

A FUNCTIONAL CENTRAL LIMIT THEOREM FOR RECURSIVE RESIDUALS AND APPLICATIONS IN ASYMPTOTIC STATISTICS

– DISSERTATION –

SUBMITTED BY KARSTEN EVERS
IN PARTIAL FULFILMENT OF THE REQUIREMENTS
FOR THE DEGREE OF
DOCTOR RERUM NATURALIUM
(Dr. rer. nat.)

Veitsbronn (Germany), 2022/04/26



KATHOLISCHE UNIVERSITÄT
EICHSTÄTT-INGOLSTADT

Chair of Mathematics – Statistics

First reviewer: Prof. Dr. Wolfgang Bischoff
Katholische Universität Eichstätt-Ingolstadt
Second reviewer: Prof. Dr. Enkelejd Hashorva
Université de Lausanne
Date of oral examination: 2022/07/14

Preface

In this thesis, we study stochastic processes based on recursive and ordinary residuals.

In Part A of this thesis we study partial sums of recursive residuals. We consider a classical linear regression model, where the regression functions are defined on a compact domain in the real numbers and sample data at specific points from the design area in the form of a triangular array. Next we define recursive least squares residuals and prove a number of their interesting properties. These have many remarkable advantages over classical residuals. We then consider partial sums of recursive residuals, define a recursive residual partial sum process and prove that this process converges weakly against Brownian motion. For not normally distributed errors, the limit process has been known only for time series samples under assumptions that are difficult to verify, or in the case of triangular schemes of design points only under strong assumptions on the regression functions. For the first time, we determine the limit process for triangular schemes without the restrictive assumption of normally distributed errors and under very mild assumptions on the regression functions (they have to be left continuous, of bounded variation and linearly independent with respect to the L2 norm) that are of great benefit for practical applications. Crucial for the proof are Donsker's invariance principle for triangular schemes, a technique to factorize simultaneously a family of functions and Rubin's famous continuous mapping theorem. Our approach, based on Rubin's theorem, further allows us to compute the distribution of the limit process under local alternatives. With the help of these asymptotic results we are then able to define and study asymptotic tests and we give an example of an asymptotically uniformly most powerful test.

In Part B we give an introduction to the theory of weak convergence of finite measures, and present these classical results in a more general form than usual. We consider measures on perfectly normal spaces instead on metric spaces, and we consider filters of measures instead of sequences of measures. To our knowledge, this approach has not yet been published.

In Part C, we summarize the main basics needed to understand the main statements from Part A, and we develop some technical tools needed to understand the corresponding proofs. Some of these results are also interesting in their own right, such as the simultaneous factorization of a whole family of functions mentioned above or the estimation of the eigenvalues of certain Gram matrices. Finally, all proofs of all statements of the previous sections can be found here.

Part D, the appendices, contains all R scripts of the simulations, a list of all new results, and a list of symbols and abbreviations.

Acknowledgments

I would like to thank my mentor Wolfgang Bischoff. Without his guidance, the many conversations and his patience, this project would not have been possible. And I would like to thank Enkelejd Hashorva for taking on the second review. Finally, I would also like to thank Thomas Heindl for the many helpful discussions, which often led to new insights.

Contents

Preface	2
Part A. Sums of recursive residuals and their asymptotics	4
1 Introduction	4
2 Ordinary residuals in linear regression	6
3 Recursive residuals	7
4 Recursive residual partial sum process	9
5 Recursive residual partial sum limit process	10
6 Reversed partial sums of recursive residuals	17
7 An asymptotically uniformly most powerful test	20
Part B. Weak convergence of finite measures	21
1 Basic notations and definitions	22
2 Finite measures in perfectly normal spaces	23
3 Weak convergence of finite measures	24
4 Convergence in distribution	27
Part C. Technical statements and proofs	29
1 Brownian motion and Donskers Theorem	29
2 Functions with existing limits from left and right	31
3 Asymptotic- F -designs	35
4 Bounds for largest and smallest eigenvalues of a Gram matrix	36
5 A theorem on continuous convergence in $C[0, 1]$	38
6 Proofs	39
Part D. Appendices	75
1 List of figures and tables and R scripts generating them	75
2 New results	79
3 List of abbreviations and symbols	80
Bibliography	82

Written in L^AT_EX.

Part A. Sums of recursive residuals and their asymptotics

1 Introduction

We consider regression models of the form

$$Y = g + \varepsilon \quad (1)$$

where $g : \mathcal{E} \rightarrow \mathbb{R}$ is an unknown deterministic regression function, $\varepsilon : \Omega \rightarrow \mathbb{R}$ is a random variable (the error) with $E(\varepsilon) = 0$ and $Var(\varepsilon) = \sigma^2 < \infty$, where (Ω, \mathcal{A}, P) is the unknown probability space, $\mathcal{E} \subseteq \mathbb{R}$ is the experimental region and Y is the observable result of the experiment. To get information on the unknown function g we sample data at certain design points $t_{n_0,1} \leq t_{n_0,2} \leq \dots \leq t_{n_0,n_0}$ of the experimental region \mathcal{E} and embed these into a triangular array of design points $t_{n,1} \leq t_{n,2} \leq \dots \leq t_{n,n}$, $n = 1, 2, 3, \dots$, to obtain asymptotic information on the regression model as n goes to infinity. To do this, we assume that the empirical distribution function F_n of the design array converges to a limit design F (limit distribution), in a sense precisely defined later. Especially, we are interested in the case that the experimental region \mathcal{E} is a compact interval. Therefore, we must consider a triangular array of design points to get asymptotic results, rather than design points $t_1 < t_2 < \dots < t_n < \dots$ from a time series sampling which can be regarded as specific triangular array. Without loss of generality we can assume for our experimental region $\mathcal{E} = [a, b]$ with $a < b$. We are interested in testing whether the above model is a linear regression, i.e., more exactly, in testing

$$H_0 : g = \sum_{k=1}^d \beta_k f_k \quad \text{vs.} \quad H_1 : g \neq \sum_{k=1}^d \beta_k f_k, \quad (2)$$

where $f_1, \dots, f_d : [a, b] \rightarrow \mathbb{R}$ are known, left continuous functions of bounded variation such that $f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]}$ are linearly independent in $L_2([a,b], F)$ for some $c \in (a, b)$ and $\beta = (\beta_1, \dots, \beta_d)^\top \in \mathbb{R}^d$ is the vector of the unknown regression parameters.

Recursive residuals play a major role in problems concerning the change point of a sequence of random variables, such as the CUSUM and the CUSUM of squares tests introduced by Brown et al. (1975). The CUSUM test is based on the sum of the recursive residuals. If this exceeds a critical bound, it is concluded that there is a structural break. The CUSUM of squares test plots the cumulative sum of squared recursive residuals divided by the squared sum over all observations. These are an alternative to ordinary residual-based regression diagnostics. In fact, recursive residuals have been known since 1891, cf. Farebrother (1978) and are very popular among statistical users and researchers today. Google Scholar lists nearly 2000 (!) articles containing “recursive residuals” from 2020 to the present alone. In contrast, recursive residuals seem relatively less discussed in books. Some notable recent and older monographs and textbooks that contain some material on recursive residuals are Hawkins

(1980), Sen (1985), Kramer and Sonnberger (1986), Harvey (1989), Wells (1996), Csörgo and Horvath (1997), Hawkins and Olwell (1998), Herrmann (2005), Kennedy (2008), Baltagi (2011), Young (2011), Hackl (2012) and Paoletta (2019).

We consider partial sums of recursive residuals and use them to define stochastic processes and we are interested in the asymptotics of these processes for $n \rightarrow \infty$. To determine the limit process, we proceed in two steps. In the first step, using Rubin's continuous mapping theorem, we determine the limit process for continuous regression functions. In the second step, we decompose the given regression functions $f_i = g_i \circ G^-$, which we now only assume to be left continuous, into continuous functions g_i and a quantile function G^- , where we "move" the discontinuity points of all f_i in G^- and G is a continuous distribution function. This way, we finally succeed in computing the limit process even under the weaker assumptions of left continuous regression functions. It turns out that, under suitable assumptions, the limit process depends only on the limit design (i.e., on F). Thus, we have an invariance principle.

Such results were obtained by MacNeill (1978a, b) for ordinary least squares residuals w_n , given a triangular array of equidistant design points. Bischoff (1998) generalized this result for arbitrary triangular design points and less restrictive regularity assumptions on the regression functions. Both results were proved by using Rubin's continuous mapping theorem. Corresponding results for recursive residuals were proved by Sen (1982), Csörgo and Horvath (1997), Rabovski (2003), Otto (2019) and Sakhanenko et al. (2021). However, Sen proved his result for a time series sampling $t_1 < t_2 < \dots$, instead for a triangular array (note that most results on recursive residuals are developed for time series samplings). Moreover, the conditions (2.9) to (2.11) in Theorems 1 resp. 2 in Sen (1982) are given by complicated functions of the model variables and even in the simple example of a polynomial regression considered there, they can only be checked with considerable effort. Whether this can also be shown for more complicated examples or even in our general model is unknown. In contrast, the conditions given by us refer directly to the regression functions of the model and can therefore be easily checked (and are obviously fulfilled in the example mentioned). Furthermore, he mentioned that his results cannot be proved using the continuous mapping theorem, instead he used the concept of tightness. The result in Csörgo and Horvath (1997), Theorem 3.4.3, is more similar to our results. However, only equidistant design points $t_{n,i} = \frac{i}{n}$ are considered there and, in addition, strong restrictive regularity conditions are imposed on the regression functions (e.g. these must be differentiable and satisfy $\max_{1 \leq j \leq d} \sup_{0 \leq t \leq 1} |f'_j(t)| < \infty$). In Rabovski (2003), only the special case of a constant regression function is studied (i.e., $d = 1$ and $f_1(t) = 1$) and in Sakhanenko et al. (2021) only the special case for one-dimensional β is considered. Finally, Otto (2021) goes in the same direction as Sen and considers time series samples, although it should be noted that the underlying model differs from ours anyway.

Results on the limit process under a local alternative are known in our setting so far only in special cases, cf. Rabovski (2003). For time series samplings and in another context (testing for changes in regression coefficients β), corresponding statements can be found in Ploberger and Krämer (1990) and Otto (2019), although our results cannot be derived from theirs. Our approach, based on Rubin's continuous mapping theorem, allows us to compute the distribution of the limit process, under the local alternative given in (2) for an f_{d+1} of bounded variation, for the first time in the form of an explicit formula. With the help of these asymptotic results we are then able to define and study asymptotic tests. By analogy with partial sums of recursive residuals, we also consider reverse partial sums and their limits. It turns out that tests based on these processes have greater power and we give an example of such an asymptotically uniformly most powerful test.

2 Ordinary residuals in linear regression

We consider model (1) under H_0 given in (2) with $\mathcal{E} = [a, b]$. Sequentially sampled observations $Y_{n,1}, \dots, Y_{n,n}$ at the design points $a \leq t_{n,1} \leq t_{n,2} \leq \dots \leq t_{n,n} \leq b$ leads to n linear models

$$\underbrace{\begin{pmatrix} Y_{n,1} \\ \vdots \\ Y_{n,i} \end{pmatrix}}_{=: Y_i} = \underbrace{\begin{pmatrix} f_1(t_{n,1}) & \dots & f_d(t_{n,1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{n,i}) & \dots & f_d(t_{n,i}) \end{pmatrix}}_{=: X_{n,i} \in \mathbb{R}^{i \times d}} \cdot \underbrace{\begin{pmatrix} \beta_1 \\ \vdots \\ \beta_d \end{pmatrix}}_{=: \beta} + \underbrace{\begin{pmatrix} \varepsilon_{n,1} \\ \vdots \\ \varepsilon_{n,i} \end{pmatrix}}_{=: \varepsilon_i}, \quad i \in \{1, \dots, n\}. \quad (3)$$

We assume

$$\varepsilon_{n,1}, \dots, \varepsilon_{n,n} \text{ i.i.d. with } E(\varepsilon_{n,i}) = 0 \text{ and } \text{Var}(\varepsilon_{n,i}) = \sigma^2, \quad i = 1, \dots, n, \quad (4)$$

holds and $f_1, \dots, f_d : [a, b] \rightarrow \mathbb{R}$ are of bounded variation, left continuous on $(a, b]$ and continuous in a and furthermore $f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]}$ are linearly independent in $L_2([a, b], F)$, for some $c \in (a, b]$, where F , in a sense precisely defined later, is the distribution function of the limit design of the design sequence $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$. By the assumption that the observations $Y_{n,1}, \dots, Y_{n,n}$ are sequentially sampled, the unknown parameter β can be uniquely estimated, as soon as $\text{rank}(X_{n,i}) = d$. The linear independence of $f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]}$ in $L_2([a, b], F)$, together with appropriate conditions we impose on the design sequence, guarantees, that there exists an $n_0 \in \mathbb{N}$ such that all design matrices $X_{n,k}$, $n \geq n_0$, relevant to the recursive residual partial sum process to be defined later (cf. Definition 4), have full rank (cf. Remark 76). Therefore, in order not to complicate the presentation unnecessarily, we assume

$$\text{rank}(X_{n,d}) = d. \quad (5)$$

So, according to the sequentially sampled observations we have $n - d + 1$ linear models

$$Y_i = X_{n,i} \beta + \varepsilon_i, \quad i \in \{d, \dots, n\}, \quad (6)$$

where the ordinary least squares (OLS) estimation for β can be updated after each observation $Y_{n,i}, i = d, \dots, n$, by

$$\hat{\beta}_{n,i} = (X_{n,i}^\top X_{n,i})^{-1} X_{n,i}^\top Y_i = \beta + (X_{n,i}^\top X_{n,i})^{-1} X_{n,i}^\top \varepsilon_i, \quad i \in \{d, \dots, n\}.$$

By the Gauss-Markov theorem, $\hat{\beta}_{n,i}$ is the best linear unbiased estimation (BLUE) for β given the observation $Y_{n,1}, \dots, Y_{n,i}$. In the sequel we use the notation

$$f : [a, b] \rightarrow \mathbb{R}^d, \quad t \mapsto (f_1(t), \dots, f_d(t))^\top$$

for the vector of known regression functions. The least squares estimation $\hat{\beta}_{n,i}$ for β using the observations $Y_{n,1}, \dots, Y_{n,i}$ can be applied to forecast the expectation $E(Y_{n,i+1}) = f(t_{n,i+1})^\top \beta$ of the next observation $Y_{n,i+1}, i = d, \dots, n - 1$ by the best linear unbiased estimation

$$f(t_{n,i+1})^\top \hat{\beta}_{n,i} = f(t_{n,i+1})^\top (X_{n,i}^\top X_{n,i})^{-1} X_{n,i}^\top Y_i, \quad i \in \{d, \dots, n - 1\}.$$

We abbreviate this estimator by

$$a_{n,i+1}^\top := f(t_{n,i+1})^\top (X_{n,i}^\top X_{n,i})^{-1} X_{n,i}^\top \in \mathbb{R}^i, \quad i = d, \dots, n - 1.$$

It holds

$$\begin{aligned} a_{n,i+1}^\top Y_i &= f(t_{n,i+1})^\top (X_{n,i}^\top X_{n,i})^{-1} X_{n,i}^\top (X_{n,i} \beta + \varepsilon_i) \\ &= f(t_{n,i+1})^\top \beta + f(t_{n,i+1})^\top (X_{n,i}^\top X_{n,i})^{-1} X_{n,i}^\top \varepsilon_i, \quad i \in \{d, \dots, n-1\}, \end{aligned} \quad (7)$$

implying for $i \in \{d, \dots, n-1\}$

$$E(a_{n,i+1}^\top Y_i) = f(t_{n,i+1})^\top \beta = E(Y_{n,i+1}), \quad (8)$$

$$\text{Var}(a_{n,i+1}^\top Y_i) = \sigma^2 f(t_{n,i+1})^\top (X_{n,i}^\top X_{n,i})^{-1} f(t_{n,i+1}) = \sigma^2 \|a_{n,i+1}\|^2, \quad (9)$$

where $\|\cdot\|$ is the Euclidian norm. We consider the linear model given in (3) for all n observations $Y_{n,1}, \dots, Y_{n,n}$, i.e. for $i=n$,

$$Y_n = X_{n,n} \beta + \varepsilon_n. \quad (10)$$

Then the ordinary (linear) regression residuals are defined for model (10) as the difference between the observed value $Y_{n,j}$ and the least squares estimation $f(t_{n,j})^\top \hat{\beta}_{n,n}$ using the information of all data to estimate $E(Y_{n,j}) = f(t_{n,j})^\top \beta$ in (10), i.e.

$$w_{n,j} := Y_{n,j} - f(t_{n,j})^\top \hat{\beta}_{n,n} = Y_{n,j} - f(t_{n,j})^\top (X_{n,n}^\top X_{n,n})^{-1} X_{n,n}^\top Y_n, \quad j = 1, \dots, n. \quad (11)$$

Let $\text{pr}_{U,V} : U \oplus V \rightarrow U, u + v \mapsto u$ be the projection on the first summand of a direct sum. Thus

$$w_n := (w_{n,1}, \dots, w_{n,n})^\top = Y_n - X_{n,n} \hat{\beta}_{n,n} = \left(I_n - X_{n,n} (X_{n,n}^\top X_{n,n})^{-1} X_{n,n}^\top \right) Y_n = \text{pr}_{U_n^\perp} Y_n, \quad (12)$$

where $U_n := \text{im}(X_{n,n})$ and U_n^\perp is the subspace orthogonal to U_n and furthermore we write abbreviatively $\text{pr}_U := \text{pr}_{U, U^\perp}$ for a subspace U of \mathbb{R}^n . The proof of (12) is basic linear algebra (projections), which we have summarized for the reader in Lemma 47.

3 Recursive residuals

Recursive residuals are defined as the difference between the present observation and its forecasting obtained by the observations sampled before. Furthermore they are standardized (thus they are standardized forecast errors) and have many nice properties (e.g. they are homoscedastic and uncorrelated), as shown below. Recursive residuals can be considered as the result of transforming the correlated and heteroscedastic ordinary regression residuals to uncorrelated and homoscedastic random variables, see below.

Definition 1. For the $n - d$ linear models given in (6), under the assumptions (4) and (5), recursive residuals $r_{n,j}$ are defined by

$$r_{n,j} := \frac{Y_{n,j} - a_{n,j}^\top Y_{j-1}}{(1 + \|a_{n,j}\|^2)^{\frac{1}{2}}}, \quad j = d + 1, \dots, n,$$

with $a_{n,j}^\top = f(t_{n,j})^\top (X_{n,j-1}^\top X_{n,j-1})^{-1} X_{n,j-1}^\top \in \mathbb{R}^{1 \times (j-1)}$.

