\,10}11\}_e\Big[\mathit{\Omega} - \{\overline{01234}\}_e\Big]\Big[...\Big] \\
&= \{5\overline{6789\,10}11\}_e\Big[\{4\}_e + \{3\overline{4}\}_e + \{2\overline{34}\}_e + \{1\overline{234}\}_e + \{0\overline{1234}\}_e\Big]\Big[...\Big] \\
&= \{5678\overline{9\,10}11\,12\,13\,14\}_d\Big[\{4\}_d + \{0123\overline{4}\}_d\Big]\Big[...\Big] \\
&= \{5...8\,\overline{9\,10}\,11...14\}_d\Big[\mathit{\Omega} - \{\overline{34}\}_d - \{\overline{2}3\overline{4}\}_d - \{\overline{1}23\overline{4}\}_d - \{\overline{0}123\overline{4}\}_d\Big]\Big[...\Big] \\
&= \{8\overline{9\,10\,11\,12}13\}_c\Big[\mathit{\Omega} - \{\overline{567}\}_c\Big]\Big[\mathit{\Omega} - \{\overline{3456}\}_c - \{\overline{012}3\overline{456}\}_c\Big]\Big[...\Big]\Big[...\Big] \\
&= \{8\overline{9\,10\,11\,12}13\}_c\Big[\mathit{\Omega} - \{\overline{567}\}_c - \{\overline{3456}7\}_c - \{\overline{012}3\overline{456}7\}_c\Big]\Big[...\Big] \\
&= \{8\overline{9\,10\,11\,12}13\}_c\Big[\{7\}_c + \{6\overline{7}\}_c + \{5\overline{67}\}_c - \{7\}_c + \{67\}_c + \{5\overline{6}7\}_c \\
&\quad + \{4\overline{56}7\}_c + \{3\overline{456}7\}_c - \{3\overline{456}7\}_c + \{23\overline{456}7\}_c + \{1\overline{2}3\overline{456}7\}_c \\
&\quad + \{0\overline{12}3\overline{456}7\}_c\Big]\Big[...\Big] \\
&= \{8\overline{9\,10\,11\,12}13\}_c\Big[\{6\overline{7}\}_c + \{5\overline{67}\}_c + \{67\}_c + \{5\overline{6}7\}_c + \{4\overline{56}7\}_c \\
&\quad + \{23\overline{456}7\}_c + \{1\overline{2}3\overline{456}7\}_c + \{0\overline{12}3\overline{456}7\}_c\Big]\Big[...\Big] \\
&= \{89\overline{1012}13\,14\}_x\Big[\{56\overline{7}\}_x + \{67\}_x + \{45\overline{6}7\}_x + \{234\overline{56}7\}_x \\
&\quad + \{01\overline{2}34\overline{56}7\}_x\Big]\Big[...\Big]
\end{aligned}
$$

$$
\begin{aligned}
G_6 &= F^4(1-F)^2\Big[F^2 + F^2(1-F) + F^3(1-F) + F^4(1-F)^2 \\
&\quad + F^5(1-F)^3\Big]^2 \\
&= \binom{4}{2}\left[\begin{pmatrix} 2\,2\,3\,4\,5 \\ 0\,1\,1\,2\,3 \end{pmatrix}\right]^2
\end{aligned}
$$

Similarly

$$
\begin{aligned}
G_7 &= F^4(1-F)^2\Big[1 - F^2 - F^3(1-F) - F^4(1-F)^2 - F^4(1-F)^2 \\
&\quad - F^4(1-F) - F^5(1-F)^2 - F^6(1-F)^3 - F^7(1-F)^4\Big]^2 \\
&= \binom{4}{2}\left[1 - \begin{pmatrix} 2\,3\,4\,4\,4\,5\,6\,7 \\ 0\,1\,2\,2\,1\,2\,3\,4 \end{pmatrix}\right]^2
\end{aligned}
$$

$$
\begin{aligned}
G_8 &= F^4(1-F)^2(1-F^2)\Big[F^3 + F^2(1-F) + F^3(1-F) + F^4(1-F)^2 \\
&\quad + F^5(1-F)^3 + F^5(1-F)^2 + F^6(1-F)^3 + F^4(1-F) \\
&\quad + F^6(1-F)^3 + F^6(1-F)^2 + F^7(1-F)^3 + F^8(1-F)^4 \\
&\quad + F^9(1-F)^5\Big]^2 \\
&= \binom{4}{2}\left(1-F^2\right)\left[\begin{pmatrix} 3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9 \\ 0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5 \end{pmatrix}\right]^2
\end{aligned}
$$