(8) and (9) imply immediately $E(r_{n,j}) = 0$, $\text{Var}(r_{n,j}) = \sigma^2$, $j = d+1, \dots, n$, and by (7),

$$r_{n,j} := \frac{\varepsilon_{n,j} - a_{n,j}^\top \varepsilon_{j-1}}{(1 + \|a_{n,j}\|^2)^{\frac{1}{2}}}, \quad j = d+1, \dots, n. \quad (13)$$

We now rewrite (13) as a matrix equation. To do so, we define

$$M_{n,k}^\top := \underbrace{\begin{pmatrix} \frac{1}{\sqrt{1+\|a_{n,d+1}\|^2}} & \cdots & 0 \\ \vdots & \ddots & \vdots \\ 0 & \cdots & \frac{1}{\sqrt{1+\|a_{n,k}\|^2}} \end{pmatrix}}_{\in \mathbb{R}^{(k-d) \times (k-d)}} \cdot \underbrace{\begin{pmatrix} -a_{n,d+1}^\top & 1 & 0 & \cdots & 0 \\ \vdots & & \ddots & \ddots & \vdots \\ \vdots & & & \ddots & 0 \\ -a_{n,k}^\top & & & & 1 \end{pmatrix}}_{\in \mathbb{R}^{(k-d) \times k}} \in \mathbb{R}^{(k-d) \times k}.$$

Because of

$$(-a_{n,j}^\top \ 1)X_{n,j} = \left(-f(t_{n,j})^\top (X_{n,j-1}^\top X_{n,j-1})^{-1} X_{n,j-1}^\top \ 1 \right) \cdot \underbrace{\begin{pmatrix} X_{n,j-1} \\ f(t_{n,j})^\top \end{pmatrix}}_{=X_{n,j}} = (0 \dots 0) \in \mathbb{R}^d$$

we get

$$M_{n,k}^\top X_{n,k} = 0 \quad \text{and} \quad M_{n,k}^\top M_{n,k} = I_{k-d} \quad (14)$$

and (13) thus becomes

$$r_n = M_{n,n}^\top \cdot \varepsilon_n = M_{n,n}^\top \cdot Y_n, \quad (15)$$

where $r_n := (r_{n,d+1}, \dots, r_{n,n})^\top$ is the vector of recursive residuals. If $\varepsilon_n \sim N(0, \sigma^2 I_n)$ holds, (14) and (15) together imply

$$r_n \sim N(0, \sigma^2 I_{n-d}) \quad \text{and} \quad r_{n,d+1}, \dots, r_{n,n} \quad \text{are independent.} \quad (16)$$

Furthermore

$$M_{n,k} M_{n,k}^\top = \text{pr}_{U_k^\perp} \quad \text{with} \quad U_k := \text{Im}(X_{n,k}), \quad (17)$$

which follows immediately from Lemma 47. Next, we state a relation between the vector of ordinary residuals $w_n = (w_{n,1}, \dots, w_{n,n})^\top$, defined in (12) and r_n from (15). It holds

$$M_{n,n}^\top w_n = r_n \quad \text{and} \quad M_{n,n} r_n = w_n \quad (18)$$

since $w_n = \text{pr}_{U_n^\perp} Y_n = \text{pr}_{U_n^\perp} \varepsilon_n = M_{n,n} M_{n,n}^\top \varepsilon_n = M_{n,n} r_n$. Almost all of the results in this section so far follow more or less directly from the definition and can be found, for example, in Brown et al. (1975). We close this section with a theorem (which can be found in Brown et al. (1975) too, but goes back in part to Plackett (1950) and Bartlett (1951)), where we summarize some technical properties that are useful for the efficient computation of recursive residuals.

Theorem 2. *Given the linear models (6) under conditions (4) and (5).*

$$1. \ X_{n,j}^\top X_{n,j} = X_{n,j-1}^\top X_{n,j-1} + f(t_{n,j}) f(t_{n,j})^\top = \sum_{i=1}^j f(t_{n,i}) f(t_{n,i})^\top$$

2. $(X_{n,j}^\top X_{n,j})^{-1} = (X_{n,j-1}^\top X_{n,j-1})^{-1} - \frac{1}{1+\|a_{n,j}\|^2} (X_{n,j-1}^\top X_{n,j-1})^{-1} f(t_{n,j}) f(t_{n,j})^\top (X_{n,j-1}^\top X_{n,j-1})^{-1}$
3. $\hat{\beta}_{n,j} = \hat{\beta}_{n,j-1} + \sqrt{1+\|a_{n,j}\|^2} r_{n,j} (X_{n,j}^\top X_{n,j})^{-1} f(t_{n,j}), j \in \{1, \dots, n\}$.
4. $\|w_{d+1}\|^2 = \|r_{n,d+1}\|^2$ and $\|w_j\|^2 = \|w_{j-1}\|^2 + r_{n,j}^2$ for $j = d+2, \dots, n$ where $w_j := Y_j - X_{n,j} \hat{\beta}_{n,j}, j \in \{d+1, \dots, n\}$.

The proof can be found on page 39.

4 Recursive residual partial sum process

All processes considered in this work are defined with the help of the following operator.

Definition 3. The partial sum operator $T_n : \mathbb{R}^n \rightarrow C[0, 1], a := (a_1, \dots, a_n)^\top \mapsto T_n(a)$,

$$T_n(a)(x) := \begin{cases} a_k(nx - k) + \sum_{i=1}^k a_i & \text{for } x \in [\frac{k-1}{n}, \frac{k}{n}), k \in \{1, \dots, n\}, \\ \sum_{i=1}^n a_i & \text{for } x = 1. \end{cases} \quad (19)$$

$T_n(a)$ is a continuous piecewise linear function on $[0, 1]$, where the line segments connect the points $(\frac{k}{n}, \sum_{i=1}^k a_i), k = 0, \dots, n$. Note $\sum_{i=1}^m a_i = 0$ for $m \leq 0$. Thus for (19) we can write

$$T_n(a)(x) = (nx - \lfloor nx \rfloor) a_{\lfloor nx \rfloor + 1} + \sum_{i=1}^{\lfloor nx \rfloor} a_i, x \in [0, 1]. \quad (20)$$

Definition 4. Let $\delta \in (0, 1)$ be fixed. The recursive residual partial sum process is defined as

$$B_{n,\delta} := \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \circ r_n, \quad (21)$$

where σ is defined in (4), r_n is defined in (15) and $g = g_\delta : \mathbb{R}^{n-d} \rightarrow \mathbb{R}^{n-d}$ is defined by

$$a = (a_1, \dots, a_{n-d})^\top \mapsto (0, \dots, 0, a_{\lfloor (n-d)\delta \rfloor + 1}, \dots, a_{n-d})^\top. \quad (22)$$

Later $\delta = F(c)$ will hold, where the c comes from the assumption

$$f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]} \text{ are linearly independent in } L_2([a,b], F) \text{ for some } c \in (a,b) \quad (23)$$

and F is the distribution function of the limit design of the design sequence $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$, in a sense precisely defined later (cf. Definition 9). We need condition (23) on the one hand to ensure in the proofs that the eigenvalues of certain Gram matrices are bounded (cf. Lemma 75 and Lemma 78) and on the other hand to ensure that the process exists at all. Note that $\lfloor (n-d)F(c) \rfloor \geq \lfloor nF(c) \rfloor - d$ holds and thus, according to Remark 76, there is $n_0 \in \mathbb{N}$ such that all recursive residuals relevant to the process from (21) actually exist (i.e., the corresponding design matrix $X_{n,k}$ has full rank for all $n \geq n_0$ and all $k \geq \lfloor (n-d)F(c) \rfloor + 1$). In practical applications, usually only functions f_1, \dots, f_d that satisfy (23) for each $c \in (a,b]$ occur. Provided that the limit designs are chosen such that no mass lies in a , i.e. $F(a) = 0$, from a theoretical

point of view δ can be chosen arbitrarily small ($\delta \rightarrow 0$). The composition of T_{n-d} and g is computed by means of (20) to

$$T_{n-d}(g(a))(x) = ((n-d)x - \lfloor (n-d)x \rfloor) a_{\lfloor (n-d)x \rfloor + 1} + \sum_{i=\lfloor (n-d)\delta \rfloor + 1}^{\lfloor (n-d)x \rfloor} a_i \quad (24)$$

for $x \geq \frac{\lfloor (n-d)\delta \rfloor}{n-d}$ and $T_{n-d}(g(a))(x) = 0$ for $x < \frac{\lfloor (n-d)\delta \rfloor}{n-d}$. Note that $\delta \in \left[\frac{\lfloor (n-d)\delta \rfloor}{n-d}, \frac{\lfloor (n-d)\delta \rfloor + 1}{n-d} \right)$.

$r_n : \Omega \rightarrow \mathbb{R}^{n-d}$ from (15) is a real multivariate random variable and $B_{n,\delta} : \Omega \rightarrow C[0, 1]$ is a random variable with values in $C[0, 1]$, cf. the sample paths of $B_{n,\delta}$ in Figure 1.

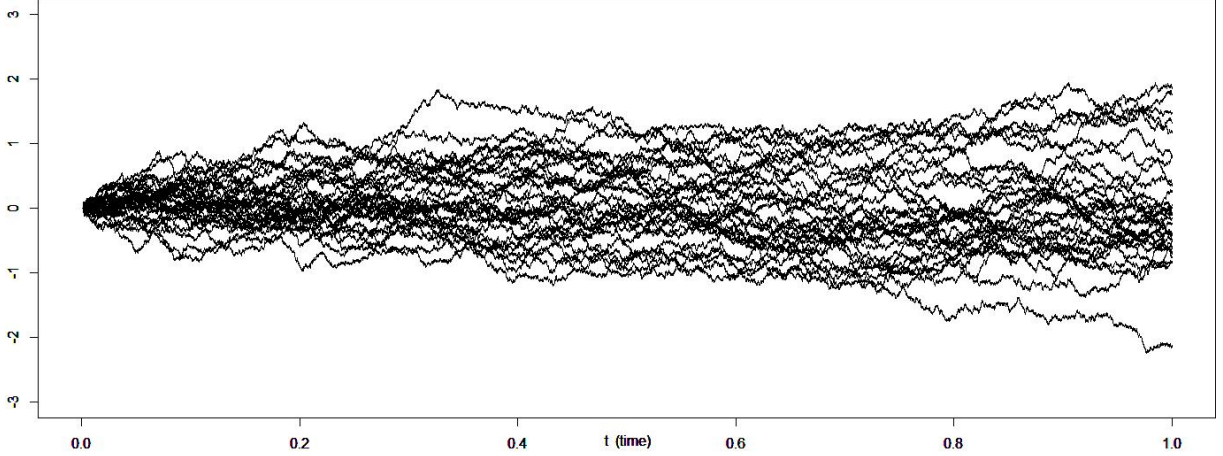


Figure 1: 30 randomly selected sample paths of $B_{n,\delta}$, whereby $n = 10000$, $f(t)^\top = (0.5, t, t^2)$ for $t \in [0, 0.25]$, $f(t)^\top = (1, t, t^2)$ for $t \in (0.25, 0.5]$, $f(t)^\top = (1, t, (t - 3/2)^2)$ for $t \in (0.5, 1]$, $d = 3$, $t_{n,i} = \frac{i}{n}$, $\varepsilon_{n,i} \sim \text{unif}([-1/2, 1/2])$ and $\delta = 0.001$, cf. R-script on p. 75.

Remark 5. If $\varepsilon_n \sim N(0, \sigma^2 I_n)$, then $r_n \sim N_{n-d}(0, \sigma^2 I_{n-d})$ by (16) and

$$\frac{1}{\sigma\sqrt{n-d}} T_{n-d} \circ g \circ r_n \xrightarrow{\mathcal{D}} B' \quad \left(P \left(\frac{1}{\sigma\sqrt{n-d}} T_{n-d} \circ g \circ r_n \right) \xrightarrow{w} P^{B'} \text{ , cf. definition 43} \right) \quad (25)$$

follows by Theorem 53, where B' , $(B'_t)_{t \in [0,1]}$ is a shifted Brownian motion, i.e. $B'_t(\omega) = 0$ for all $t \in [0, \delta)$ and $B'_t(\omega) = B_{t-\delta}(\omega)$, whereby B , $(B_t)_{t \in [0,1]}$ is a Brownian motion on $[0, 1]$.

If ε_n is not normally distributed, the limit process is known only for time series under conditions that are difficult to check, or for triangular schemes under strong restrictions on the regression functions, or only in special cases (cf. Introduction on page 4). Our goal in the following is to determine the limit process (i.e., the weak limit) of $B_{n,\delta}$ for a triangular array of design points under much weaker restrictions on the regression functions.

5 Recursive residual partial sum limit process

Each design $a \leq t_{n,1} \leq t_{n,2} \leq \dots \leq t_{n,n} \leq b$ corresponds to a probability measure $(t_{n,1}, \dots, t_{n,n}) \mapsto P_n := \frac{1}{n} \sum_{i=1}^n \delta_{t_{n,i}}$ with Dirac measures $\delta_{t_{n,i}}$ on $t_{n,i}$. Let F_n be the distribution function of P_n , i.e.

$$F_n : \mathbb{R} \rightarrow [0, 1], \quad t \mapsto \frac{1}{n} \sum_{i=1}^n \mathbf{1}_{[t_{n,i}, \infty)}(t). \quad (26)$$

Note this implies $F_n(t) = 0$ for all $t < a$ and $F_n(t) = 1$ for all $t \geq b$.

Definition 6. For a measure P on \mathbb{R} , we define the support by

$$\begin{aligned} \text{support}(P) &:= \{x \in \mathbb{R} \mid \forall \varepsilon > 0 : P((x - \varepsilon, x + \varepsilon)) > 0\} \\ &= \mathbb{R} \setminus \bigcup_{O \in \tau^*} O, \quad \tau^* := \{O \subseteq \mathbb{R} \mid O \text{ is open and } P(O) = 0\}. \end{aligned}$$

For a measure-defining function F , cf. Theorem 63, let $\text{support}(F) := \text{support}(\mu_F)$.

Definition 7. Let F be a distribution function. We call $(F^-(1/n), F^-(2/n), \dots, F^-(1))$ upper F -design for n observations, $n \in \mathbb{N}$, where F^- is the quantil function (cf. Definition 56).

Lemma 8. Let $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$ be an upper F -design for n observations. Then

$$F_n(t) \leq F(t) \leq F_n(t) + \frac{1}{n} \quad \text{for all } t \in \mathbb{R}. \quad (27)$$

The proof can be found on page 40.

Definition 9. We call a sequence $(t_{n,1}, \dots, t_{n,n})$, $n \in \mathbb{N}$, asymptotic- F -design if

$$\sup_{t \in \mathbb{R}} |F_n(t) - F(t)| \rightarrow 0 \quad \text{for } n \rightarrow \infty. \quad (28)$$

If we write asymptotic F -design in $[a, b]$, this means $a \leq t_{n,1} \leq \dots \leq t_{n,n} \leq b$ for all $n \in \mathbb{N}$. In this case $F(t) = 0$ for $t < a$ and $F(t) = 1$ for $t \geq b$, according to (28). If F is continuous, then (28) holds if $F_n(t) \xrightarrow{n \rightarrow \infty} F(t)$ holds at every t where F is continuous (cf. Kannan and Krueger (1996), Theorem 1.4.2 or Roussas (1997), §8.6* Polya's Lemma). Eisenberg and Shixin (1983) give a characterization of (28) in terms of the convergence of characteristic functions. Note in this context also Theorem 42, Definition 61 and Lemma 62.

Limit process of $B_{n,F(c)}$ under null hypothesis

To simplify notation, we suppress the dependence of $F(c)$ in $B_{n,F(c)}$ and write $B_n := B_{n,F(c)}$.

Theorem 10. Let $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$ be an asymptotic F -design in $[a, b]$ and consider model (3) with $f_1, \dots, f_d : [a, b] \rightarrow \mathbb{R}$ all of bounded variation, left continuous on $(a, b]$, continuous in a , let $f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]}$ be linearly independent (l.i.) in $L_2([a, b], F)$ for some $c \in (a, b)$ and let $\delta := F(c) > 0$. Under the assumptions of H_0 given in (2), $B_n \xrightarrow{\mathcal{D}} B'$ holds true, whereby B' is a shifted Brownian motion (cf. Remark 5 for the definition of B').

The proof can be found on page 40. Thus, using this theorem, we can use the test statistic B_n directly for an asymptotic size α test. The corresponding procedure is well known. We only need appropriate knowledge about the boundary crossing probability for a Brownian motion. These probabilities are well known, see Theorem 51.

Theorem 11. Under the assumptions of Theorem 10, an asymptotic test of size α is given by each of the following rules.

$$\text{reject } H_0 \text{ given in (2)} \quad :\Leftrightarrow \quad \exists s \in [0, 1] \text{ with } \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d}(g \circ r_n)(s) > x, \quad (29)$$

$$\text{reject } H_0 \text{ given in (2)} \quad :\Leftrightarrow \quad \exists s \in [0, 1] \text{ with } \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d}(g \circ r_n)(s) < y, \quad (30)$$

$$\text{reject } H_0 \text{ given in (2)} \quad :\Leftrightarrow \quad \exists s \in [0, 1] \text{ with } \left| \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d}(g \circ r_n)(s) \right| \geq z, \quad (31)$$

where $x := (1 - \delta) \cdot \Phi^{-1}(1 - \frac{\alpha}{2})$, $y := (1 - \delta) \cdot \Phi^{-1}(\frac{\alpha}{2}) = -x$, $z := (1 - \delta)z'$, with z' chosen such that $\sum_{n=-\infty}^{\infty} (\Phi((4n+1)z') - \Phi((4n-1)z') - \Phi(-(4n-3)z') + \Phi(-(4n-1)z')) \geq 1 - \alpha$ holds, δ is defined in Theorem 10 and $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$.

The proof can be found on page 40.

Remark 12. To determine x, y and z in Theorem 11, one needs $\Phi^{-1}(1 - \frac{\alpha}{2})$, $\Phi^{-1}(\frac{\alpha}{2})$ and z' . For the significance levels $\alpha = 0.01$ and $\alpha = 0.05$ these are given in the table. For the computation we use $\Phi^{-1}(1 - \frac{\alpha}{2}) = -\Phi^{-1}(\frac{\alpha}{2})$ and the R Script on page 76. See also Figure 2.

α	$\Phi^{-1}(1 - \frac{\alpha}{2})$	$\Phi^{-1}(\frac{\alpha}{2})$	z'
0.05	1.959964	-1.959964	2.2414028
0.01	2.575829	-2.575829	2.8070338

From the definition of the partial sum operator (19), it follows that the test statistics (29), (30) and (31) can be computed (i.e., updated) stepwise with each new recursive residual, and H_0 can be rejected as soon as the bounds are crossed the first time.

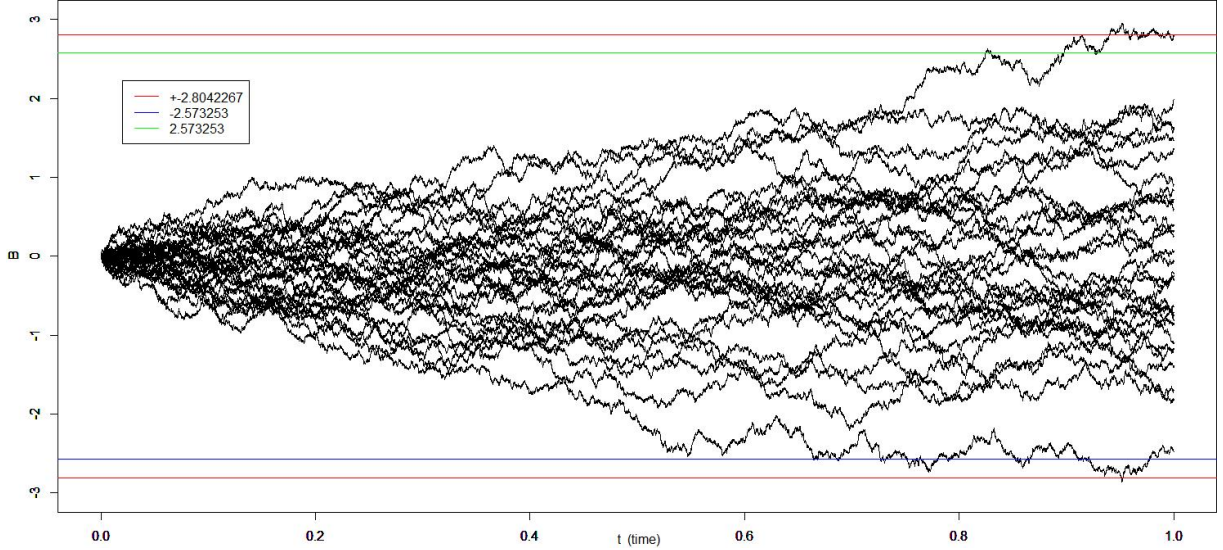


Figure 2: A very rare event: two out of 30 sample paths of a Brownian motion on $[\delta, 1]$ cross both outer boundaries (i.e. z). The horizontal lines represent the boundaries x , y ($= -x$) and z from Theorem 11 for $\delta = 0.001$ and $\alpha = 0.01$. R-script on p. 76

Limit process of B_n under a local alternative

We now consider the assumption that the (true) observations Y in fact did not occur under H_0 , but under the local alternative H_1 , given in (32),

$$H_0 : Y = \sum_{k=1}^d \beta_k f_k + \varepsilon \quad \text{vs.} \quad H_1 : Y = \sum_{k=1}^{d+1} \beta_k f_k + \varepsilon, \quad \beta_{d+1} \neq 0, \quad (32)$$

where $f_{d+1} : [a, b] \rightarrow \mathbb{R}$ is a known function of bounded variation.

Example 13. In Figure 1 (p. 10) we plotted 30 randomly chosen sample paths of B_n , assuming that the observations occurred under $H_0 : Y(t) = \sum_{k=1}^3 \beta_k \cdot f_k(t) + \varepsilon(t)$. In Theorem 10 we proved that $B_n \xrightarrow{\mathcal{D}} B$ holds (assuming H_0). How does B_n and its limit process B change if the observations were in fact made under the local alternative $H_1 : Y(t) = \sum_{k=1}^4 \beta_k \cdot f_k(t) + \varepsilon(t)$, $f_4(t) = t^3$, but the recursive residuals are computed under H_0 (cf. Figure 3)?

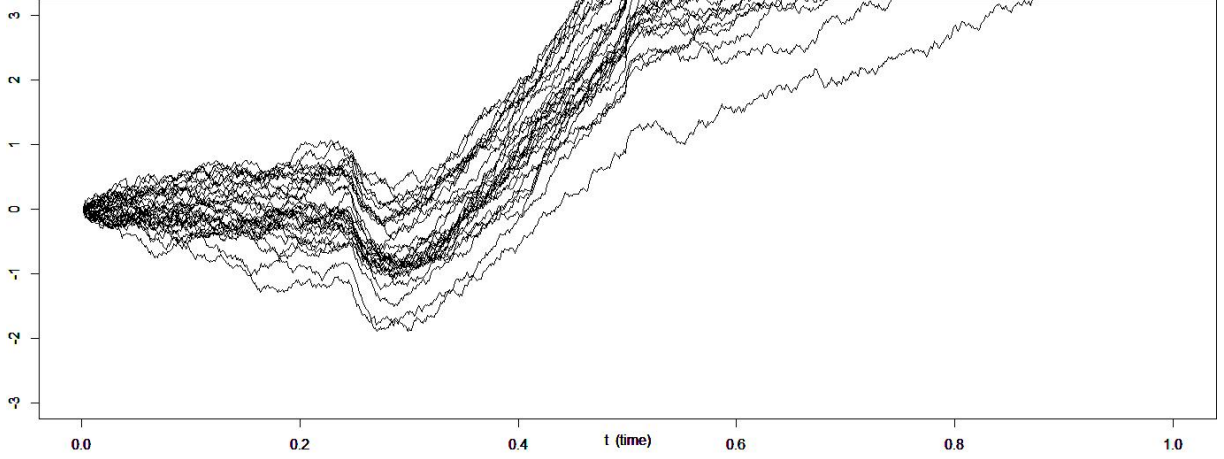


Figure 3: 30 randomly selected sample paths of B_n , where the observations in truth occurred under H_1 and f_1 to f_3 are defined as in Figure 1 and $f_4(t) := t^3$ with $n = 1000$, $t_{n,i} = \frac{i}{n}$, $\varepsilon_{n,i} \sim \text{unif}([-\frac{1}{2}, \frac{1}{2}])$ and $\delta = 0.001$. It looks like a somewhat strange but quite clearly drifted Brownian motion (cf. R-script on p. 76).

How can the trend from Figure 3 be explained? To do so, we follow Bischoff and Miller (2000) and Rabovski (2003) and compute B_n under the assumption that the true observations in fact occurred under the local alternative H_1 from (32), but the recursive residuals were computed under the false assumption H_0 :

$$\begin{aligned}
B_n &= \frac{1}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \circ r_n = \frac{1}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \circ \left(\underbrace{M_{n,n}^\top}_{\text{computed under } H_0} \cdot \underbrace{\left(\tilde{X}_{n,n} \tilde{\beta} + \varepsilon_n \right)}_{\text{observations under } H_1} \right) \\
&= \frac{1}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot (X_{n,n} \beta + \varepsilon_n) + \beta_{d+1} M_{n,n}^\top \cdot \xi_n \right) \\
&= \frac{1}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot (X_{n,n} \beta + \varepsilon_n) \right) + \frac{n \cdot \beta_{d+1}}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n \right) \right) \quad (33)
\end{aligned}$$

with $M_{n,n}^\top$ from (15), $\beta = (\beta_1, \dots, \beta_d)^\top$, $\tilde{\beta} = (\beta_1, \dots, \beta_{d+1})^\top$, $\xi_n = (f_{d+1}(t_{n,1}), \dots, f_{d+1}(t_{n,n}))^\top$, $X_{n,n} := \begin{pmatrix} f_1(t_{n,1}) & \dots & f_d(t_{n,1}) \\ \vdots & \ddots & \vdots \\ f_1(t_{n,n}) & \dots & f_d(t_{n,n}) \end{pmatrix}$ and $\tilde{X}_{n,n} := \begin{pmatrix} f_1(t_{n,1}) & \dots & f_d(t_{n,1}) & f_{d+1}(t_{n,1}) \\ \vdots & \ddots & \vdots & \vdots \\ f_1(t_{n,n}) & \dots & f_d(t_{n,n}) & f_{d+1}(t_{n,n}) \end{pmatrix}$.

According to Theorem 10, $\frac{1}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot (X_{n,n} \beta + \varepsilon_n) \right) \xrightarrow{\mathcal{D}} B'$ holds true. It remains to determine whether and against what $\frac{n \cdot \beta_{d+1}}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n \right) \right)$ from equation (33) converges. The following lemma is a generalization of a result from Rabovski (2003). However, the proof given there is incorrect (the mappings u_n defined there on p. 20 are not necessarily continuous, cf. our Example 80 on page 42).

Lemma 14. Under the premises of Theorem 10 and if $f_{d+1} : [a, b] \rightarrow \mathbb{R}$ is of bounded variation,

$$\sup_{s \in [0, 1]} \left| T_{n-d} \circ g \left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n \right) \right) (s) - \Gamma(s) \right| \xrightarrow{n \rightarrow \infty} 0 \quad (34)$$

holds, where $\Gamma \in C[0, 1]$ with $\Gamma(s) := 0$ for $s \in [0, \delta)$ and otherwise

$$\begin{aligned} \Gamma(s) := & u(s) - u(\delta) \\ & + \int_{\delta}^s f(F_0^-(t))^\top \left(\int_{[0, t]} (ff^\top) \circ F_0^- d\lambda \right)^{-1} \left[\int_{(a, F_0^-(t))} (u \circ F_0) d\mu_f - f(F_0^-(t))u(t) \right] dt, \end{aligned}$$

applies, where $F_0 := F_{|[a, b]}$ and $u \in C[0, 1]$ is defined by $u(s) := \int_0^s f_{d+1} \circ F_0^- d\lambda$.

The proof can be found on page 41. Because of $\frac{n}{\sqrt{n-d}} \xrightarrow{n \rightarrow \infty} \infty$ and Lemma 14, $\frac{n \cdot \beta_{d+1}}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n \right) \right)$ literally seems to explode for $n \rightarrow \infty$ (see Figure 4).

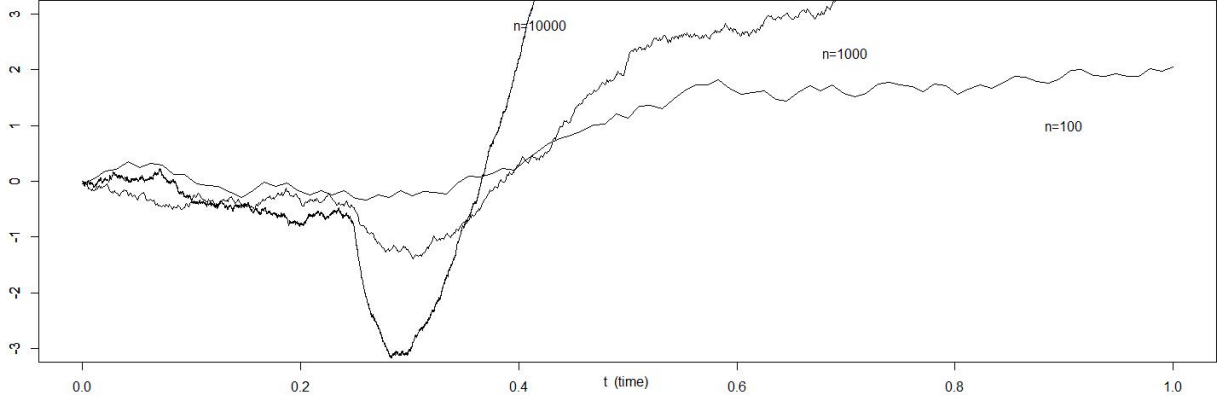


Figure 4: Single sample paths of B_n , for different n , where the observations in truth occurred under H_1 . $t_{n,i}$, $\varepsilon_{n,i}$, δ , and f as in Figure 3 (cf. R-script on p. 77).

Hence we need to re-parameterize the parameter space $\beta_{d+1}^* := \frac{\sqrt{n-d}}{n} \cdot \beta_{d+1}$ to obtain a r.v. whose limit distribution we can use for a test statistic. Therefore we define the linear model

$$Y_n^* = X_{n,n} \beta + \beta_{d+1}^* \xi_n + \varepsilon_n. \quad (35)$$

Note that the test problem (32) is equivalent to

$$H_0 : Y = \sum_{k=1}^d \beta_k f_k + \varepsilon \quad \text{vs.} \quad H_1 : Y = \sum_{k=1}^d \beta_k f_k + \beta_{d+1}^* f_{d+1} + \varepsilon, \quad \beta_{d+1}^* \neq 0 \quad (36)$$

for all n , where f_{d+1} is of bounded variation. With this we have the following

Theorem 15. Under the assumptions of Theorem 10 and Lemma 14, for B_n under the alternative H_1 given in (36), with respect to model (35), $B_n \xrightarrow{\mathcal{D}} B' + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma$ holds.

The proof can be found on page 42.

Example 16. Let $d = 1$, $[a, b] = [0, 1]$, $f_1(t) = 1$ and $t_{n,i} = \frac{i}{n}$. We consider the test problem

$$H_0 : Y(t) = \beta_1 + \varepsilon \quad \text{vs.} \quad H_1 : Y(t) = \beta_1 + \beta_2^* \cdot f_2(t) + \varepsilon, \quad \beta_2^* \neq 0. \quad (37)$$

We compute Γ according to Lemma 14 and obtain

$$\begin{aligned} \Gamma(s) &= u(s) - u(\delta) - \int_{\delta}^s \frac{1}{t} \cdot u(t) dt \\ &= \int_{\delta}^s \left(f_2(t) - \frac{1}{t} \int_0^t f_2(\tau) d\tau \right) dt, \quad s \geq \delta. \end{aligned} \quad (38)$$

For $f_2(t) = t$, $\Gamma(s) = \max(0, \frac{1}{4}(s^2 - \delta^2))$ follows. We simulate the partial sum process under the alternative H_1 , $\delta = 0.001$, $\varepsilon_{n,i} \sim \text{unif}([-\frac{1}{2}, \frac{1}{2}])$ and assuming $(\beta_1, \beta_2) = (0.5, 1)$ holds for the true parameters and compare with the limit process. Cf. Figure 5.

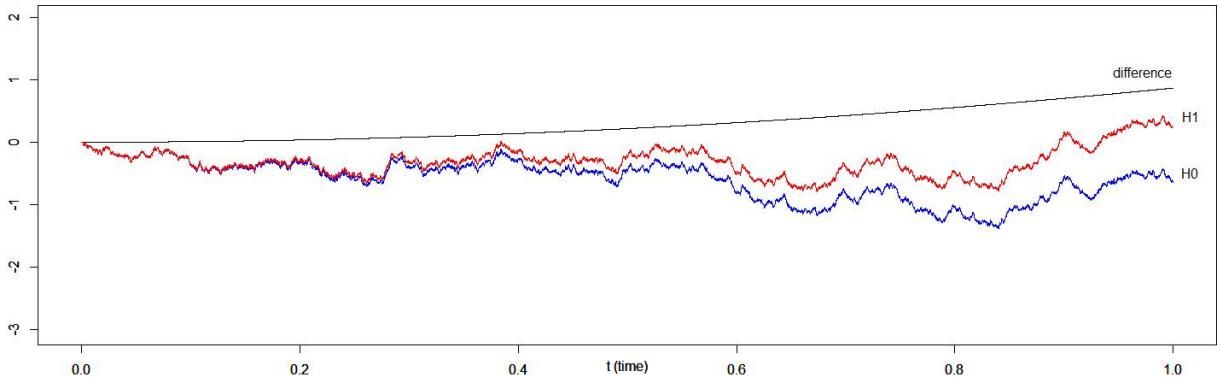


Figure 5: We compute a sample path of the partial sum process under H_0 (blue) and under H_1 (red), each with the same simulated error vector ε_n . The difference is the black path, which exactly fits Γ (cf. R-script on p. 77).

For the power $P(\text{reject } H_0 | H_1 \text{ is true})$ of the three tests from Theorem 11, under H_1 given in (36), we obtain

$$\begin{aligned} \text{power(29)} &= P \left(\sup_{0 \leq s \leq 1-\delta} \left(B_s + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma(s + \delta) \right) > x \right), \\ \text{power(30)} &= P \left(\inf_{0 \leq s \leq 1-\delta} \left(B_s + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma(s + \delta) \right) < y \right), \\ \text{power(31)} &= P \left(\sup_{0 \leq s \leq 1-\delta} \left| B_s + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma(s + \delta) \right| \geq z \right). \end{aligned} \quad (39)$$

Example 17. Following Bischoff (2016), we give a simple application from quality control in which we use (39) to determine suitable designs to detect change-points in the quality of a manufactured product. We consider the test problem of a constant product quality $Y(t)$ vs. a decreasing one, more precisely

$$H_0 : Y(t) = \beta_1 + \varepsilon \quad \text{vs.} \quad H_1 : Y(t) = \beta_1 + \beta_2^* \cdot g_{t_0}(t) + \varepsilon, \quad \beta_2^* < 0, \quad (40)$$

(Constant vs. step function)

where $t_0 \in (0, 1)$, $g_{t_0} : [0, 1] \rightarrow \mathbb{R}$, $g_{t_0}(t) := 1_{(t_0, 1]}(t)$. Let $(t_{n,1}, \dots, t_{n,n})$ be an asymptotic F -design in $[0, 1]$ with $F(t_0) =: q \in (\delta, 1)$. We call this an asymptotic q - F -design. First, we determine $\Gamma = \Gamma_{g_{t_0}, F}$ from Lemma 14. Note $g_{t_0} \circ F_0^- = 1_{(q, 1]}$, because $r \leq q$ implies $F_0^-(r) \leq t_0$ and $r > q$ implies $F_0^-(r) > t_0$. With some short calculations we get (note $\delta < q$)

$$u(s) = \int_{[0, s]} g_{t_0} \circ F_0^- d\lambda = (s - q) \cdot 1_{(q, 1]}(s), \quad s \in [0, 1]$$

and

$$\begin{aligned} \Gamma(s) &= u(s) - u(\delta) + \int_{[\delta, s]} \frac{1}{t} [0 - u(t)] dt \\ &= q \cdot \ln\left(\frac{s}{q}\right) \cdot 1_{(q, 1]}(s), \quad s \in [0, 1]. \end{aligned}$$

With respect to power(30) given in (39), we call an asymptotic q_1 - F -design uniformly better than an asymptotic q_2 - F -design if $\beta_2^* < 0$ and

$$\beta_2^* q_1 \ln\left(\frac{s}{q_1}\right) \cdot 1_{(q_1, 1]}(s) \leq \beta_2^* q_2 \ln\left(\frac{s}{q_2}\right) \cdot 1_{(q_2, 1]}(s) \quad (41)$$

holds for all $s \in [\delta, 1]$. With this we have the following statement.

(U) For $\beta_2^* < 0$ and $e^{-1} \leq q_1 < q_2$, an asymptotic q_1 - F -design is uniformly better than an asymptotic q_2 - F -design.

Proof of (U): For $s \leq q_1$ and $q_1 < s \leq q_2$ the assertion is clear. For $q_2 < s \leq 1$ we note

$$\begin{aligned} &\beta_2^* q_1 \ln\left(\frac{s}{q_1}\right) \cdot 1_{(q_1, 1]}(s) \leq \beta_2^* q_2 \ln\left(\frac{s}{q_2}\right) \cdot 1_{(q_2, 1]}(s) \\ \Leftrightarrow &\beta_2^* q_1 \ln\left(\frac{s}{q_1}\right) \leq \beta_2^* q_2 \ln\left(\frac{s}{q_2}\right) \\ \Leftrightarrow &q_1 \ln\left(\frac{q_1}{s}\right) \leq q_2 \ln\left(\frac{q_2}{s}\right). \end{aligned} \quad (42)$$

For $e^{-1} \leq q < 1$ we consider $f(q) := q \ln\left(\frac{q}{s}\right)$. Obviously, $f'(q) = \ln\left(\frac{q}{s}\right) + 1 \geq 0$ and thus f is strictly monotonically increasing. Consequently, (42) and thus overall (41) holds. \square

Least-squares residual partial sum process

The technique of the factorization of the regression functions used in the proof of Theorem 10 (cf. Lemma 67) can of course be profitably used in the study of other stochastic processes. The closest example to the recursive residual partial sum processes we considered is the ordinary least squares residual partial sum process. In analogy to Definition 4, we thus come to

Definition 18. *The ordinary least-squares residual partial sum process is defined as*

$$D_n := \frac{1}{\sigma\sqrt{n}} \cdot T_n \circ w_n, \quad (43)$$

where σ is defined in (4) and w_n is defined in (12).