$$
\begin{aligned}
G_9 &= F^4(1-F)^2\Big[1 - F^2 - F^3(1-F) - F^4(1-F)^2 - F^4(1-F)^2 \\
&\quad - F^4(1-F) - F^5(1-F)^2 - F^6(1-F)^3 - F^7(1-F)^4 \\
&\quad - (1-F^2)F^2(1-F)\Big[F^3 + F^2(1-F) + F^3(1-F) + F^4(1-F) \\
&\quad + F^4(1-F)^2 + F^4(1-F)^2 + F^5(1-F)^3 + F^6(1-F)^3 \\
&\quad + F^6(1-F)^3 + F^6(1-F)^2 + F^7(1-F)^3 + F^8(1-F)^4 \\
&\quad + F^9(1-F)^5\Big]\Big]^2 \\
&= \binom{4}{2}\left[1 - \begin{pmatrix} 2\,3\,4\,4\,4\,5\,6\,7 \\ 0\,1\,2\,2\,1\,2\,3\,4 \end{pmatrix} - \left(1-F^2\right)\binom{2}{1}\right. \\
&\quad \left.\times \begin{pmatrix} 3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9 \\ 0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5 \end{pmatrix}\right]^2
\end{aligned}
$$

$$
G_{10} = \binom{4}{2}\left(1 - 2F^2 + F^3\right)\left[\begin{pmatrix} 3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9 \\ 0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5 \end{pmatrix} - \binom{2}{1}\Big[1 -\right.
$$

$$\binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4} - \left(1 - F^2\right)\binom{2}{1}\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5}\Bigg]\Bigg]^2$$

$$\begin{aligned}G_{11} = &\binom{4}{2}\left(2F - F^2\right)\Bigg[1 - \binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4} - \left(1 - F^2\right)\binom{2}{1}\\ &\times\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} - \left(1 - 2F^2 + F^3\right)\binom{2}{1}\\ &\times\Bigg[\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} + \binom{2}{1}\binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4}\\ &- \binom{2}{1} + \left(1 - F^2\right)\binom{4}{2}\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5}\Bigg]\Bigg]^2\end{aligned}$$

$$\begin{aligned}G_{12} = &\binom{4}{2}\left(1 - 2F^2 + F^4\right)\Bigg[\Bigg(\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} - \binom{2}{1} + \binom{2}{1}\\ &\times\binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4} + \binom{4}{2}\left(1 - F^2\right)\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5}\\ &+ \left(2F - F^2\right)\Bigg[-\binom{2}{0} + \binom{3}{0} + \binom{2}{1}\binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4} + \binom{4}{2}\\ &\times\left(1 - F^2\right)\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} + \binom{4}{2}\left(1 - 2F^2 + F^3\right)\\ &\times\Bigg[\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} - \binom{2}{1} + \binom{2}{1}\binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4}\\ &+ \binom{4}{2}\left(1 - F^2\right)\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5}\Bigg]\Bigg]\Bigg]\Bigg]^2\end{aligned}$$

The first few distributions of C_n are now calculated.

$$\begin{aligned}F_{C_0} &= (0)_x\\ &= F\end{aligned}$$

$$\begin{aligned}F_{C_1} &= F_{C_0} + (G_2 - G_1)\\ &= 2\,F^2 - F^3 + F^4 - F^5\end{aligned}$$

$$\begin{aligned}F_{C_2} &= F_{C_1} + 2(G_4 - G_3)\\ &= 3\,F^3 + 3\,F^4 - 9\,F^5 + 4\,F^6 + 4\,F^7 - 10\,F^8 + 4\,F^9 + 8\,F^{10}\end{aligned}$$

$$- 8\,F^{11} + 2\,F^{12}$$

$$\begin{aligned} F_{C_3} &= F_{C_2} + 3(G_6 - G_5) \\ &= 9\,F^4 - 6\,F^5 - 2\,F^6 - 5\,F^7 - 4\,F^8 + 16\,F^9 + 8\,F^{10} - 35\,F^{11} \\ &\quad - 4\,F^{12} + 78\,F^{13} - 75\,F^{14} + 27\,F^{15} - 21\,F^{16} - 9\,F^{17} + 93\,F^{18} \\ &\quad - 126\,F^{19} + 78\,F^{20} - 24\,F^{21} + 3\,F^{22} \end{aligned}$$

$$\begin{aligned} F_{C_4} &= F_{C_3} + 4(G_8 - G_7) \\ &= 5\,F^4 + 2\,F^5 + 2\,F^6 - 13\,F^7 + 4\,F^8 - 40\,F^9 + 88\,F^{10} - 59\,F^{11} \\ &\quad - 32\,F^{12} + 174\,F^{13} - 307\,F^{14} + 299\,F^{15} - 145\,F^{16} - 73\,F^{17} \\ &\quad + 437\,F^{18} - 870\,F^{19} + 930\,F^{20} - 536\,F^{21} - 85\,F^{22} + 992\,F^{23} \\ &\quad - 1916\,F^{24} + 2104\,F^{25} - 1292\,F^{26} - 112\,F^{27} + 1424\,F^{28} \\ &\quad - 1952\,F^{29} + 1552\,F^{30} - 792\,F^{31} + 256\,F^{32} - 48\,F^{33} + 4\,F^{34} \end{aligned}$$