This leads to a generalization of Theorem 2.2 from Bischoff (1998), which can also be found as Theorem 2.2.11 in Heindl (2022), since it was developed jointly by Heindl and Evers.

Theorem 19. *Let $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$ be an asymptotic F -design in $[a, b]$ and consider model (3) with $f_1, \dots, f_d : [a, b] \rightarrow \mathbb{R}$ all of bounded variation and left continuous and let f_1, \dots, f_d be linearly independent in $L_2([a, b], F)$. Under these assumptions and H_0 given in (32), $D_n \xrightarrow{\mathcal{D}} B_{f,F}$, where $B_{f,F}$ is the Gaussian process on $[0, 1]$ given by*

$$B_{f,F}(\omega, t) := B(\omega, t) + \left(\int_{(0,t)} f \circ F_0^- d\lambda \right)^\top \cdot \left(\int_{[a,b]} f f^\top d\mu_F \right)^{-1} \cdot \left(\int_{[a,b]} B(\omega, F) d\mu_f - B(\omega, 1) f(F^-(1)) \right), \quad (44)$$

where B is a Brownian motion on $[0, 1]$.

The proof can be found on page 43. Note that the proof is based on techniques developed for the proof of Theorem 79.

6 Reversed partial sums of recursive residuals

Definition 20. *The reversed recursive residual partial sum process is defined as*

$$B_{n,\delta}^{(b)} := \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \circ r_n, \quad (45)$$

where σ is defined in (4) and $g = g_\delta : \mathbb{R}^{n-d} \rightarrow \mathbb{R}^{n-d}$ is defined by

$$a = (a_1, \dots, a_{n-d})^\top \mapsto (0, \dots, 0, a_{n-d}, \dots, a_{\lfloor (n-d)\delta \rfloor + 1})^\top. \quad (46)$$

Theorem 21. *Under the assumptions of Theorem 10, $B_{n,\delta}^{(b)} \xrightarrow{\mathcal{D}} B'$ holds under H_0 from (36), whereby B' is a shifted Brownian motion (cf. Remark 5 for the definition of B').*

The proof can be found on page 45. As with B_n , for abbreviation we write $B_n^{(b)}$ instead of $B_{n,\delta}^{(b)}$.

Theorem 22. *Under the assumptions of Theorem 21, an asymptotic test of size α is given by each of the following rules.*

$$\text{reject } H_0 \text{ given in (32)} \quad :\Leftrightarrow \quad \exists s \in [0, 1] \text{ with } \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d}(g \circ r_n)(s) > x, \quad (47)$$

$$\text{reject } H_0 \text{ given in (32)} \quad :\Leftrightarrow \quad \exists s \in [0, 1] \text{ with } \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d}(g \circ r_n)(s) < y, \quad (48)$$

$$\text{reject } H_0 \text{ given in (32)} \quad :\Leftrightarrow \quad \exists s \in [0, 1] \text{ with } \left| \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d}(g \circ r_n)(s) \right| \geq z, \quad (49)$$

where x , y and z are defined as in Theorem 11, but the mapping g is given here by (46).

The proof is the same as that already given for Theorem 11.

Theorem 23. For $B_n^{(b)}$ under the alternative H_1 given in (36), with respect to model (35),

$$B_n^{(b)} \xrightarrow{\mathcal{D}} B' + \frac{\beta_{d+1}}{\sigma} \cdot \Delta \quad (\text{for } n \rightarrow \infty) \quad (50)$$

holds under the assumptions of Theorem 10, with $\Delta \in C[0, 1]$, $\Delta(s) := 0$ for $s \in [0, \delta)$ and

$$\begin{aligned} \Delta(s) := & u(1) - u(1 - s + \delta) \\ & + \int_{1-s+\delta}^1 f(F_0^-(t))^\top \left(\int_{[0,t]} (ff^\top) \circ F_0^- d\lambda \right)^{-1} \left[\int_{(a, F_0^-(t))} (u \circ F_0) d\mu_f - f(F_0^-(t))u(t) \right] dt, \end{aligned} \quad (51)$$

applies, where $u \in C[0, 1]$ is defined by $u(s) := \int_0^s f_{d+1} \circ F_0^- d\lambda$ and $F_0 := F_{[a,b]}$.

The proof can be found on page 46.

Example 24. We again consider Example 16 and compute the reversed recursive residual partial sum process under H_0 and H_1 , cf. Figure 6. For the trend function Δ we get

$$\Delta(s) = u(1) - u(1 - s + \delta) - \int_{1-s+\delta}^1 \frac{1}{t} \cdot u(t) dt = \frac{1}{4}(s - \delta)(2 + \delta - s), \quad s \geq \delta.$$

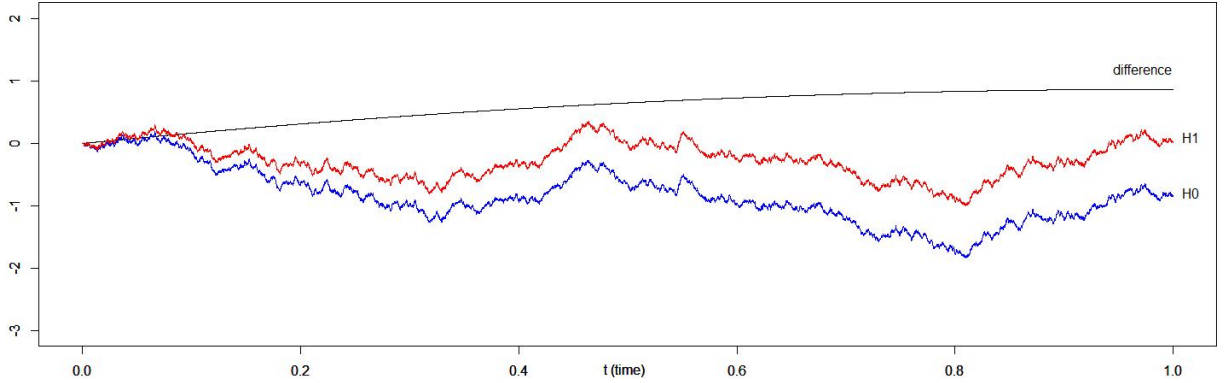


Figure 6: Sample path of the reversed partial sum process under H_0 (blue) and under H_1 (red), each with the same simulated error vector ε_n . The difference (black path) fits Δ .

For the power of the tests (47), (48), and (49) under H_1 given in (36), we obtain, completely analogous to the “normal” recursive residual partial sum process,

$$\begin{aligned} \text{power(47)} &= P \left(\sup_{0 \leq s \leq 1-\delta} \left(B_s + \frac{\beta_{d+1}}{\sigma} \cdot \Delta(s + \delta) \right) > x \right), \\ \text{power(48)} &= P \left(\inf_{0 \leq s \leq 1-\delta} \left(B_s + \frac{\beta_{d+1}}{\sigma} \cdot \Delta(s + \delta) \right) < y \right), \\ \text{power(49)} &= P \left(\sup_{0 \leq s \leq 1-\delta} \left| B_s + \frac{\beta_{d+1}}{\sigma} \cdot \Delta(s + \delta) \right| \geq z \right). \end{aligned}$$

Example 25. We follow Bischoff et al. (2005) and consider the particular importend test problem

$$H_0 : Y(t) = \beta_1 + \varepsilon \quad \text{vs.} \quad H_1 : Y(t) = \beta_1 + \beta_2^* \cdot g(t) + \varepsilon, \quad \beta_2^* > 0 \quad (52)$$

(Constant vs. non-constant)

with $g : [a, b] \rightarrow \mathbb{R}$ of bounded variation and we want to determine Γ of Theorem 15 (cf. Lemma 14) and Δ of Theorem 23. In both cases $u(s) = \int_0^s g \circ F_0^-(t) dt$, where $F_0 := F_{|[a,b]}$. Thus

$$\Gamma(s) = \int_{\delta}^s \left(g \circ F_0^-(t) - \frac{1}{t} \cdot \int_0^t g \circ F_0^-(\tau) d\tau \right) dt \quad (53)$$

$$\Delta(s) = \int_{1-s+\delta}^1 \left(g \circ F_0^-(t) - \frac{1}{t} \cdot \int_0^t g \circ F_0^-(\tau) d\tau \right) dt, \quad (54)$$

with $\Gamma(s) = \Delta(s) = 0$ for all $s \in [0, \delta]$. To compare the power of the associated tests, it would be helpful if we could show $\Gamma(s) \leq \Delta(s)$ for all s (or vice versa). To do this, we set $f := g \circ F^-$ and analyze the integrand $I(t) := f(t) - \frac{1}{t} \cdot \int_0^t f(\tau) d\tau$ from (53) and (54), respectively. Because of $1 - (1 - s + \delta) = s - \delta$, it would suffice if we specify conditions under which I is nondecreasing. If f is nondecreasing and $t < t'$ holds,

$$\begin{aligned} \frac{1}{t'} \cdot \int_0^{t'} f(\tau) d\tau - \frac{1}{t} \cdot \int_0^t f(\tau) d\tau &\leq \frac{1}{t'} \left(\int_0^t f(\tau) d\tau + (t' - t)f(t') \right) - \frac{1}{t} \cdot \int_0^t f(\tau) d\tau \\ &\leq \left(\frac{1}{t'} - \frac{1}{t} \right) \int_0^t f(\tau) d\tau + \left(1 - \frac{t}{t'} \right) f(t') \\ &\leq \left(\frac{1}{t'} - \frac{1}{t} \right) \cdot t \cdot f(t) + \left(1 - \frac{t}{t'} \right) f(t') \\ &= \left(1 - \frac{t}{t'} \right) \cdot (f(t') - f(t)) \leq f(t') - f(t) \end{aligned}$$

follows. Hence $I(t) = f(t) - \frac{1}{t} \cdot \int_0^t f(\tau) d\tau$ would be non-decreasing. Thus

$$\forall s \in [\delta, 1] : \Gamma(s) \leq \Delta(s) \quad (55)$$

holds if $g : [a, b] \rightarrow \mathbb{R}$ is non-decreasing (since $f = g \circ F^-$ is non-decreasing as well and $1 - (1 - s + \delta) = s - \delta$). So, if we consider test problem (52) for a non-decreasing g , we can either use test (29) based on process (21), or test (47) based on process (45). Since

$$\begin{aligned} \text{power}(47) &= P \left(\sup_{0 \leq s \leq 1-\delta} \left(B_s + \frac{\beta_2}{\sigma} \cdot \Delta(s + \delta) \right) > x \right) \\ &\stackrel{(55)}{\geq} P \left(\sup_{0 \leq s \leq 1-\delta} \left(B_s + \frac{\beta_2}{\sigma} \cdot \Gamma(s + \delta) \right) > x \right) = \text{power}(29), \quad (56) \end{aligned}$$

the test has a larger power when based on reversed sums of recursive residuals. Furthermore, $\Gamma : [0, 1] \rightarrow \mathbb{R}$ is non-decreasing and convex and $\Delta : [0, 1] \rightarrow \mathbb{R}$ is non-decreasing and concave, if g is non-decreasing. This follows from

$$\Gamma'(s) = I(s) \quad \text{and} \quad \Delta'(s) = I(1 - s + \delta)$$

and because $s \mapsto I(s)$ is non-decreasing and $s \mapsto I(1 - s + \delta)$ is non-increasing. Moreover, $\Gamma(1) = \Delta(1)$ and $\Gamma(s) = \Delta(s)$ holds for all $s \in [0, \delta]$.

7 An asymptotically uniformly most powerful test

For technical reasons, in this section we do not consider the processes B_n , $B_n^{(b)}$ and B' (defined in (21), (45) resp. Remark 5), but we shift them to the left by δ and cut off the uninteresting part at the beginning, cf. (24) and the proof of Theorem 21).

Definition 26. For $X : \Omega \rightarrow \mathbb{R}^{[0,1]}$, $u : [0, 1] \rightarrow \mathbb{R}$ and $\delta := F(c)$ we define

$$\begin{aligned} X^\bullet : \Omega &\rightarrow C[0, 1 - \delta], & X^\bullet(\omega)(t) &:= X(\omega)(t + \delta) \\ u^\bullet : [0, 1 - \delta] &\rightarrow \mathbb{R}, & u^\bullet(t) &:= u(t + \delta). \end{aligned}$$

Note that B'^\bullet is an ordinary Brownian motion on $[0, 1 - \delta]$ (the dash at B'^\bullet is not to be confused with the derivative). It is obvious that all convergence results remain valid for the transformed processes, although one has to shift functions like Γ and Δ at appropriate places. For example, under the alternative H_1 from (36) for B_n^\bullet and $B_n^{(b)\bullet}$ with respect to model (35),

$$B_n^\bullet \xrightarrow{\mathcal{D}} B'^\bullet + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma^\bullet \quad \text{and} \quad B_n^{(b)\bullet} \xrightarrow{\mathcal{D}} B'^\bullet + \frac{\beta_{d+1}}{\sigma} \cdot \Delta^\bullet$$

holds. In the following we consider the one-sided test problem

$$H_0 : Y = \sum_{k=1}^d \beta_k f_k + \varepsilon \quad \text{vs.} \quad H_1 : Y = \sum_{k=1}^d \beta_k f_k + \beta_{d+1}^* f_{d+1} + \varepsilon, \quad \beta_{d+1}^* > 0, \quad (57)$$

with respect to model (35)

Definition 27. We call a sequence of functions $\varphi_n : C[0, 1 - \delta] \rightarrow \{0, 1\}$, $n \in \mathbb{N}^{\geq 1}$ an asymptotically uniformly most powerful size α test for recursive residuals, resp. reversed recursive residuals, with respect to model (35) if

$$\varphi_n \left(\frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \circ (M_{n,n}^\top Y_n^*) \right)^\bullet \xrightarrow{\mathcal{D}} \varphi \left(B'^\bullet + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma^\bullet \right) \quad (58)$$

$$\text{resp.} \quad \varphi_n \left(\frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \circ (M_{n,n}^\top Y_n^*) \right)^\bullet \xrightarrow{\mathcal{D}} \varphi \left(B'^\bullet + \frac{\beta_{d+1}}{\sigma} \cdot \Delta^\bullet \right), \quad (59)$$

where the g in (58) comes from (22) and that in (59) comes from (46) and $\varphi : C[0, 1 - \delta] \rightarrow \{0, 1\}$ is a uniformly most powerful size α test for the test problem (57) observing $B'^\bullet + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma^\bullet$, resp. observing $B'^\bullet + \frac{\beta_{d+1}}{\sigma} \cdot \Delta^\bullet$.

Theorem 28. We define the test statistic $\rho : C[0, 1 - \delta] \rightarrow \mathbb{R}$ with

$$\rho(u) := \left(- \int_{[0, 1-\delta]} \Delta^\bullet(s) d(\Delta^\bullet)'(t) \right)^{-\frac{1}{2}} \int_{[0, 1-\delta]} u(t) d(\Delta^\bullet)'(t) \quad (60)$$

and the statistical test $\varphi := 1_{\rho^{-1}((\Phi^{-1}(1-\alpha), \infty))}$ with the decision rule

$$\text{reject } H_0 \text{ given in (57)} \quad :\Leftrightarrow \quad \varphi \left(\left(\frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \circ (M_{n,n}^\top Y_n^*) \right)^\bullet \right) = 1, \quad (61)$$

where $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$ and g in (61) comes from (46) and the dashes $'$ in (60) denote the derivative. Under the assumptions of Theorem 10 and Lemma 14, whereby we additionally assume that $f_{d+1}(x) = 0$ holds for all $x \in [a, F_0^-(\delta)]$, φ ($= \varphi_n$ for all n) is an asymptotically uniformly most powerful size α test for (57) with respect to model (35).

The proof can be found on page 46.

Part B. Weak convergence of finite measures

Here we describe the main principles of the theory of weak convergence of finite measures. The results in this section are, in principle, facts that have been known for almost 60 years. Therefore, to make the presentation more interesting, we present these classical results in a more general way than usual. On the one hand, we let the underlying spaces very general and consider measures on perfectly normal spaces instead on metric spaces and on the other hand, we consider filters of measures instead of sequences of measures.

It is sometimes like this: One considers certain structures on topological spaces (e.g. measures) and has a notion of convergence at hand. Nevertheless, one often considers only sequences of these structures (and correspondingly the convergence of sequences). However, if one considers the set of all these structures, one can sometimes naturally define a topology that respects convergence for sequences of these structures, which then gives a filter convergence anyway. If one then also finds that the theory for filters is basically no more difficult to develop than for sequences, then there is really no reason to restrict oneself to sequences and one can develop the notion of convergence directly for filters.

In the literature one can find presentations of the theory of weak convergence of measures using nets, cf. Topsoe (1970) and Bogachev (2018), but although nets and filters are both sufficient structures to describe topological properties, we think that filters are better and more natural to study problems concerning convergence. A net is defined as a function whose domain is a directed set I . Therefore one has to deal with indexed values all the time. Filters are defined much more directly and are virtually their own directed set. Although there is in principle a canonical correspondence between filters and nets, many arguments and concepts can be formulated more simply and naturally with filters. Finally, general convergence theory is and has been developed mainly in the language of filters, see, e.g., the monographs Bourbaki (1966), Binz (1975), Beattie and Butzmann (2002), Preuss (2002), Hart et al. (2004), Mynard and Pearl (2009), Dolecki and Mynard (2016), Nel (2016) and filters are better studied than nets with respect to their set-theoretic and algebraic properties and resulting applications in topological problems, cf. e.g. Comfort and Negrepointis (1974), Walker (1974), Kunen and Vaughan (1984), Jech (2003), Zelenyuk (2011), Hindman and Strauss (2012), to name just a few of the best-known books in this field. In fact, the author is not aware of any work in which nets are studied in a similar depth as filters in these books.

This generalization of the theory of weak convergence of measures for filters is of no direct relevance for the understanding of Part A of this thesis. Rather, it has arisen only as a by-product of the author's familiarization with the subject. However, since it is not found in this form in the literature, we decided to include it as a separate part.

The classical theory of weak convergence necessary to understand Part A for sequences of measures in the context of metric spaces is developed, for example, in Billingsley (1968, 1999), Kallenberg (2002), Borovkov (2013), Shorack (2017), and Klenke (2020).

1 Basic notations and definitions

Definition 29. We call a subset $\varphi \subseteq \mathcal{P}(X)$ of the powerset of a set X filter (on X), if

1. $\emptyset \notin \varphi$ and $X \in \varphi$
2. $P, Q \in \varphi$ implies $P \cap Q \in \varphi$
3. $P \in \varphi$ and $P \subseteq Q \subseteq X$ implies $Q \in \varphi$.

Filter that are maximal with respect to subset relation are called ultrafilter. Let $\mathcal{F}(X)$ (resp. $\mathcal{F}_U(X)$) the set of all filters (resp. ultrafilters) on X . $\varphi_0 \subseteq \varphi$ is called filterbase or base (for φ) if for all $P \in \varphi$ exists some $Q \in \varphi_0$ with $Q \subseteq P$. A map $f : X \rightarrow Y$ has a natural extension $F : \mathcal{F}(X) \rightarrow \mathcal{F}(Y)$ (resp. $F : \mathcal{F}_U(X) \rightarrow \mathcal{F}_U(Y)$) defined by

$$F(\varphi) := \{Q \subseteq Y \mid \exists P \in \varphi \text{ with } f(P) \subseteq Q\}. \quad (62)$$

For $x \in X$ we define $\dot{x} := \{P \subseteq X \mid x \in P\}$. If we identify each $x \in X$ with \dot{x} , then F is an extension of X to all of $\mathcal{F}(X)$. If $\alpha \subseteq \mathcal{P}(X)$ we define $[\alpha] := \{P \subseteq X \mid \exists A \in \alpha \text{ with } A \subseteq P\}$.

Definition 30. We call $\tau \subseteq \mathcal{P}(E)$ a topology on E if

1. $\emptyset, E \in \tau$
2. $U, V \in \tau$ implies $U \cap V \in \tau$
3. $\tau' \subseteq \tau$ implies $\bigcup \tau' \in \tau$

The pair (E, τ) is called topological space. Elements of τ are called open and their complements are called closed. To say that a filter φ on E converges to $a \in E$, denoted by $\varphi \xrightarrow{\tau} a$, means that for all open U with $a \in U$ exists $P \in \varphi$ with $P \subseteq U$.

A filter contain those subsets that are sufficiently large in some sense. For example let $x = (x_n)_{n \in \mathbb{N}}$ be a sequence in some set X and define $\varphi(x)$ by

$$\varphi(x) := \{P \subseteq X \mid \exists n \in \mathbb{N} \forall k \geq n : x_k \in P\}.$$

We can think of φ as the set of end-pieces of $(x_n)_{n \in \mathbb{N}}$. Note that $\varphi : X^{\mathbb{N}} \rightarrow \mathcal{F}(X)$ is not necessarily injective. For example $1, \frac{1}{2}, \frac{1}{3}, \frac{1}{4}, \frac{1}{5}, \frac{1}{6}, \dots$ and $\frac{1}{2}, 1, \frac{1}{4}, \frac{1}{3}, \frac{1}{6}, \frac{1}{5}, \dots$ are two sequences $x = (x_n)_{n \in \mathbb{N}}$ and $y = (y_n)_{n \in \mathbb{N}}$ with $x_n \neq y_n$ for all $n \in \mathbb{N}$, but $\varphi(x) = \varphi(y)$. Nevertheless all topological properties of $x = (x_n)_{n \in \mathbb{N}}$ can be reformulated in terms of $\varphi(x)$. For example

$$x_n \text{ converges to } a \iff \{U \subseteq X \mid U \text{ is a neighbourhood of } a\} \subseteq \varphi(x) \iff \varphi(x) \xrightarrow{\tau} a$$

Filters were introduced by Cartan (1937) and brought to popularity by Bourbaki. For more details concerning general topological spaces and filters cf. Bourbaki (1966).

Definition 31. We call $(\Omega, \mathcal{A}, \mu)$ measure space if

1. Ω is a nonempty set
2. \mathcal{A} is a σ -algebra on Ω , that mean (a) $\Omega \in \mathcal{A}$, (b) $\Omega \setminus A \in \mathcal{A}$ whenever $A \in \mathcal{A}$ and (c) $\bigcup_{i \in \mathbb{N}} A_i \in \mathcal{A}$ whenever $A_i \in \mathcal{A}$ for all $i \in \mathbb{N}$

3. $\mu : \mathcal{A} \rightarrow [0, \infty) \cup \{\infty\}$ is a measure on (Ω, \mathcal{A}) , that mean (a) $\mu(\emptyset) = 0$ and (b)

$$\mu \left(\bigcup_{i \in \mathbb{N}} A_i \right) = \sum_{i \in \mathbb{N}} \mu(A_i)$$

whenever $A_i \in \mathcal{A}$, $i \in \mathbb{N}$ are pairwise disjoint sets.

For ∞ and $-\infty$ the usual properties apply (for example $a + \infty = \infty$, $\infty + \infty = \infty$ and so on). A measure space with $\mu(\Omega) < \infty$ is called finite measure space. If $\mu(\Omega) = 1$, $(\Omega, \mathcal{A}, \mu)$ is called probability space. In this case μ is called probability measure.

Definition 32. Let $\alpha \subseteq \mathcal{P}(\Omega)$. We call

$$\sigma(\alpha) := \bigcap_{\substack{\alpha \subseteq \mathcal{A} \subseteq \mathcal{P}(\Omega) \\ \mathcal{A} : \sigma\text{-algebra}}} \mathcal{A} \quad (63)$$

the σ -algebra generated by α . If a topology τ is given on Ω , we call $\sigma(\tau)$ the Borel σ -algebra.

For basics from measure and integration theory in general and probability theory in particular, we refer to popular textbooks on these topics, such as Bauer (2001), Kallenberg (2002), Shorack (2017), Elstrodt (2018), and Klenke (2020).

2 Finite measures in perfectly normal spaces

Definition 33. A closed subset B of a topological space is called G_δ -set if there exists open sets O_n , $n \in \mathbb{N}$ with $B = \bigcap_{n \in \mathbb{N}} O_n$. A topological space E is said to be a G_δ -space if every closed subspace B of E is a G_δ -set. A topological space E is said to be normal if for disjoint, closed A, B exists disjoint, open U, V with $A \subseteq U$ and $B \subseteq V$. A normal G_δ -space is called perfectly normal.

Lemma 34. For a topological space X the following is equivalent.

1. X is perfectly normal.
2. For all closed A in X exists a continuous $f : X \rightarrow [0, 1]$ with $A = f^{-1}(0)$.
3. For all closed, disjoint A, B in X exists a continuous $f : X \rightarrow [0, 1]$ with $A = f^{-1}(0)$ and $B = f^{-1}(1)$.

The proof can be found on page 48. All metrizable spaces are perfectly normal (but not vice versa¹) and all CW-complexes are perfectly normal (of course not vice versa), cf. Lundell and Weingram (1969), Proposition 4.3. But there exists nice spaces (e.g. hereditarily normal² compact Hausdorff spaces) that are not perfectly normal³. Nonetheless, in many cases

¹There is a prominent example called Sorgenfrey line: (\mathbb{R}, τ) , with $\tau := \{\bigcup \mathcal{B}' \mid \mathcal{B}' \subseteq \mathcal{B}\}$ and $\mathcal{B} := \{[a, b) \mid a, b \in \mathbb{R}\}$, cf. Bourbaki (1966), Part 2, Ch. IX. §4 Exercise 8.

²A space X is called hereditarily normal iff every subspace is normal. In fact this is equivalent to the condition: for all $A, B \subseteq X$ with $\bar{A} \cap B = \emptyset = A \cap \bar{B}$ exists open and disjoint U, V with $A \subseteq U$ and $B \subseteq V$. The proof of this statement is not important for us and at the same time very simple and therefore omitted.

³Let X uncountable, $x \in X$, $\tau := \{U \subseteq X \mid x \notin U \text{ or } X \setminus U \text{ is finite}\}$. Then (X, τ) is not perfectly normal, since $\{x\}$ is closed but not a G_δ -set but (X, τ) is hereditarily normal (for $\bar{A} \cap B = \emptyset = A \cap \bar{B}$ exists open and disjoint U, V with $A \subseteq U$ and $B \subseteq V$).

perfectly normal is exactly the property we need to prove a theorem in this context of weak convergence. This is mainly due to the following Lemma 35 (there especially the second point), which in combination with the dominated convergence theorem is crucial for the proof of the most basic results (cf. Lemma 36, Lemma 38 and Theorem 40).

Lemma 35.

1. Let (E, d) be a metric space, F a closed subspace and $\varepsilon > 0$. Define

$$h : [0, \infty) \rightarrow [0, 1], \quad h(x) := \begin{cases} -\varepsilon^{-1}x + 1 & \text{if } x \in [0, \varepsilon] \\ 0 & \text{if } x > \varepsilon \end{cases}$$

$$f_{F, \varepsilon} : E \rightarrow [0, 1], \quad f_{F, \varepsilon}(x) := h(d(x, F))$$

Then $f_{F, \varepsilon}$ has the following properties:

- a) $0 \leq f_{F, \varepsilon}(x) \leq 1$ for all $x \in E$
 - b) $f_{F, \varepsilon}(x) = 1$ for all $x \in F$
 - c) $f_{F, \varepsilon}(x) = 0$ for all $x \in E$ with $d(x, F) \geq \varepsilon$
 - d) $f_{F, \varepsilon}$ is $\frac{1}{\varepsilon}$ -Lipschitz-continuous
2. Let (E, τ) be normal and F a closed G_δ -set in E . Then there exists two sequences $(B_n)_{n \in \mathbb{N}}$ in τ and $(f_n)_{n \in \mathbb{N}}$ in $C(E, [0, 1])$ with

$$\bigcap_{n \in \mathbb{N}} B_n = F, \quad \overline{B_{n+1}} \subseteq B_n, \quad f_n(\overline{B_{n+1}}) \subseteq \{1\}, \quad f_n(E \setminus B_n) \subseteq \{0\}$$

and therefore $f_n(x) \xrightarrow{\tau_{\mathbb{R}}} 1_F(x)$ for $n \rightarrow \infty$ and for all $x \in E$.

The proof can be found on page 49. The aim of the next Lemma is to show that in a perfectly normal space a finite measure is determined by its values for closed and open sets.

Lemma 36. Let (E, τ) be perfectly normal and P a finite measure defined on $\mathcal{B}(E)$, the Borel σ -algebra. Then

$$\forall A \in \mathcal{B}(E) \forall \varepsilon > 0 \exists F : \text{closed}, G : \text{open}, \text{ with } F \subseteq A \subseteq G \text{ and } P(G \setminus F) < \varepsilon.$$

This means P is regular.

The proof can be found on page 49.

3 Weak convergence of finite measures

Intuitively, one would say of a series $(\mu_n)_{n \in \mathbb{N}}$ of measures to converge against μ , if

$$\mu_n(A) \rightarrow \mu(A) \quad \text{as } n \rightarrow \infty, \quad \text{for every } A \text{ from the related } \sigma\text{-algebra} \quad (64)$$

holds. On the other hand if μ_n is the Dirac measure at $\frac{1}{n}$ (In the measuring space $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$), one would intuitively expect that the sequence converges to μ , the Dirac measure in 0. But for $A = (-\infty, 0]$ the above requirement for the convergence of the measures is violated, because

$$\lim_{n \rightarrow \infty} \mu_n(A) = 0 \neq \mu(A).$$

This concept of convergence seems to be too strong. An equivalent formulation of the above, intuitive convergence concept of measures is (cf. Elstrodt (2018), VIII exercise 4.3):

$$\lim_{n \rightarrow \infty} \int_{\Omega} f d\mu_n = \int_{\Omega} f d\mu, \quad \text{for all bounded and measurable } f. \quad (65)$$

The proof of the equivalence of (65) and (64) can be found on page 50.

On the basis of this characterization we seek for weaker functional classes \mathcal{F} and sets of measures \mathcal{M} , so that the above equation (65) holds for these measures and these classes of functions. Roughly speaking the average value of sufficiently nice functions should converge. Furthermore we want \mathcal{F} to be a separating family for \mathcal{M} . That means

$$\int_{\Omega} f d\mu = \int_{\Omega} f d\nu \quad \text{for all } f \in \mathcal{F} \implies \mu = \nu$$

for all $\mu, \nu \in \mathcal{M}$. This guarantees the uniqueness of the limit value. If one chooses the finite measures on the borel σ -algebra as \mathcal{M} and the bounded continuous functions as \mathcal{F} , one obtains the weak convergence described below. It will turn out that this kind of convergence is exactly the kind of convergence we encounter in the central limit theorem. Another choice of function classes and sets of measures provides, for example, the vague convergence (which will not be discussed here).

Let (E, τ) be a topological space, $M(E) := \{p : \mathcal{B}(E) \rightarrow \mathbb{R} \mid p \text{ is a finite measure}\}$ where $\mathcal{B}(E)$ is the Borel σ -algebra. Let τ_w be the coarsest topology on $M(E)$ such that all mappings

$$\alpha_f : M(E) \rightarrow \mathbb{R}, \quad p \mapsto \int_E f dp, \quad f \in C(E, \mathbb{R}) \cap B(E, \mathbb{R}),$$

continuous, where $C(E, \mathbb{R})$ are continuous and $B(E, \mathbb{R})$ are bounded functions. In other words: τ_w is the initial topology on $M(E)$, with respect to $\{\alpha_f \mid f \in C(E, \mathbb{R}) \cap B(E, \mathbb{R})\}$. If E is a metrizable space, that is separable and complete (as metric space), then this topology can even be defined by a metric, the Prokhorov metric. We study here only the topological case. It is widely known that a topology is completely characterized by its convergent filter. Also well known is that

$$\psi \xrightarrow{\tau_w} p \iff \forall f \in C(E, \mathbb{R}) \cap B(E, \mathbb{R}) : \alpha_f(\psi) \xrightarrow{\tau_{\mathbb{R}}} \alpha_f(p) \quad (66)$$

holds for any filter ψ on $M(E)$ with respect to τ_w , cf. Bourbaki (1966), Part 1, Chap. I. §7.6 Proposition 10. However, in the next definition we will use the more intuitive notation $\int_E f d\psi$ instead of $\alpha_f(\psi)$, i.e. we set $\int_E f d\psi := \alpha_f(\psi)$ from now on.

Definition 37. *The initial topology τ_w on $M(E)$, with respect to $\{\alpha_f \mid f \in C(E, \mathbb{R}) \cap B(E, \mathbb{R})\}$ is called topology of weak convergence on $M(E)$. Let ψ be a filter on $M(E)$. We say ψ converges weakly to $p \in M(E)$ iff*

$$\int_E f d\psi \xrightarrow{\tau_{\mathbb{R}}} \int_E f dp \quad \text{for all } f \in C(E, \mathbb{R}) \cap B(E, \mathbb{R}).$$

In this case we write $\psi \xrightarrow{w} p$ and call $\int_E f dp$ weak limit of ψ . A sequence $(q_n)_{n \in \mathbb{N}}$ in $M(E)$ converges weakly to $p \in M(E)$ iff the filter induced by $(q_n)_{n \in \mathbb{N}}$ converges weakly to p , i.e.

$$\int_E f dq_n \xrightarrow[n \rightarrow \infty]{\tau_{\mathbb{R}}} \int_E f dp \quad \text{for all } f \in C(E, \mathbb{R}) \cap B(E, \mathbb{R}).$$

In this case we write $(q_n)_{n \in \mathbb{N}} \xrightarrow{w} p$ or simply $q_n \xrightarrow{w} p$.

According to (66), $\psi \xrightarrow{\tau_w} p$ and $\psi \xrightarrow{w} p$ are two equivalent formulations. One more remark: If $p \in M' \subseteq M(E)$ and ψ is a filter on M' then obviously

$$\psi \xrightarrow{\text{in } M'} p \iff i(\psi) \xrightarrow{\text{in } M(E)} p \quad \text{for inclusion } i: M' \hookrightarrow M(E).$$

Therefore we will not distinguish between ψ and $i(\psi)$.

Lemma 38. *The weak limit of filters ψ on $M(E)$ is unique if (E, τ) is perfectly normal*

$$\psi \xrightarrow{w} P \quad \text{and} \quad \psi \xrightarrow{w} Q \quad \text{implies} \quad P = Q$$

In other words $(M(E), \tau_w)$ is Hausdorff provided (E, τ) is perfectly normal⁴.

The proof can be found on page 50.

Lemma 39. *Let ψ be a filter on \mathbb{R} , ψ_0 a filterbase of ψ and $x \in \mathbb{R}$.*

1. a) $\sup \inf \psi := \sup\{\inf P \mid P \in \psi\} \leq \inf\{\sup P \mid P \in \psi\} =: \inf \sup \psi$ and
b) $\bigcap_{P \in \psi} \bar{P} \subseteq [\sup \inf \psi, \inf \sup \psi]$

Note that the two sets $\inf \bigcap_{P \in \psi} \bar{P}$ and $\sup\{\inf P \mid P \in \psi\}$ can be different⁵. The same holds for $\sup \bigcap_{P \in \psi} \bar{P}$ and $\inf\{\sup P \mid P \in \psi\}$.

2. $\psi \xrightarrow{\tau_{\mathbb{R}}} x \iff \sup \inf \psi = \inf \sup \psi = x$
3. a) $\sup\{\inf P \mid P \in \psi_0\} = \sup\{\inf P \mid P \in \psi\}$
b) $\inf\{\sup P \mid P \in \psi_0\} = \inf\{\sup P \mid P \in \psi\}$

The proof can be found on page 51. Using Definition 37 directly for checking if a filter ψ of finite Borel measures converge weakly or not seems to be somewhat elusive, because it characterizes $\psi \xrightarrow{w} P$ in terms of a whole bunch of conditions (for all bounded and continuous f) instead of using the definitions of the involved measures directly. In the next theorem (referred to as the Portmanteau Theorem) we therefore develop more suitable ways of defining weak convergence.

Theorem 40. *(Portmanteau-Theorem)*

For any perfectly normal space E , $P \in M(E)$ and filter ψ on $M(E)$ the following statements are equivalent.

1. $\int_E f d\psi \xrightarrow{\tau_{\mathbb{R}}} \int_E f dP$ for each continuous map $f: E \rightarrow [0, 1]$ ($\psi \xrightarrow{w} P$)
2. $\int_E f d\psi \xrightarrow{\tau_{\mathbb{R}}} \int_E f dP$ for each Lipschitz-continuous map $f: E \rightarrow [0, 1]$, if E is a metric space (E, d) .
3. $\int_E f d\psi \xrightarrow{\tau_{\mathbb{R}}} \int_E f dP$ for each measurable and bounded map $f: E \rightarrow \mathbb{R}$ with $P(D_f) = 0$, where $D_f := \{x \in E \mid \exists \text{ open } V, f(x) \in V \text{ such that } \forall \text{ open } U, x \in U: f(U) \not\subseteq V\}$ is the set of points of discontinuity.

⁴It is interesting to ask whether the converse also holds true.

⁵For example let ψ_0 be an ultrafilter on $(-\infty, 0]$ that does not converge in $(-\infty, 0]$. Define $\psi := \{R \subseteq \mathbb{R} \mid \exists P \in \psi_0 \text{ with } P \cup [1, 2] \subseteq R\}$. Then $\bigcap_{P \in \psi} \bar{P} = [1, 2]$ and $\sup\{\inf P \mid P \in \psi\} \leq 0 < 1 = \inf(\bigcap_{P \in \psi} \bar{P})$.