The output distribution of F_n can be calculated as follows:

$$F_{F_n} = 1 - F_{C_n}(1 - F)$$

4.6 An expression for G_n

In this section we look at the results of the previous section for G_n. Lead by the form of the expression for $G_n(x)$, define

$$H_n(F) = \sum_i F^{a_{n,i}}(1 - F)^{b_{n,i}}$$

so that

$$G_n(x) = \begin{cases} F^{\frac{n+1}{2}}(1 - F)^2\Big[1 - H_n(F)\Big]^2, & n = 3, 5, 7 \\ F^4(1 - F)^2 H_n^2(F), & n = 4, 6 \end{cases}$$

From the results for $G_n(F)$ we can get the values of $a_{n,i}$ en $b_{n,i}$. It is shown in the next table. (In each row $a_{n,i}$ is listed above $b_{n,i}$). The number of

n	$a_{n,i}$ $b_{n,i}$
2	0 0
3	2 0
4	1 2 0 1
5	2 3 4 0 1 2
6	2 2 3 4 5 0 1 1 2 3
7	2 3 4 4 4 5 6 7 0 1 2 1 2 2 3 4

Table 4.2: The exponents in the expansion of $H_n(F)$

terms in the expansion of $H_n(F)$ above grows as the Fibonacci numbers defined by

$$u_0 = u_1 = 1 \text{ and } u_n = u_{n-1} + u_{n-2}.$$

Thus we aim to write $H_n(F)$ in the form

$$H_n = f(H_{n-2}, H_{n-1}).$$

Next we derive such a recursive expression. Define $\delta_{n\ even}$ as follows.

$$\delta_{n\ even} = \begin{cases} 1, & n\ even \\ 0, & n\ uneven \end{cases}$$

$$\begin{aligned} H_n(F) &= f\Big(H_{n-2}(F), H_{n-1}(F)\Big) \\ &= H_{n-2}(F) + F^{a_{n-2,1}}(1-F)H_{n-1}(F) \\ &\quad +(F.F^{a_{n-2,1}} - F^{a_{n-2,1}})\delta_{n\ even} \\ &= H_{n-2}(F) + F^{a_{n-2,1}}(1-F)H_{n-1}(F) + F^{a_{n-2,1}}(F-1)\delta_{n\ even} \\ &= H_{n-2}(F) + F^{a_{n-2,1}}(1-F)\Big[H_{n-1}(F) - \delta_{n\ even}\Big] \end{aligned}$$

Now distinguish between even and uneven n.

$$\begin{aligned} H_{2n}(F) &= H_{2n-2}(F) + F^{a_{2n-2,1}}(1-F)\Big[H_{2n-1}(F) - 1\Big] \\ H_{2n+1}(F) &= H_{2n-1}(F) + F^{a_{2n-1,1}}(1-F)H_{2n}(F) \end{aligned}$$

Notice that

$$\begin{aligned} a_{2n-1,1} &= 2 \\ a_{2n,1} &= n-1 \end{aligned}$$

Substituting for these values gives us the following set of equations to calculate $H_n(F)$ and $G_n(F)$ for $n = 1 \ldots 7$.

$$\begin{aligned} H_2(F) &= 1 \\ H_3(F) &= x^2 \\ H_{2n}(F) &= H_{2n-2}(F) + F^{n-2}(1-F)\Big[H_{2n-1}(F) - 1\Big] \\ H_{2n+1}(F) &= H_{2n-1}(F) + F^2(1-F)H_{2n}(F) \end{aligned}$$

$$\begin{aligned} G_1(F) &= x(1-F)^2 \\ G_2(F) &= x^4(1-F) \\ G_{2n-1}(F) &= F^n(1-F)^2\Big[1 - H_{2n-1}(F)\Big]^2, \; n > 2. \\ G_{2n}(F) &= F^4(1-F)^2 H_{2n}^2(F), \; n > 2. \end{aligned}$$

Now we will rewrite the expressions for $G_n(F)$ for $n = 8 \ldots 12$ in a similar form.

$$\begin{aligned} &G_8(F) \\ &= \binom{4}{2}\left(1-F^2\right)\left[\begin{pmatrix} 3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9 \\ 0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5 \end{pmatrix}\right]^2 \\ &= \binom{4}{2}\left(1-F^2\right)\left[H_6(F) + \binom{2}{1}\left[H_7(F) - 1\right]\right]^2 \\ &\equiv \binom{4}{2}\left(1-F^2\right)\Big[H_8(F)\Big]^2 \end{aligned}$$

$$\begin{aligned} &G_9(F) \\ &= \binom{4}{2}\left[1 - \begin{pmatrix} 2\,3\,4\,4\,4\,5\,6\,7 \\ 0\,1\,2\,2\,1\,2\,3\,4 \end{pmatrix} - \left(1-F^2\right)\binom{2}{1} \right. \\ &\quad \left. \times \begin{pmatrix} 3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9 \\ 0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5 \end{pmatrix}\right]^2 \\ &= \binom{4}{2}\left[1 - H_7(F) - \left(1-F^2\right)\binom{2}{1} H_8(F)\right]^2 \end{aligned}$$

$$\equiv \binom{4}{2}\Big[1 - H_9(F)\Big]^2$$

$G_{10}(F)$

$$\begin{aligned}
&= \binom{4}{2}\Big(1-2F^2+F^3\Big)\Bigg[\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} - \binom{2}{1}\Bigg[1 - \\
&\quad \binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4} - \Big(1-F^2\Big)\binom{2}{1}\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5}\Bigg]\Bigg]^2 \\
&= \binom{4}{2}\Big(1-2F^2+F^3\Big) \\
&\quad \times\Bigg[H_8(F) - \binom{2}{1}\Bigg[1 - H_7(F) - \Big(1-F^2\Big)\binom{2}{1}H_8(F)\Bigg]\Bigg]^2 \\
&= \binom{4}{2}\Big(1-2F^2+F^3\Big)\Bigg[H_8(F) + \binom{2}{1}\Big[H_9(F) - 1\Big]\Bigg]^2 \\
&\equiv \binom{4}{2}\Big(1-2F^2+F^3\Big)\Big[H_{10}(F)\Big]^2
\end{aligned}$$

$G_{11}(F)$

$$\begin{aligned}
&= \binom{4}{2}\Big(2F-F^2\Big)\Bigg[1 - \binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4} - \Big(1-F^2\Big)\binom{2}{1} \\
&\quad \times\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} - \Big(1-2F^2+F^3\Big)\binom{2}{1} \\
&\quad \times\Bigg[\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5} + \binom{2}{1}\binom{2\,3\,4\,4\,4\,5\,6\,7}{0\,1\,2\,2\,1\,2\,3\,4} \\
&\quad -\binom{2}{1} + \Big(1-F^2\Big)\binom{4}{2}\binom{3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9}{0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5}\Bigg]\Bigg]^2 \\
&= \binom{4}{2}\Big(2F-F^2\Big)\Bigg[1 - H_7(F) - \Big(1-F^2\Big)\binom{2}{1}H_8(F) \\
&\quad -\Big(1-2F^2+F^3\Big)\binom{2}{1}\Bigg[H_8(F) + \binom{2}{1}H_7(F) - \binom{2}{1} \\
&\quad +\Big(1-F^2\Big)\binom{4}{2}H_8(F)\Bigg]\Bigg]^2 \\
&= \binom{4}{2}\Big(2F-F^2\Big)\Bigg[1 - H_9(F) - \Big(1-2F^2+F^3\Big)\binom{2}{1} \\
&\quad \times\Bigg[H_8(F) + \binom{2}{1}\Big[H_9(F) - 1\Big]\Bigg]\Bigg]^2 \\
&= \binom{4}{2}\Big(2F-F^2\Big)\Bigg[1 - H_9(F) - \Big(1-2F^2+F^3\Big)\binom{2}{1}H_{10}(F)\Bigg]^2
\end{aligned}$$