4. $\inf\{\sup Q(F) \mid Q \in \psi\} \leq P(F)$ for each closed $F \subseteq E$
5. $\sup\{\inf Q(G) \mid Q \in \psi\} \geq P(G)$ for each open $G \subseteq E$
6. $\psi(A) \xrightarrow{\mathbb{R}} P(A)$ for each $A \subseteq E$ with $P(\partial A) = 0$. These A are called P -continuity sets.

In 4. and 5. note that

$$\sup\{\inf Q(G) \mid Q \in \psi\} = \liminf p_n(G) \quad \text{and} \quad \inf\{\sup Q(G) \mid Q \in \psi\} = \limsup p_n(G)$$

if ψ is induced by a sequence $(p_n)_{n \in \mathbb{N}}$ of finite Borel measures.

The proof can be found on page 51. The next theorem gives a criterion for a filter ψ to converge weakly, by showing $\psi(A) \xrightarrow{w} P(A)$ for all $A \in \mathcal{U}$ for a suitable subclass $\mathcal{U} \subseteq \mathcal{B}(E)$.

Theorem 41. Let E be perfectly normal, $P \in M(E)$, $\psi \in \mathcal{F}(M(E))$ and $\mathcal{U} \subseteq \mathcal{B}(E)$ with

1. $A_1, \dots, A_n \in \mathcal{U} \Rightarrow \bigcap_{i=1}^n A_i \in \mathcal{U}$,
2. each open $G \subseteq E$ is a countable union of elements of \mathcal{U} ,
3. $\psi(A) \xrightarrow{\mathbb{R}} P(A)$ for all $A \in \mathcal{U}$.

Then $\psi \xrightarrow{w} P$.

The proof can be found on page 53. The last theorem in this paragraph provides a picturesque idea of weak convergence, especially in the context of sequences of measures and connects it to a famous example of weak convergence, the central limit theorem, cf. the introduction in Billingsley (1968, 1999) and Proposition 5.9 in Kallenberg (2002). Let $F : M(\mathbb{R}) \rightarrow \mathbb{R}^{\mathbb{R}}$, $F(q)(x) := q((-\infty, x])$. If ψ is a filter on $M(\mathbb{R})$, then $F(\psi)$ is a filter on $\mathbb{R}^{\mathbb{R}}$, cf. (62). Given $x \in \mathbb{R}$, $F(\psi)(x) := \{\{f(x) \mid f \in Q\} \mid Q \in F(\psi)\}$ is a filter on \mathbb{R} . If ψ is generated by a sequence $(F_n)_{n \in \mathbb{N}}$ of distribution functions, then $F(\psi)(x)$ is generated by the sequence $(F_n(x))_{n \in \mathbb{N}}$. Note in this context Definition 61 and Lemma 62.

Theorem 42. Let $p \in M(\mathbb{R})$ and ψ be a filter on $M(\mathbb{R})$ such that there exists some $Q \in \psi$ with $q(\mathbb{R}) = p(\mathbb{R})$ for all $q \in Q$. Then

$$\psi \xrightarrow{w} p \Leftrightarrow \forall x \in c(F(p)) : F(\psi)(x) \rightarrow F(p)(x),$$

where $c(F(p))$ is the set of continuity points of $F(p)$.

The proof can be found on page 54.

4 Convergence in distribution

We define the convergence of a filter of random variables in terms of the weak convergence of the distributions of these random variables. Aim of this paragraph is to state and prove Rubin's continuous mapping theorem⁶ (Theorem 46).

⁶This was first proven by H. Rubin in an unpublished paper, see the discussion in the appendix of Anderson (1963) and it was brought to popularity by Billingsley (1968).

Definition 43. Let $X : \Omega \rightarrow E$ be a map from a probability space (Ω, \mathcal{A}, P) to a perfectly normal space (E, τ) that is measurable (i.e. $X^{-1}(B) \in \mathcal{A}$ for all $B \in \mathcal{B}(E)$). The distribution of the random variable X is the probability measure $P^X : \mathcal{B}(E) \rightarrow \mathbb{R}$ defined by $P^X(B) := P(X^{-1}(B))$. I. e., we have a canonical mapping $\mathcal{D} : R(\Omega, E) \rightarrow M(E)$, $\mathcal{D}(Y) := P^Y$. Let

$$R(\Omega, E) := \{Y \in E^\Omega \mid Y^{-1}(B) \in \mathcal{A} \text{ for all } B \in \mathcal{B}(E)\}$$

be the set of all random variables. We say a filter ϕ on $R(\Omega, E)$ converges in distribution to $X \in R(\Omega, E)$ iff

$$\mathcal{D}(\phi) \xrightarrow{w} \mathcal{D}(X) \quad (\text{the image filter converges weakly})$$

and we write $\phi \xrightarrow{\mathcal{D}} X$.

Definition 44. Let (X, τ) and (Y, σ) topological Spaces. A filter Φ on Y^X converges continuously in x to $h \in Y^X$ iff

$$\psi \xrightarrow{\tau} x \text{ implies } \Phi(\psi) \xrightarrow{\sigma} h(x) \text{ for all filter } \psi \text{ on } X, \quad (67)$$

whereby $\Phi(\psi) := \{A \subseteq Y \mid \exists P \in \Phi, Q \in \psi : P(Q) \subseteq A\}$ with $P(Q) := \{g(x) \mid g \in P, x \in Q\}$.

Since $\psi \xrightarrow{\tau} x \Leftrightarrow \dot{x} \cap \tau \subseteq \psi$, (67) is equivalent to

$$\phi([\dot{x} \cap \tau]) \xrightarrow{\sigma} h(x). \quad (68)$$

This is useful in that we have defined the weak convergence of a filter ψ on $M(E)$ with respect to the convergence of a topology τ_w , that is, $\psi \xrightarrow{w} P \Leftrightarrow \psi \xrightarrow{\tau_w} P$.

Lemma 45. Let (X, τ) and (Y, σ) topological spaces, $(f_n)_{n \in \mathbb{N}}$ be a sequence in Y^X and $x \in X$ be a point with a countable neighbourhoodbase (cnb). Then the following is equivalent:

1. The filter associate with $(f_n)_{n \in \mathbb{N}}$ converge continuously in x to f .
2. \forall sequences $(x_n)_{n \in \mathbb{N}}$ in X with $x_n \xrightarrow{\tau} x$ it follows $f_n(x_n) \xrightarrow{\sigma} f(x)$.

The proof can be found on page 55. The next theorem (continuous mapping theorem) is tremendously important for this thesis. We will use the following notation. Let $\gamma : F^E \rightarrow M(F)^{M(E)}$, $\gamma(f)(P) := P^f$, ϕ be a filter on F^E and $\Phi := \gamma(\phi)$ be the image filter of ϕ under γ .

Theorem 46. Let (E, τ) , (F, σ) be perfectly normal, $P \in M(E)$, $h : E \rightarrow F$ and ϕ be a filter on F^E with a countable base. If $P(D) = 0$, with $D := \{x \in E \mid \phi([\dot{x} \cap \tau]) \not\xrightarrow{\sigma} h(x)\}$ and if $\psi \xrightarrow{w} P$ holds as well, then $\Phi(\psi) \xrightarrow{w} h(P)$ follows (note that $h(P) = P^h$ denotes the image measure).

The proof can be found on page 55. Note that in Theorem 46 neither the mapping h needs to be continuous, nor $C(E, F) \in \phi$ was assumed. Moreover, any filter induced by a sequence has a countable basis, but not vice versa.

Part C. Technical statements and proofs

In this part we collect some important concepts and theorems necessary for understanding Part A or the proofs of some statements from it. Depending on the needs, the reader can thus refer here (references to this part are given at all appropriate places). For all statements given here, we provide detailed proofs or extensive references to standard works on the respective topic. We start with a simple lemma about projections, which we needed at the very beginning of the thesis when defining the residuals.

- Lemma 47.** 1. Let $B \in \mathbb{R}^{n \times n}$ be idempotent (i.e., $BB = B$). Then $B = pr_{U,V}$, $B^\top = pr_{V^\perp, U^\perp}$ and $I_n - B = pr_{V,U}$, where $U := im(B)$ and $V := ker(B)$.
2. Let $X \in \mathbb{R}^{k \times d}$ and $U := im(X)$. Then $pr_{U, U^\perp} = X(X^\top X)^- X^\top$, where $(X^\top X)^-$ denotes a generalized inverse⁷ of $X^\top X$. Note that $(X^\top X)^- = (X^\top X)^{-1}$ if $rank(X) = d$.
3. Let $X \in \mathbb{R}^{k \times d}$, $M \in \mathbb{R}^{k \times m}$, $rank(M) + rank(X) \geq k$, $M^\top X = 0$ and $M^\top M = I_m$. Then $MM^\top = pr_{U^\perp, U}$, where $U := im(X)$.

The proof can be found on page 56.

1 Brownian motion and Donskers Theorem

In this small paragraph we give a few important results on Brownian motion that are fundamental to this work and refer to classical and contemporary literature for proofs and further studies.

Brownian motion

Definition 48. We call $(X, (\Omega, \mathcal{A}, \mu), (E, \mathcal{B}), I)$ stochastic process if $(\Omega, \mathcal{A}, \mu)$ is a probability space, (E, \mathcal{B}) a measure space, I is some (index) set and $X : \Omega \times I \rightarrow E$ is a map, such that $X_t : \Omega \rightarrow E$, $X_t(\omega) := X(\omega, t)$ is measurable for all $t \in I$.

Definition 49. A real-valued stochastic process $(B, (\Omega, \mathcal{A}, \mu), (\mathbb{R}, \mathcal{B}(\mathbb{R})), I)$ with the following properties

1. $I \subseteq \mathbb{R}$ is an Intervall of positive length with $0 = \inf(I) \in I$
2. $B_0(\omega) = 0$ a.s.

⁷ $A^- \in \mathbb{R}^{d \times k}$ is a generalized inverse of $A \in \mathbb{R}^{k \times d}$ if $AA^-A = A$ holds, which in turn is equivalent to $\forall y \in im(A) : A(A^-y) = y$. One possible generalized inverse is given, for example, by the Moore-Penrose inverse.

3. $B_{t_n} - B_{t_{n-1}}, \dots, B_{t_2} - B_{t_1}$ are independent for all $0 \leq t_1 < \dots < t_n \in I$
4. $B_t - B_s \sim N(0, t - s)$ for all $0 \leq s < t \in I$
5. the paths $I \ni t \mapsto B(\omega, t) \in \mathbb{R}$ are continuous for μ -almost all $\omega \in \Omega$

is called *Brownian motion*.

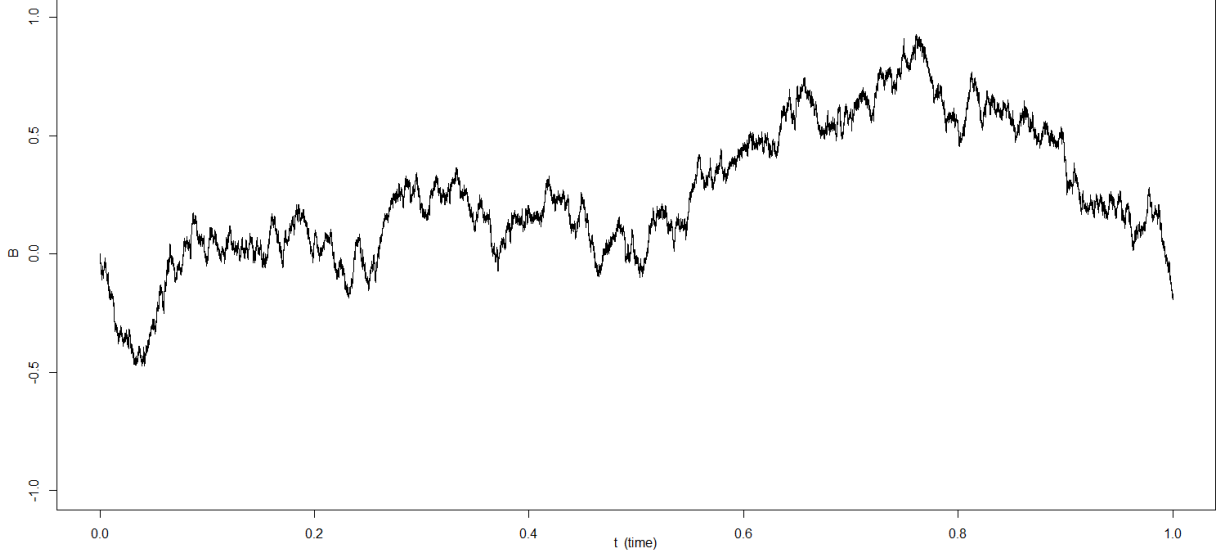


Figure 7: A typical path of a Brownian motion.

Because of 5. we can think of B as a random variable $B : \Omega \rightarrow C(I, \mathbb{R})$, $\omega \mapsto (t \mapsto B(\omega, t))$, where in $C(I, \mathbb{R})$ we take as σ algebra the trace of the product σ algebra $\prod_{i \in I} \mathcal{B}(\mathbb{R})$ on $C(I, \mathbb{R})$. The corresponding image measure $W := \mu^B$ is called *Wiener measure*. That this procedure makes sense and that there are stochastic processes at all which satisfy Definition 49, we summarize with the following short theorem.

Theorem 50. *A Brownian motion exists.*

A proof of this Theorem can be found for example in Billingsley (1968), Ch. 2 or in Hida (1980), Ch. 2 or in Partzsch and Schilling (2012), Ch. 3 and 4. The following equations in the next theorem were first proved by P. Levy. Equation (69) is called *Levi's triple law*.

Theorem 51. *Let $I \subseteq (a, b)$, $a < 0 < b$, $t > 0$, $m_t := \inf_{0 \leq s \leq t} B_s$ and $M_t := \sup_{0 \leq s \leq t} B_s$. Then*

$$\begin{aligned} \mu(M_t > b) &= 2 \cdot \mu(B_t > b) \\ \mu(m_t < a) &= 2 \cdot \mu(B_t < a) \\ \mu(m_t > a \wedge M_t < b \wedge B_t \in I) &= \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi t}} \int_I \left(e^{-\frac{(x+2n(b-a))^2}{2t}} - e^{-\frac{(x-2a-2n(b-a))^2}{2t}} \right) dx \quad (69) \end{aligned}$$

For a proof see for example Hida (1980), Proposition 2.9 and Proposition 2.10, respectively or Partzsch and Schilling (2012), Theorem 6.9 and Theorem 6.18, respectively.

Donsker's Theorem

Definition 52. We call $(\xi_{n,i})_{\substack{n \in \mathbb{N} \\ 1 \leq i \leq n}}$ a triangular array of random variables if $\xi_{n,1}, \dots, \xi_{n,n} : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ are i.i.d. r.v. with $E(\xi_{n,i}) = 0$ and $\text{Var}(\xi_{n,i}) = \sigma^2$ for all $n \in \mathbb{N}$ and $1 \leq i \leq n$.

Theorem 53. Suppose $(\xi_{n,i})_{n \in \mathbb{N}, 1 \leq i \leq n}$ is a triangular array of random variables. Then

$$\frac{1}{\sigma\sqrt{k_n}} T_n(\xi_{n,1}, \dots, \xi_{n,n}) \xrightarrow{\mathcal{D}} B \quad \left(P_{\frac{1}{\sigma\sqrt{k_n}} T_n(\xi_{n,1}, \dots, \xi_{n,n})} \xrightarrow{w} P^B, \text{ cf. Definition 43} \right)$$

For a proof see extension to Theorem 14.1 in Billingsley (1999), Theorem 20.1.1 and Remark 20.1.1 in Borovkov (2013), Satz 10.2 in Bischoff (2014) or Theorem 27.14 in Davidson (1994), the latter being proved for a martingale difference array.

2 Functions with existing limits from left and right

We study monotone functions in terms of their invertibility and define generalized inverses. We then summarize the most important results for us on functions of bounded variation and on measure-defining functions. Finally, we consider functions which have left and right limits at every point and we give a new and interesting construction to factorizes a family $\tilde{f}_i, i \in J$ of such functions simultaneously into continuous functions f_i and a strictly monotone function h , i.e., $\tilde{f}_i = f_i \circ h$ for all $i \in J$. This factorization will become important for weakening the regularity assumptions for the regression functions.

Monotone functions

Each monotone function can be assigned a generalized inverse function, which under certain conditions is the inverse function. In case of distribution functions, these are called quantile function. The corresponding construction for distribution functions is given in practically every textbook on statistics. However, we need this construction for a somewhat larger class of non-decreasing real functions. Curiously, this more general case is not really treated in the literature, although the underlying technique does not change. We try to overcome this deficiency by specifying of all classical properties of the quantile function exactly under which conditions they remain valid in the general case. Then we discuss by a short lemma how to undo “generalized inversion” and we pick up a new and interesting result about the probability integral transformation from Heindl (2022), Lemma 7.1.2, with an alternative and somewhat shorter proof. Finally, we give an extension of Theorem 42.

Definition 54. Let $f : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}$ and t be a cluster point of D . We call L (resp. R) the left (resp. right) limit of f at t if $f(t_n) \rightarrow L$ holds (resp. $f(t_n) \rightarrow R$) for all sequences $(t_n)_{n \in \mathbb{N}}$ in $D \cap (-\infty, t)$ (resp. in $D \cap (t, \infty)$) with $t_n \rightarrow t$.

Lemma 55. Let $f : D \rightarrow \mathbb{R}$, $D \subseteq \mathbb{R}$ and t be a cluster point of D . Then it is equivalent:

1. For all nondecreasing sequences $(t_n)_{n \in \mathbb{N}}$ in $D \cap (-\infty, t)$ with $t_n \rightarrow t$, $f(t_n) \rightarrow f(t)$ holds.
2. For all sequences $(t_n)_{n \in \mathbb{N}}$ in $D \cap (-\infty, t)$ with $t_n \rightarrow t$, $f(t_n) \rightarrow f(t)$ holds.

An analogous equivalence holds for right-sided limits.

The proof can be found on page 56.

Definition 56. Let $I \subseteq \mathbb{R}$ be an interval and $F : I \rightarrow \mathbb{R}$ be non-decreasing. We define the generalized inverse of F by $F^- : Q_F \rightarrow \bar{I} \cup \{-\infty\}$, $F^-(q) := \inf\{t \in I \mid F(t) \geq q\}$ where $Q_F := \{q \in \mathbb{R} \mid \exists a, b \in I \text{ with } F(a) \leq q \leq F(b)\}$. If F is a distribution function, F^- is called quantile function.

Note that Q_F is an interval and $\inf(I) = -\infty$ if I is a left unbounded interval.