$$\equiv \binom{4}{2}\left(2F-F^2\right)\left[1-H_{11}(F)\right]^2$$

$G_{12}(F)$

$$\begin{aligned}
&= \binom{4}{2}\left(1-2F^2+F^4\right)\left[\begin{pmatrix}3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9\\0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5\end{pmatrix}-\binom{2}{1}+\binom{2}{1}\right.\\
&\quad\times\begin{pmatrix}2\,3\,4\,4\,4\,5\,6\,7\\0\,1\,2\,2\,1\,2\,3\,4\end{pmatrix}+\binom{4}{2}\left(1-F^2\right)\begin{pmatrix}3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9\\0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5\end{pmatrix}\\
&\quad+\left(2F-F^2\right)\left[-\binom{2}{0}+\binom{3}{0}+\binom{2}{1}\begin{pmatrix}2\,3\,4\,4\,4\,5\,6\,7\\0\,1\,2\,2\,1\,2\,3\,4\end{pmatrix}+\binom{4}{2}\right.\\
&\quad\times\left(1-F^2\right)\begin{pmatrix}3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9\\0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5\end{pmatrix}+\binom{4}{2}\left(1-2F^2+F^3\right)\\
&\quad\times\left[\left(\begin{pmatrix}3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9\\0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5\end{pmatrix}-\binom{2}{1}+\binom{2}{1}\begin{pmatrix}2\,3\,4\,4\,4\,5\,6\,7\\0\,1\,2\,2\,1\,2\,3\,4\end{pmatrix}\right.\right.\\
&\quad\left.\left.\left.\left.+\binom{4}{2}\left(1-F^2\right)\begin{pmatrix}3\,2\,3\,4\,5\,5\,6\,4\,6\,6\,7\,8\,9\\0\,1\,1\,2\,3\,2\,3\,1\,3\,2\,3\,4\,5\end{pmatrix}\right)\right]\right]\right]^2\\
&= \binom{4}{2}\left(1-2F^2+F^4\right)\left[H_8(F)-\binom{2}{1}+\binom{2}{1}H_7(F)\right.\\
&\quad+\binom{4}{2}\left(1-F^2\right)H_8(F)+\left(2F-F^2\right)\left[-\binom{2}{0}+\binom{3}{0}+\binom{2}{1}H_7(F)\right.\\
&\quad+\binom{4}{2}\left(1-F^2\right)H_8(F)+\binom{4}{2}\left(1-2F^2+F^3\right)\\
&\quad\left.\left.\times\left[H_8(F)-\binom{2}{1}+\binom{2}{1}H_7(F)+\binom{4}{2}\left(1-F^2\right)H_8(F)\right]\right]\right]^2\\
&= \binom{4}{2}\left(1-2F^2+F^4\right)\left[H_8(F)+\binom{2}{1}\left[H_9(F)-1\right]\right.\\
&\quad+\left(2F-F^2\right)\left[-\binom{2}{1}+\binom{2}{1}H_7(F)+\binom{4}{2}\left(1-F^2\right)H_8(F)\right.\\
&\quad\left.\left.+\binom{4}{2}\left(1-2F^2+F^3\right)\left[H_8(F)+\binom{2}{1}\left[H_9(F)-1\right]\right]\right]\right]^2\\
&= \binom{4}{2}\left(1-2F^2+F^4\right)\left[H_{10}(F)+\left(2F-F^2\right)\left[\binom{2}{1}\left[H_9(F)-1\right]\right.\right.\\
&\quad\left.\left.+\binom{4}{2}\left(1-2F^2+F^3\right)H_{10}(F)\right]\right]^2\\
&= \binom{4}{2}\left(1-2F^2+F^4\right)\left[H_{10}(F)+\left(2F-F^2\right)\binom{2}{1}\left[H_{11}(F)-1\right]\right]^2\\
&\equiv \binom{4}{2}\left(1-2F^2+F^4\right)\left[H_{12}(F)\right]^2
\end{aligned}$$

Define $I_n(F)$ as follows

$$I_n(F) = \binom{2}{1}\left[H_n(F) - n \bmod 2\right]$$

Then $H_n(F)$ and $G_n(F)$ are given by

$$\begin{aligned} H_n(F) &= H_{n-2}(F) + I_{n-1}(F)J_{n-1}(F) \\ G_n(F) &= I_n^2(F)J_n(F) \end{aligned}$$

where $J_n(F)$ for $n = 1\ldots 12$ is listed in table 4.3.

n	$J_n(F)$	n	$J_n(F)$
1	F^{-3}	2	$(1-F)^{-1}$
3	F^{-2}	4	1
5	F^{-1}	6	1
7	1	8	$1-F^2$
9	1	10	$1-2F^2+F^3$
11	$2F-F^2$	12	$1-2F^2+F^4$

Table 4.3: Values of $J_n(F)$

Looking at the values for $J_n(F)$ it seems possible that

$$J_{2n+3}(F) = J_{2n}(1-F)$$

Together with the starting values

$$\begin{aligned} H_1 &= 0 \\ H_2 &= 1 \end{aligned}$$

we have a set of equations to calculate $G_n(F)$ for $n = 1\ldots 12$. We can now calculate F_{C_n} recursively for $n = 1\ldots 6$. The distributions can be calculated very fast on a computer using the above recursive equations. The calculations was done and table 4.6 shows the degree of F_{C_n}.