Lemma 57. Let $I \subseteq \mathbb{R}$ be an interval and $F : I \rightarrow \mathbb{R}$ be non-decreasing. Then

1. F^- is non-decreasing on Q_F .
2. $F^-(q) = \min\{t \in I \mid F(t) \geq q\}$ if $F^-(q) \in I$ and F is right continuous in $F^-(q)$.
3. $F(F^-(q)) \geq q$ if $F^-(q) \in I$ and F is right continuous in $F^-(q)$.
4. $F^-(F(t)) \leq t$ for all $t \in I$.
5. $q \leq F(t) \Rightarrow F^-(q) \leq t$ for all $t \in I$ and $q \in Q_F$.
6. $F^-(q) \leq t \Rightarrow q \leq F(t)$ if $F^-(q) \in I$ and F is right continuous in $F^-(q)$.
7. $F(F^-(q)) = q$ if $q \in F(I)$, $F^-(q) \in I$ and F is right continuous in $F^-(q)$.
8. F^- is left continuous on Q_F if F is right continuous on I .
9. $F^-|_B$ is the inverse of F if F is strictly increasing, where $B := F(I)$.
10. Let J be an interval, $F : I \rightarrow J$ and let $G : J \rightarrow \mathbb{R}$ be non-decreasing. Then $(G \circ F)^-(q) = F^-(G^-(q))$ applies to any $q \in Q_G$ with $G^-(q) \in Q_F$ and $F^-(G^-(q)) \in I$.

Remark 58. If we interchange the terms “left” and “right” and “ \leq ” with “ \geq ”, all statements from above hold if we also interchange F^- with $F^+(q) := \sup\{t \in I \mid F(t) \leq q\}$.

The proof of Lemma 57 can be found on page 56.

Lemma 59. Let $h : [a, b] \rightarrow [h(a), h(b)]$ be a non-decreasing and left continuous function. Then $h^+ : [h(a), h(b)] \rightarrow [a, b]$ is non-decreasing and right continuous and $(h^+)^- = h$ holds. If $h : [a, b] \rightarrow [h(a), h(b)]$ is right continuous instead, then $(h^-)^+ = h$ holds.

The proof can be found on page 58.

Lemma 60. Let $X : (\Omega, \mathcal{A}, P) \rightarrow \mathbb{R}$ be a r.v. and let F be its distribution function.

1. If U is a uniformly distributed r.v. on $(0, 1)$, then $F^-(U)$, called quantile transformation, also has distribution function F .
2. Define $U := F \circ X$, which is called the probability integral transformation. Then

$$F_U(r) := P(F(X) \leq r) = \begin{cases} 0 & \text{if } r \leq 0 \\ \lim_{x \nearrow F^-(r)} F(x) & \text{if } r \notin F(\mathbb{R}) \text{ and } r \in (0, 1) \\ r & \text{if } r \in F(\mathbb{R}) \text{ and } r \in (0, 1) \\ 1 & \text{if } r \geq 1. \end{cases} \quad (70)$$

Moreover,

$$P^U|_{\sigma(F^-)} = \lambda|_{\sigma(F^-)}, \quad (71)$$

where $P^U|_{\sigma(F^-)}$ resp. $\lambda|_{\sigma(F^-)}$ denotes the restriction of P^U (resp. λ) to the σ algebra $(F^-)^{-1}(\mathcal{B}(\mathbb{R}))$ that is generated by F^- and $\mathcal{B}(\mathbb{R})$. Finally, for $\tilde{X} := F^-(U)$,

$$P(X = \tilde{X}) = 1 \quad \text{and} \quad F(\tilde{X}) = U. \quad (72)$$

The proof can be found on page 57.

Definition 61. A sequence of cumulative distribution functions (respectively quantil functions) is defined to converge weakly to a cumulative distribution function (resp. a quantil function), denoted by $F_n \rightsquigarrow F$ (resp. $F_n^- \rightsquigarrow F^-$), if and only if $F_n(t) \xrightarrow{n \rightarrow \infty} F(t)$ at every t where F is continuous (resp. $F_n^-(u) \xrightarrow{n \rightarrow \infty} F^-(u)$ at every u where F^- is continuous).

Lemma 62. For any sequence of distribution functions, $F_n \rightsquigarrow F$ if and only if $F_n^- \rightsquigarrow F^-$.

The proof can be found on page 58.

Measure defining functions and functions of bounded variation

An important class of real functions directly related to monotone functions are functions of bounded variation. This class plays a significant role in many areas of mathematics. It is interesting for us with respect to its properties in measure and integration theory, since they are directly related to signed measures. The next Theorem is from Tao (2011), Theorem 1.7.9.

Theorem 63. Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be non-decreasing. Then there exists a unique Borel measure $\mu_F : \mathcal{B}(\mathbb{R}) \rightarrow [0, \infty]$ such that $\mu_F([a, b]) = F(b^+) - F(a^-)$, $\mu_F((a, b)) = F(b^-) - F(a^-)$, $\mu_F((a, b]) = F(b^+) - F(a^+)$, $\mu_F([a, b)) = F(b^-) - F(a^+)$ and $\mu_F(\{a\}) = F(a^+) - F(a^-)$ for all $-\infty < a < b < \infty$.

Definition 64. For $a \leq b$ and $f : [a, b] \rightarrow \mathbb{R}$,

$$V(f, a, b) := \sup \left\{ \sum_{i=1}^n |f(x_i) - f(x_{i-1})| \mid n \in \mathbb{N}^{>0} \text{ and } a = x_0 \leq \dots \leq x_n = b \right\}$$

is the variation of f on $[a, b]$.

Lemma 65. Let $a \leq b$ and $f : [a, b] \rightarrow \mathbb{R}$ be given. It holds:

1. $[c, d] \subseteq [a, b] \Rightarrow V(f, c, d) \leq V(f, a, b)$
2. $a \leq c \leq b \Rightarrow V(f, a, b) = V(f, a, c) + V(f, c, b)$
3. $V(f + g, a, b) \leq V(f, a, b) + V(g, a, b)$
4. $V(f, a, b) < \infty \Leftrightarrow \exists g, h$ non decreasing with $f = g - h$.

The representation $f = g - h$ with $g(x) := \frac{1}{2}(V(f, a, x) + f(x))$ and $h(x) := \frac{1}{2}(V(f, a, x) - f(x))$ is called canonical Jordan decomposition.

5. $V(f, a, b) < \infty \Rightarrow f$ is Borel-measurable

6. Let f be left continuous in $c \in (a, b]$ (resp. right continuous in $c \in [a, b)$). Then the function $[a, b] \ni x \mapsto V(f, a, x)$ is left (resp. right) continuous in c too. In particular, g and h from the proof of 4 can be chosen appropriately.

The proof can be found on page 59.

Lemma 66. Let $g : \mathbb{R} \rightarrow \mathbb{R}$ be Borel-measurable and $a \leq b$.

1. Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be non-decreasing and right continuous. Then

$$\int_{\mathbb{R}} 1_{(a,b]} \cdot g \, d\mu_F = \int_{[F(a), F(b)]} g \circ F_0^- \, d\lambda, \quad (73)$$

where $F_0 := F|_{[a,b]} : [a, b] \rightarrow [F(a), F(b)]$ is the restriction of F on $[a, b]$.

2. Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a distribution function. Then

$$\int_{\mathbb{R}} 1_{(a,b]} \cdot g \, d\mu_F = \int_{(F(a), F(b))} g \circ (F^-)|_{(0,1)} \, d\lambda. \quad (74)$$

The proof can be found on page 60.

A simultaneous factorization of a family of functions

Lemma 67. 1. All functions $\tilde{f}_i : [0, 1] \rightarrow \mathbb{R}$, $i \in J$ have an existing left limit in all $x \in (0, 1]$ and an existing right limit in all $x \in [0, 1)$ and $D' := \{x \in [0, 1] \mid \exists i \in J \text{ such that } \tilde{f}_i \text{ is discontinuous at } x\}$ is at most countable. Then there are continuous $f_i : [0, 1] \rightarrow \mathbb{R}$, $i \in J$ and a strictly monotonically increasing $h : [0, 1] \rightarrow [0, 1]$ (i.e. $h(r_1) < h(r_2)$ for all $r_1 < r_2$) with

$$\tilde{f}_i = f_i \circ h \text{ for all } i \in J.$$

2. If all \tilde{f}_i are additionally left continuous (resp. right continuous), then h can be chosen to be additionally left continuous (resp. right continuous) and h can be written as the generalized inverse of a continuous nondecreasing function $F : [0, 1] \rightarrow [0, 1]$, i.e. $h = F^-$ (resp. $h = F^+$), with $F(0) = 0$ and $F(1) = 1$.

3. If in addition all \tilde{f}_i are continuous at $t = 0$ (resp. $t = 1$ if the \tilde{f}_i are right continuous), then

$$F(t) > 0 \text{ for all } t \in (0, 1] \text{ and } F(t) < 1 \text{ for all } t \in [0, 1). \quad (75)$$

4. Regardless of the previous addition, the construction of the f_i yields that all f_i are of bounded variation if all \tilde{f}_i are of bounded variation.

The proof can be found on page 61. It is clear that all implications stated in Lemma 67 are in fact equivalences. That is, if, for example, $\tilde{f}_i = f_i \circ F^-$ holds for all $i \in J$ and continuous f_i and F is a continuous distribution function, then D' is countable, all \tilde{f}_i are left continuous and have right limits. An analogous formulation holds for all other stated implications. The usefulness of this lemma is then shown in the proof of Theorem 79 and the stochastic applications to the (recursive) residual partial sum processes given below. The basic idea is that regression functions \tilde{f}_i which are not continuous (but satisfy the conditions of Lemma 67) are factorized $\tilde{f}_i = f_i \circ F^-$. Certain computations can then be carried out for the “new” continuous

regression functions, while F can be integrated into the remaining statistical model assumptions. In the resulting formulas of the limit processes considered there, f_i and F now occur. By certain transformations these reduce to the “old” regression functions \tilde{f}_i and the factorization no longer appears in the formulas. However, we are sure that this lemma (together with the related Lemma 69) has applications in other areas as well.

Remark 68. *Note that a function $f : [0, 1] \rightarrow \mathbb{R}$ with existing left (or right) limits in every $x \in [0, 1]$ can have at most countably many points of discontinuity. Thus it follows that D' from Lemma 67 is countable in any case if J is countable. But of course D' can still be countable even if J is uncountable. This remains then to be checked for the concrete $\tilde{f}_i, i \in J$.*

The proof of Remark 68 can be found on page 63. We give a second version of Lemma 67 with a different proof. It was the version from Lemma 69 that was first conjectured by Thomas Heindl (with f_i being continuous and F being right continuous) and shortly thereafter proved independently by Heindl (in a version similar to that from Lemma 69 with f_i being Lipschitz continuous and F being right continuous, as Lemma 7.1.5 in his dissertation) and Evers (in the version given here in Lemma 67). For Lemma 69 we give two slightly different proofs, both based on Heindl’s original measure-theoretic argument (cf. Remark 81). Note that the construction of F in the proof of Lemma 67 depends only on D' and not on the mappings $\tilde{f}_i, i \in J$ and thus works for infinite J as well. The construction of F in the proof of Lemma 69, on the other hand, depends explicitly on all $\tilde{f}_i, i \in J$. Moreover, J must be finite and all \tilde{f}_i must be of bounded variation. But for this the second proof yields Lipschitz continuous functions, instead of continuous functions.

Lemma 69. *Let $\tilde{f}_1, \dots, \tilde{f}_d : [0, 1] \rightarrow \mathbb{R}$ be of bounded variation. Then there are Lipschitz continuous functions $f_1, \dots, f_d : [0, 1] \rightarrow \mathbb{R}$ and a strictly monotonically increasing $h : [0, 1] \rightarrow [0, 1]$ with $\tilde{f}_i = f_i \circ h$ for all $i \in \{1, \dots, d\}$. If all \tilde{f}_i are left continuous, then h is left continuous too and furthermore $h = F^-$ for some non-decreasing continuous function $F : [0, 1] \rightarrow [0, 1]$ with $F(0) = 0$ and $F(1) = 1$. Moreover, if every \tilde{f}_i is continuous in 0, then (75) also holds.*

The proof can be found on page 63.

3 Asymptotic- F -designs

Lemma 70. *Let $s \in (0, 1]$ and $(t_{n,1}, \dots, t_{n,n})$ be an asymptotic F -design in $[a, b]$, i.e. $a \leq t_{n,1} \leq \dots \leq t_{n,n} \leq b$ for all $n \in \mathbb{N}$. Then $(t_{n, \lfloor sn \rfloor})_{n \in \mathbb{N}}$ has cluster points and for each such cluster point $x', x' \geq F^-(s)$ holds. If F^- is continuous in s , then $t_{n, \lfloor sn \rfloor} \xrightarrow{n \rightarrow \infty} F^-(s)$ holds.*

The proof can be found on page 65. For the upper F -design we can strengthen Lemma 70 and we obtain $t_{n, \lfloor sn \rfloor} \rightarrow F^-(s)$ for every $s \in (0, 1)$. [Proof: Since $t_{n, \lfloor sn \rfloor} = F^-(\frac{\lfloor sn \rfloor}{n})$, $\frac{\lfloor sn \rfloor}{n} \rightarrow s$ and $\frac{\lfloor sn \rfloor}{n} \leq s$, we are done due to the left continuity of F^- and Lemma 55. \square] The nearby conjecture $t_{n, \lfloor sn \rfloor} \rightarrow F^-(t)$ for all s , is wrong. In fact, $F_n^-(s_n)$ may even be divergent, even if $(s_n)_{n \in \mathbb{N}}$ is a sequence convergent in $[0, 1]$. This can be seen in the following example.

Example 71. *We define on $[a, b] = [0, 2]$ the designs $(t_{n,1}, \dots, t_{n,n})$ with*

$$(t_{n,1}, \dots, t_{n,n}) = \begin{cases} \left(\underbrace{0, \dots, 0}_{n/2-1}, \underbrace{2, \dots, 2}_{n/2+1} \right) & \text{if } n \text{ is even,} \\ \left(\underbrace{0, \dots, 0}_{(n+1)/2}, \underbrace{2, \dots, 2}_{(n-1)/2} \right) & \text{if } n \text{ is odd.} \end{cases}$$

To these correspond $F_n : \mathbb{R} \rightarrow [0, 1]$ for n even and odd, respectively, see (26)

$$F_n(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} - \frac{1}{n} & \text{if } 0 \leq x < 2, \\ 1 & \text{if } x \geq 2 \end{cases}, \quad \text{and} \quad F_n(x) = \begin{cases} 0 & \text{if } x < 0 \\ \frac{1}{2} + \frac{1}{2n} & \text{if } 0 \leq x < 2, \\ 1 & \text{if } x \geq 2. \end{cases}$$

Obviously $\sup_{x \in [0, 2]} |F_n(x) - F(x)| \xrightarrow{n \rightarrow \infty} 0$ holds for $F(x) := \begin{cases} 0 & \text{if } x < 0 \\ 1/2 & \text{if } 0 \leq x < 2. \\ 1 & \text{if } 2 \leq x \end{cases}$

Thus $t_{n, \lfloor \frac{n}{2} \rfloor} = \begin{cases} 2 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$ and $F^-(\frac{1}{2}) = 0$. But even if $(s_n)_{n \in \mathbb{N}}$ converges in $[0, 1]$, in

general $F_n^-(s_n)$ need not converge at all: $F_n^-(s_n) = \begin{cases} 2 & \text{if } n \text{ is even} \\ 0 & \text{if } n \text{ is odd} \end{cases}$ for $s_n := \frac{1}{2} - \frac{1}{2n}$.

Lemma 72. Let $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$ be an asymptotic F -design in $[a, b]$ and F^- be continuous in $s \in (0, 1]$. Let $u \in C[0, 1]$ and $f \in C[a, b]$ be of bounded variation. Then

$$\sum_{i=2}^{\lfloor sn \rfloor} (f(t_{n,i}) - f(t_{n,i-1})) u \left(\frac{i-1}{n} \right) \xrightarrow{n \rightarrow \infty} \int_{[a, F^-(s)]}^{(R)} (u \circ F) df. \quad (76)$$

The proof can be found on page 65.

4 Bounds for largest and smallest eigenvalues of a Gram matrix

In the context of recursive (resp. ordinary) residuals, Gramian matrices occur in a natural way. To ensure the convergence of certain expressions, we need estimates of the eigenvalues of these matrices. We define the spectral norm by $\|A\|_2 := \max_{\|x\|_2=1} \|Ax\|_2$. Note the relations

$$\left\| \frac{1}{n} X^\top X \right\|_2^2 = \left[\chi_{\max} \left(\frac{1}{n} X^\top X \right) \right]^2, \quad \left\| \left(\frac{1}{n} X^\top X \right)^{-1} \right\|_2^2 = \left[\chi_{\min} \left(\frac{1}{n} X^\top X \right) \right]^{-2}$$

and

$$\frac{1}{n} \sum_{k=1}^m f_i(t_{n,k}) f_j(t_{n,k}) = \int_{(0, \frac{m}{n})} (f_i f_j) \circ F_n^- d\lambda,$$

where the minimum and maximum eigenvalues of a symmetric matrix are denoted by $\chi_{\min}(A)$ and $\chi_{\max}(A)$, respectively.

Lemma 73. Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a distribution function, $a < c \leq b$, $f_1, \dots, f_d \in L_2([a, b], F)$ and $H(t) := \int_{(0, t)} (f f^\top) \circ (F^-)|_{(0, 1)} d\lambda$ for all $F(c) \leq t \leq F(b)$. If (A1) or (A2) holds,

(A1) $f_1 \cdot \mathbf{1}_{(a, c]}, \dots, f_d \cdot \mathbf{1}_{(a, c]}$ are linearly independent in $L_2([a, b], F)$,

(A2) $a \leq \inf(\text{support}(F))$ and $f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]}$ are linearly independent in $L_2([a,b], F)$, then

$$\exists \chi_{\min}, \chi_{\max} \in (0, \infty) \forall t \in [F(c), F(b)] : \chi_{\min} \leq \chi_{\min}(H(t)) \leq \chi_{\max}(H(t)) \leq \chi_{\max}. \quad (77)$$

The proof can be found on page 67.

Lemma 74. Let $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$ be a sequence of designs in $[a, b]$ with distribution functions F_n (in the sense of (26)). Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a distribution function with $F_n \rightsquigarrow F$ (cf. Definition 61) and let $g : [a, b] \rightarrow \mathbb{R}$ be λ -a.e. continuous. Then $\forall \varepsilon > 0 \exists K_\varepsilon \subseteq (0, 1)$ with $\lambda(K_\varepsilon) > 1 - \varepsilon$ and $\sup_{s \in K_\varepsilon} |g(F_n^-(s)) - g(F^-(s))| \rightarrow 0$.

The proof can be found on page 68.

Lemma 75. Let $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$ be a sequence of designs in $[a, b]$ with distribution functions F_n (in the sense of (26)). Let $F : \mathbb{R} \rightarrow \mathbb{R}$ be a distribution function with $F_n \rightsquigarrow F$ (cf. Definition 61), $a < c \leq b$ with $F(c) > 0$, $K \in \mathbb{N}$ and $f_1, \dots, f_d : [a, b] \rightarrow \mathbb{R}$ be bounded. Under the assumptions of Lemma 73, $\tilde{\chi}_{\min}, \tilde{\chi}_{\max} > 0$ and $n_0 \in \mathbb{N}$ exists such that $\forall n \geq n_0 \forall t \in [F(c), F(b)]$

$$0 < \tilde{\chi}_{\min} \leq \chi_{\min} \left(\frac{1}{n} X_{n, [tn]-K}^\top X_{n, [tn]-K} \right) \leq \chi_{\max} \left(\frac{1}{n} X_{n, [tn]-K}^\top X_{n, [tn]-K} \right) \leq \tilde{\chi}_{\max}.$$

The proof can be found on page 68.