4.7 Programming

The next Mathematica program calculates the first 6 distributions of C_n recursively.

n	$\deg(F_{C_n})$
1	5
2	12
3	22
4	36
5	53
6	76

Table 4.4: The degrees of the first C_n distributions

```
n=12;
F={x};
G={};
H={0,1};
II={};
J={1/x^3,1/(1-x),1/x^2,1,1/x,1,1,1-x^2,1,1-2x^2+x^3,2x-x^2,
   1-2x^2+x^4};

For[k=1,k<=n,k++,
  If[k>2,
    H=Append[H,Extract[H,k-2]+Extract[II,k-1]Extract[J,k-1]];
  ];
  II=Append[II,x^2(1-x)(Extract[H,k]-Mod[k,2])];
  G=Append[G,Extract[II,k]^2Extract[J,k]];
  If[Mod[k,2]==0,
   F=Append[F,Extract[F,k/2]+k/2(Extract[G,k]-Extract[G,k-1])]
  ];
];

MatrixForm[Expand[Simplify[F]]]
```

4.8 Summary

In this section we devised a notation to simplify the algebra in order to derive the exact expressions of the LULU output distributions. We then found these expressions, including that of the output distribution of C_n for small n. The recursive formula in Theorem 14 can be used in further study of these distributions. Progress has been made to find an expression for G_n.

Chapter 5

Conclusion and Future Research Directions

We investigated the influence of impulsive noise on linear filters and found that they are ineffective in handling this type of noise. For an overview of the background of nonlinear smoothers, we looked at the popular median smoother.

Some of the LULU smoothers' attractive properties and other related concepts were presented. LULU smoothers are seperators, being both idempotent and co-idempotent, and they preserve the shape and variation of their input. We looked at the discrete pulse transform and showed that it decomposes sequences consistently. The DPT acts as linearly as is required for images.

We discussed the output distributions of the median and LULU smoothers and used their moments to compare them. We showed that competitive alternatives to C_n exist. We verified numerically that the first resolution level of the DPT can be used to estimate the standard deviation of an unknown error distribution. We looked at the asymptotic distributions of LULU smoothers as the window size goes to infinity, and saw how they are related to the order statistics.

This study was fruitful in that we were able to derive a recursive formula for the output distribution of C_n. This is an important theoretical result which aids in the study of the discrete pulse transform.

Avenues that are open for further investigation includes the following:

- There now exists a recursive formula for the output distribution of C_n, but whether it can be written in a closed form is still an interesting open problem. For theoretical purposes it would be very useful.
- K_n and other alternatives to C_n need to be researched, as we have shown that C_n, although it is a most natural composite LULU smoother, it is not necessarily an optimal one.
- A study of the resolution levels of the DPT, in particular, the output distribution of D_n, will yield important results.
- The use of the LULU smoothers in image processing is promising. In particular the DPT could be useful in the elusive "mathematics of vision", as envisaged prophetically, of Marr[19]. Understanding the handling of noise superimposed on signal is crucial and a difficult problem. Getting a grip on the transer of distributions in the presence of some type of signals will benefit from the results obtained for pure noise.

Appendix A

Background of standard distributions

A.1 The cardinal B-splines

We define the cardinal B-splines:

$$B_0(t) = \begin{cases} 1, & t \in [-\frac{1}{2}, \frac{1}{2}] \\ 0, & \text{elsewhere} \end{cases}$$

and let B_n ($n \in \mathbb{N}$) be defined as

$$B_n = B_0 * B_{n-1}$$

which is a convolution, defined as

$$(f * g)(t) = \int_{-\infty}^{\infty} f(t-x)g(x)dx$$

The B-splines B_n are continuous functions and B_n has compact support $\left[-\frac{n+1}{2}, \frac{n+1}{2}\right]$.

The sequence of normalized and scaled versions of B_n tends to the Gaussian function as n increases [21]:

$$\lim_{n\to\infty}\left\{\sqrt{\frac{n+1}{12}}B_n\left(\sqrt{\frac{n+1}{12}}\cdot x\right)\right\} = \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{x^2}{2}\right)$$

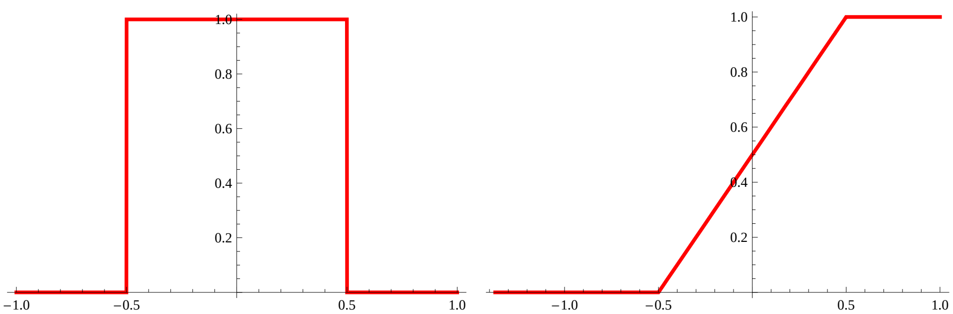
We consider distributions which have the B-splines as their density functions. We will now look at the uniform distribution B_0, the linear B-spline distribution B_1 and the normal (Gaussian) distribution, which is B_∞.

The uniform distribution

The uniform distribution has the following density and distribution function

$$f(x) = \begin{cases} 1, & x \in [-\frac{1}{2}, \frac{1}{2}] \\ 0, & \text{elsewhere} \end{cases}$$

$$F(x) = \begin{cases} 0, & x < -\frac{1}{2} \\ x + \frac{1}{2}, & x \in [-\frac{1}{2}, \frac{1}{2}] \\ 1, & x > \frac{1}{2} \end{cases}$$



(a) Density function (b) Distribution

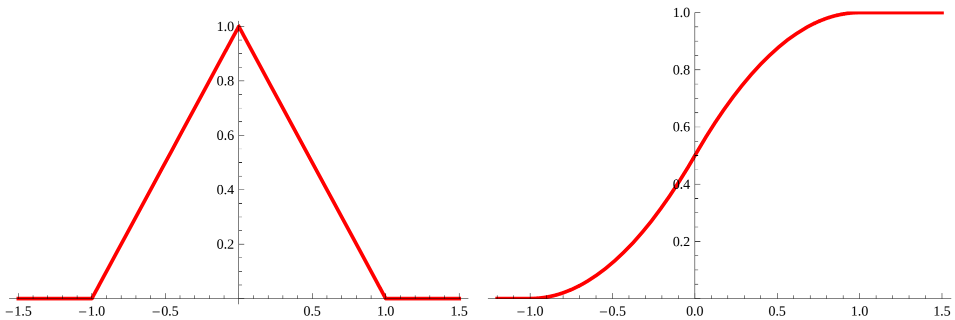
Figure A.1: The uniform distribution

The linear spline distribution

The linear distribution has the following density and distribution function

$$f(x) = \begin{cases} 1 + x, & x \in [-1, 0] \\ 1 - x, & x \in [0, 1] \\ 0, & \text{elsewhere} \end{cases}$$

$$F(x) = \begin{cases} 0, & x < -1 \\ \frac{1}{2}(1 + 2x + x^2), & x \in [-1, 0] \\ \frac{1}{2}(1 + 2x - x^2), & x \in [0, 1] \\ 1, & x > 1 \end{cases}$$



(a) Density function (b) Distribution

Figure A.2: The linear distribution

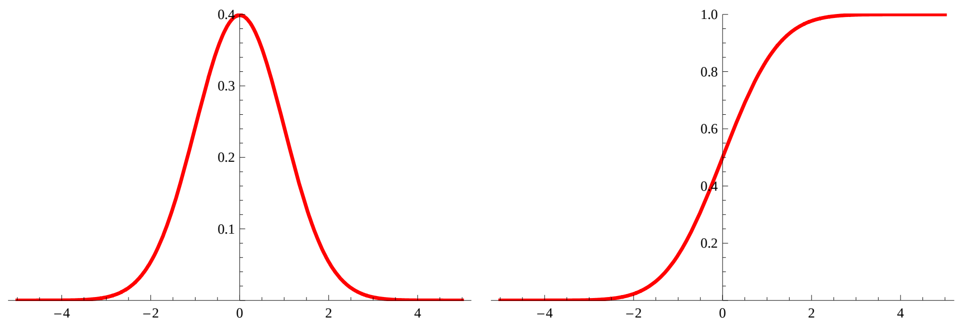
The normal distribution

We consider the normal distribution with mean $\mu = 0$ and standard deviation $\sigma = 1$. Its density function is the following:

$$f(x) = \frac{1}{\sqrt{2\pi}} \exp\left(\frac{-x^2}{2}\right)$$

Its distribution function is the error function

$$F(x) = erf(x) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} \exp\left(\frac{-x^2}{2}\right) dx$$



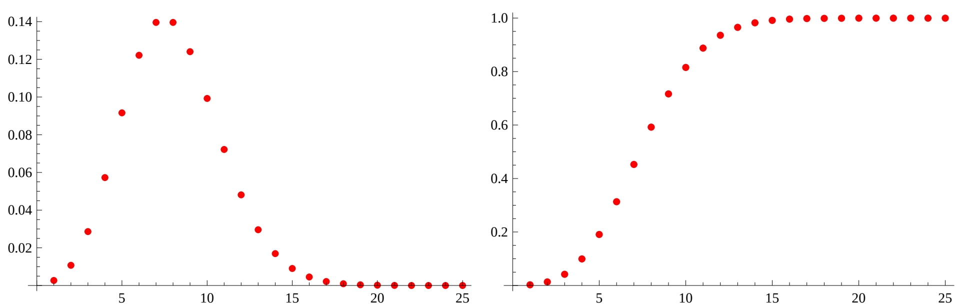
(a) Density function (b) Distribution

Figure A.3: The normal distribution

A.2 The Poisson distribution

The Poisson distribution is a discrete function. It has the following density function

$$f(\lambda, k) = \frac{\lambda^k e^{-\lambda}}{k!}, \text{ with } k \in \mathbb{N}_0$$



(a) Density function (b) Distribution

Figure A.4: The Poisson distribution

The poisson distribution has a mean and variance of λ.

Bibliography

[1] R. Anguelov, P.W. Butler, C.H. Rohwer, and M. Wild. The distribution transfers of important LULU operators and other strong filters. In preparation.

[2] M. Boodram, L. Denby, C. Mallows, and R.D. Martin. Computing linear parts of nonlinear smoothers. In *Proceedings of the Statistical Computing Section of the American Statistical Association*, 1989.

[3] C.K. Chui. *An introduction to wavelets.* Academic Press, 1992.

[4] W.J. Conradie, T. de Wet, and M. Jankowitz. Exact and asymptotic distributions of LULU smoothers. *J. Comput. Appl. Math.*, 186:253–267, 2006.

[5] J.P. du Toit. The discrete pulse transform and applications. Master's thesis, Stellenbosch University, 2007.

[6] P. Embrechts, C. Kluppelberg, and T. Mikosch. *Modelling Extremal Events for Insurance and Finance.* Springer, first edition, 1997.

[7] N.C. Gallagher and G.L. Wise. A theoretical analysis of the properties of median filters. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-29(6):1136–1141, 1981.

[8] R.W. Hamming. *Digital Filters.* Prentice-Hall, Inc., first edition, 1977.

[9] M.D. Jankowitz. *Some statistical aspects of LULU smoothers.* PhD thesis, Stellenbosch University, 2007.

[10] C.L. Mallows. Some theoretical results of Tukey's 3R smoother. In *Smoothing techniques for curve estimation*, volume 757, pages 77–90. Springer-Verlag, 1979.

[11] C.L. Mallows. Some theory of nonlinear smoothers. *Ann. Statist.*, 8(4):695–715, 1980.

[12] T.A. Nodes and N.C. Gallagher. Median Filters: Some properties and their modifications. *IEEE Trans. Acoust., Speech, Signal Processing*, ASSP-30(5):739–746, 1982.

[13] S. Perreault and P. Hébert. Median filtering in constant time. *IEEE Transactions on Image Processing*, 16(9):2389–2394, 2007.

[14] A. Restrepo and A.C. Bovic. Locally monotonic regression. *IEEE Transactions on Signal Processing*, 41(9), 1993.

[15] C.H. Rohwer. Variation reduction and LULU smoothing. *Questiones Mathematicae*, 25:163–176, 2002.

[16] C.H. Rohwer. *Nonlinear Smoothing and Multiresolution Analysis*. Birkha user Publishers, first edition, 2005.

[17] C.H. Rohwer. The estimation of moments of an unknown error distribution in the discrete pulse transform. *Numerical Algorithms*, 45:239–251, 2007.

[18] C.H. Rohwer and D.P. Laurie. The discrete pulse transform. *SIAM Journal of Mathematical Analysis*, 38(3):1012–1034, 2006.

[19] C.H. Rohwer and M. Wild. LULU theory, idempotent stack filters, and the mathematics of vision of Marr. *Advances in imaging and electron physics*, 146:57–162, 2007.

[20] I. Shmulevich, O. Yli-Harja, K. Egazarian, and J. Astola. Output distributions of recursive stack filters. *IEEE Signal processing letters*, 6(7):175–178, 1999.

[21] M. Unser, A. Aldroubi, and M. Eden. On the asymptotic convergence of B-spline wavelets to Gabor functions. *IEEE Transactions on Information Theory*, 38(2):864–872, 1992.

[22] Paul F. Velleman. Robust nonlinear data smoothers: Definitions and recommendations. *Proc. Natl. Acad. Sci. USA*, 74(2):434–435, 1977.

[23] Zhou Xing-Wei, Yang De-Yun, and Ding Run-Tao. Infinite-length roots of median filters. *Science in China*, 35(12):1496–1508, 1992.