Remark 76. From Lemma 75 it follows immediately that under its assumptions and consequences $\text{rank}(X_{n,k}) = d$ holds for all $n \geq n_0$ and all k with $\lfloor n \cdot F(c) \rfloor - d \leq k \leq n$.

Lemma 77. Under the assumptions of Lemma 75,

$$\sup_{t \in [0,1]} \left\| \frac{1}{n} X_{n, [tn]-K}^\top X_{n, [tn]-K} - \int_{[0,t]} (ff^\top) \circ F^- d\lambda \right\| \xrightarrow{n \rightarrow \infty} 0 \quad (78)$$

holds for all $K \in \mathbb{N}$ and each matrix norm $\|\cdot\|$ (note that all matrix norms are equivalent). However, not only the above relation is valid, but also

$$\sup_{t \in [F(c), 1]} \left\| \left(\frac{1}{n} X_{n, [tn]-K}^\top X_{n, [tn]-K} \right)^{-1} - \left(\int_{[0,t]} (ff^\top) \circ F^- d\lambda \right)^{-1} \right\| \xrightarrow{n \rightarrow \infty} 0. \quad (79)$$

Note that the c from the premises in Lemma 75 plays no role for (78).

The proof can be found on page 69.

Lemma 78. Under the assumptions of Lemma 75 there exists $r \geq 0$ and $n_0 \in \mathbb{N}$ with

$$1 \leq c_{n, [tn]-K} = 1 + \frac{1}{n} f(t_{n, [tn]-K})^\top \left(\frac{1}{n} X_{n, [tn]-1-K}^\top X_{n, [tn]-1-K} \right)^{-1} f(t_{n, [tn]-K}) \leq 1 + \frac{r}{n}$$

for all $n \geq n_0$ and $t \in [F(c), 1]$, where $c_{n,k} := 1 + \|a_{n,k}\|^2$ and $a_{n,k}$ is defined in Definition 1.

The proof can be found on page 70.

5 A theorem on continuous convergence in $C[0, 1]$

In this section, we formulate the technical heart to determine the distribution of the recursive residual partial sum limit process under H_0 and under H_1 . Following MacNeill (1978a) we introduce a discrete difference operator $V_n : C[0, 1] \rightarrow \mathbb{R}^n$ by

$$V_n(u) := \left(u\left(\frac{1}{n}\right) - u(0), u\left(\frac{2}{n}\right) - u\left(\frac{1}{n}\right), \dots, u(1) - u\left(\frac{(n-1)}{n}\right) \right)^\top \quad (80)$$

Next, we define⁸

$$\phi_n : C[0, 1] \rightarrow C[0, 1], \quad \phi_n := T_{n-d} \circ g \circ M_{n,n}^\top \circ V_n,$$

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{M_{n,n}^\top} & \mathbb{R}^{n-d} \\ V_n \uparrow & & \downarrow T_{n-d} \circ g \\ C[0, 1] & \xrightarrow{\phi_n} & C[0, 1] \end{array} \quad (81)$$

according to the right digram, where $M_{n,n}$ is defined in

(15) and T_{n-d} and g were defined in (19) and (22), respectively. Obviously, the following equations hold true

$$V_n \circ T_n = id_{\mathbb{R}^n}, \quad (\text{thus } T_n \text{ is injective and } V_n \text{ is surjective}) \quad (82)$$

$$T_{n-d} \circ g \circ r_n \stackrel{(18)}{=} T_{n-d} \circ g \circ M_{n,n}^\top \circ \varepsilon_n \stackrel{(82)}{=} T_{n-d} \circ g \circ M_{n,n}^\top \circ V_n \circ T_n \circ \varepsilon_n = \phi_n \circ T_n \circ \varepsilon_n. \quad (83)$$

Because of (82) one can think of T_n as a discrete integral operator.

Theorem 79. *Let $(t_{n,1}, \dots, t_{n,n})_{n \in \mathbb{N}}$ be an asymptotic F -design in $[a, b]$ and consider model (3) with $f_1, \dots, f_d : [a, b] \rightarrow \mathbb{R}$ all of bounded variation, left continuous on $(a, b]$, continuous in a and let $f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]}$ are linearly independent (l.i.) in $L_2([a, b], F)$ for some $c \in (a, b]$. Let $\delta := F(c) > 0$, ϕ_n defined as in (81). Then $\phi : C[0, 1] \rightarrow C[0, 1]$,*

$$\begin{aligned} \phi(u)(x) &:= u(x) - u(\delta) \\ &+ \int_{\delta}^x f(F_0^-(t))^\top \left(\int_{[0,t]} (ff^\top) \circ F_0^- d\lambda \right)^{-1} \left[\int_{(a, F_0^-(t))} (u \circ F_0) d\mu_f + f(F_0^-(0))u(0) - f(F_0^-(t))u(t) \right] dt, \end{aligned}$$

for $x \in [\delta, 1]$ and $\phi(u)(x) = 0$ for $x \in [0, \delta)$, is continuous with respect to the sup norm, where $F_0 := F|_{[a,b]}$ and μ_f is the Lebesgue-Stieltjes measure generated⁹ by f and

$$\|\phi(u) - \phi_n(u)\| \rightarrow 0 \quad \text{and} \quad \|\phi(u) - \phi_n(u_n)\| \rightarrow 0$$

holds for all $u, u_n \in C[0, 1]$ with $\|u - u_n\| \rightarrow 0$ (with sup-norm $\|\cdot\| = \sup_{x \in [0,1]} |\cdot|$ on $C[0, 1]$).

The proof can be found on page 70.

⁸For ordinary residuals MacNeill (1978a) and Bischoff (1998) defined a similar function ϕ_n by $\phi_n : C[0, 1] \rightarrow C[0, 1]$, $\phi_n := T_n \circ \text{pr}_{U_n^\perp} \circ V_n$ with $U_n := \text{Im}(X_{n,n})$. Note that $\text{pr}_{U_n^\perp} = M_{n,n} M_{n,n}^\top$, cf. (17).

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{\text{pr}_{U_n^\perp}} & U \leq \mathbb{R}^n \\ V_n \uparrow & & \downarrow T_n \\ C[0, 1] & \xrightarrow{\phi_n} & C[0, 1] \end{array}$$

⁹For the existence of the Lebesgue-Stieltjes measure μ_f , see Theorem 63. Note that it is formulated for non-decreasing functions. By means of Lemma 65.4, it can be formulated and proved also for functions of bounded variation.

6 Proofs

Proof of Theorem 2 from page 8:

1. A short computation immediately results in

$$\begin{aligned} X_{n,j}^\top X_{n,j} &= \left(\begin{pmatrix} X_{n,j-1}^\top & 0_{d \times 1} \end{pmatrix} + \begin{pmatrix} 0_{d \times (j-1)} & f(t_{n,j}) \end{pmatrix} \right) \left(\begin{pmatrix} X_{n,j-1} \\ 0_{1 \times d} \end{pmatrix} + \begin{pmatrix} 0_{(j-1) \times d} \\ f(t_{n,j})^\top \end{pmatrix} \right) \\ &= X_{n,j-1}^\top X_{n,j-1} + f(t_{n,j})f(t_{n,j})^\top. \end{aligned}$$

2. From 1. we obtain

$$\begin{aligned} (X_{n,j}^\top X_{n,j})^{-1} &= (X_{n,j-1}^\top X_{n,j-1} + f(t_{n,j})f(t_{n,j})^\top)^{-1} \\ &\stackrel{(*)}{=} (X_{n,j-1}^\top X_{n,j-1})^{-1} - \frac{1}{1 + \|a_{n,j}\|^2} (X_{n,j-1}^\top X_{n,j-1})^{-1} f(t_{n,j})f(t_{n,j})^\top (X_{n,j-1}^\top X_{n,j-1})^{-1} \end{aligned}$$

where (*) follows from Woodbury's formula

$$(R + STU)^{-1} = R^{-1} - R^{-1}S(T^{-1} + UR^{-1}S)^{-1}UR^{-1},$$

if $R \in \mathbb{R}^{n \times n}$, $S \in \mathbb{R}^{n \times m}$, $T \in \mathbb{R}^{m \times m}$, $U \in \mathbb{R}^{m \times n}$, and R , T and $T^{-1} + UR^{-1}S$ are nonsingular. See Harville (2008), Theorem 18.2.8 (Woodbury's formula).

3. First of all,

$$\begin{aligned} X_{n,j}^\top X_{n,j} \hat{\beta}_{n,j} &= X_{n,j}^\top Y_j = X_{n,j-1}^\top Y_{j-1} + f(t_{n,j})Y_{n,j} \\ &= X_{n,j-1}^\top X_{n,j-1} \hat{\beta}_{n,j-1} + Y_{n,j} f(t_{n,j}) \\ &\stackrel{!}{=} (X_{n,j}^\top X_{n,j} - f(t_{n,j})f(t_{n,j})^\top) \hat{\beta}_{n,j-1} + Y_{n,j} f(t_{n,j}) \\ &= X_{n,j}^\top X_{n,j} \hat{\beta}_{n,j-1} + (Y_{n,j} - f(t_{n,j})^\top \hat{\beta}_{n,j-1}) f(t_{n,j}). \end{aligned}$$

By multiplying with $(X_{n,j}^\top X_{n,j})^{-1}$ we get

$$\begin{aligned} \hat{\beta}_{n,j} &= \hat{\beta}_{n,j-1} + (Y_{n,j} - f(t_{n,j})^\top \hat{\beta}_{n,j-1}) (X_{n,j}^\top X_{n,j})^{-1} f(t_{n,j}) \\ &= \hat{\beta}_{n,j-1} + \sqrt{1 + \|a_{n,j}\|^2} r_{n,j} (X_{n,j}^\top X_{n,j})^{-1} f(t_{n,j}). \end{aligned}$$

4. By (18) we get $w_j = \text{pr}_{U_j^\perp, U_j} Y_j = M_{n,j} M_{n,j}^\top Y_j$ for $j = d+1, \dots, n$, which implies $\|w_{d+1}\|^2 = \|M_{n,d+1}^\top Y_{d+1}\|^2 = \|r_{n,d+1}\|^2$ and

$$\begin{aligned} \|w_j\|^2 &= Y_j^\top M_{n,j} M_{n,j}^\top M_{n,j} M_{n,j}^\top Y_j = \|M_{n,j}^\top Y_j\|^2 \\ &= \left\| M_{n,j-1}^\top Y_{j-1} \right\|^2 + \frac{1}{1 + a_{n,j}^2} |(-a_{n,j}^\top \mathbf{1}) Y_j|^2 \\ &= \|w_{j-1}\|^2 + r_{n,j}^2, \quad j = d+2, \dots, n, \end{aligned}$$

noting $\|(v_1, \dots, v_k)^\top\|^2 = \|(v_1, \dots, v_{k-1})^\top\|^2 + v_k^2$ for all $v_1, \dots, v_k \in \mathbb{R}$. \square

Proof of Lemma 8 from page 11:

(27) is best seen by means of the following 3 cases:

1. $t < t_{n,1}$. In this case $F_n(t) = 0 \leq F(t) < \frac{1}{n} = F_n(t) + \frac{1}{n}$.
2. $\exists i$ with $t_{n,i} \leq t < t_{n,i+1}$. Then

$$F_n(t_{n,i}) = F_n(t) = \frac{i}{n} \leq F(t_{n,i}) \leq F(t) < \frac{i+1}{n} = F_n(t_{n,i}) + \frac{1}{n} \leq F_n(t) + \frac{1}{n}.$$

3. $t_{n,n} \leq t$. In this case $F_n(t) = 1 = F_n(t_{n,n}) \leq F(t_{n,n}) \leq F(t) \leq 1 \leq F_n(t_{n,n}) + \frac{1}{n}$ follows. \square

Proof of Theorem 10 from page 11:

$P^{\frac{1}{\sigma\sqrt{n}}T_n \circ g \circ \varepsilon_n} \xrightarrow{w} P^{B'}$ in $C[0, 1]$ and $P^{\frac{1}{\sigma\sqrt{n}}T_n \circ \varepsilon_n} \xrightarrow{w} P^B$ in $C[0, 1]$ hold by Theorem 53 (Donsker's Theorem). Theorem 46, Theorem 79 and equation (83) now imply

$$P^{\left(\frac{1}{\sigma\sqrt{n-d}}T_{n-d} \circ g \circ r_n\right)} = P^{\left(\sqrt{\frac{n}{n-d}} \cdot \phi_n \circ \left(\frac{1}{\sigma\sqrt{n}}T_n \circ \varepsilon_n\right)\right)} = \left(P^{\frac{1}{\sigma\sqrt{n}}T_n \circ \varepsilon_n}\right)^{\sqrt{\frac{n}{n-d}} \cdot \phi_n} \xrightarrow{w} (P^B)^{\phi'} = P^{\phi' \circ B}$$

But $P^{\phi' \circ B}$ is always the same (independent of the concrete distribution of ε_n). Without loss of generality we assume additionally $\varepsilon_n \sim N(0, \sigma^2 I_n)$. Then r_n is independent, by (16) and we conclude with Donsker's Theorem, as in (25), $P^{\left(\frac{1}{\sigma\sqrt{n-d}}T_{n-d} \circ g \circ r_n\right)} \xrightarrow{w} P^{B'}$. \square

Proof of Theorem 11 from page 11:

Note that the limit process B is a Brownian motion on $[\delta, 1]$ with $B_t = 0$ for all $t \in [0, \delta]$. Therefore, for the relevant crossing probabilities, we consider instead a standard Brownian motion on the interval $[0, 1 - \delta]$, which we denote by B^* to distinguish them. Note that $B_n, B : \Omega \rightarrow C[0, 1]$ can be understood as random variables with values in $C[0, 1]$. Furthermore, let $\text{sup} : C[0, 1] \rightarrow \mathbb{R}$, $\text{sup}(f) := \sup_{0 \leq x \leq 1} f(x)$. Obviously sup is continuous ($C[0, 1]$ with sup norm). Therefore, by Theorem 46 it follows $(P^{B_n})^{\text{sup}} \xrightarrow{w} (P^B)^{\text{sup}}$ and together with Theorem 40.6 and Theorem 51 we conclude for the first test (29)

$$\begin{aligned} P(\text{reject } H_0 \mid H_0 \text{ is true}) &= (P^{B_n})^{\text{sup}}((x, \infty)) \xrightarrow{n \rightarrow \infty} (P^B)^{\text{sup}}((x, \infty)) \\ &= P\left(\sup_{0 \leq s \leq 1} B_s > x\right) = P\left(\sup_{0 \leq s \leq 1-\delta} B_s^* > x\right) \\ &= 2 \cdot P(B_{1-\delta}^* > x) = 2 \cdot \left(1 - \Phi\left(\frac{x}{1-\delta}\right)\right) = \alpha, \end{aligned}$$

since $(P^B)^{\text{sup}}(\partial(x, \infty)) = (P^B)^{\text{sup}}(\{x\}) = \lim_{\varepsilon \searrow 0} P(x - \varepsilon < \sup_{0 \leq s \leq 1} B_s^* \leq x) = 0$ (this follows from the distribution of $\sup_{0 \leq s \leq t} B_s^*$, cf. Partzsch and Schilling (2012), Theorem 6.9). We proceed completely analogously for the tests (30) and (31).

$$\begin{aligned} P(\text{reject } H_0 \mid H_0 \text{ is true}) &\xrightarrow{n \rightarrow \infty} P\left(\inf_{0 \leq s \leq 1-\delta} B_s^* < y\right) \\ &= 2 \cdot P(B_{1-\delta}^* < y) = 2 \cdot \Phi\left(\frac{y}{1-\delta}\right) = \alpha \end{aligned}$$

and

$$\begin{aligned}
P(\text{reject } H_0 | H_0 \text{ is true}) &\xrightarrow{n \rightarrow \infty} P\left(\inf_{0 \leq s \leq 1-\delta} B_s^* \leq -z \vee \sup_{0 \leq s \leq 1-\delta} B_s^* \geq z\right) \\
&= 1 - P\left(\inf_{0 \leq s \leq 1-\delta} B_s^* > -z \wedge \sup_{0 \leq s \leq 1-\delta} B_s^* < z \wedge B_{1-\delta}^* \in (-z, z)\right) \\
&= 1 - \sum_{n=-\infty}^{\infty} \frac{1}{\sqrt{2\pi(1-\delta)}} \int_{-z}^z \left(e^{-\frac{(x+4nz)^2}{2(1-\delta)}} - e^{-\frac{(x-2z(2n-1))^2}{2(1-\delta)}}\right) dx \\
&= 1 - \sum_{n=-\infty}^{\infty} A_n \leq \alpha,
\end{aligned}$$

respectively, where

$$\begin{aligned}
A_n &:= \frac{1}{\sqrt{2\pi(1-\delta)}} \int_{-z}^z \left(e^{-\frac{(x+4nz)^2}{2(1-\delta)}} - e^{-\frac{(x-2z(2n-1))^2}{2(1-\delta)}}\right) dx \\
&= \Phi\left(\frac{(4n+1)z}{1-\delta}\right) - \Phi\left(\frac{(4n-1)z}{1-\delta}\right) - \Phi\left(-\frac{(4n-3)z}{1-\delta}\right) + \Phi\left(-\frac{(4n-1)z}{1-\delta}\right).
\end{aligned}$$

□

Proof of Lemma 14 from page 14:

First of all, $\Gamma \in C[0, 1]$ obviously holds. We define $u, u_n \in C[0, 1]$ by

$$u(s) := \int_0^s f_{d+1} \circ F_0^- d\lambda \quad \text{and} \quad u_n(s) := \int_0^s f_{d+1} \circ F_n^- d\lambda.$$

Note that $u(0) = 0$. Thus

$$V_n(u_n) = \frac{1}{n} \cdot \xi_n,$$

since

$$u_n(k/n) = \int_0^{k/n} f_{d+1} \circ F_n^- d\lambda = \sum_{i=1}^k f_{d+1}(F_n^-(i/n)) \cdot \frac{1}{n}.$$

We infer

$$T_{n-d} \circ g\left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n\right)\right) = T_{n-d} \circ g(M_{n,n}^\top \cdot V_n(u_n)) = \phi_n(u_n).$$

Let $\varepsilon > 0$. f_{d+1} is bounded, since it is of bounded variation. Thus we can define $m := \sup_{t \in [a,b]} |f_{d+1}(t)|$. By Lemma 74

$$\begin{aligned}
\exists K_\varepsilon \subseteq [0, 1] \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \\
\lambda(K_\varepsilon) > 1 - \varepsilon \quad \text{and} \quad \sup_{s \in K_\varepsilon} |f_{d+1} \circ F_0^-(s) - f_{d+1} \circ F_n^-(s)| \leq \varepsilon.
\end{aligned}$$

Therefore

$$\begin{aligned}
|u(s) - u_n(s)| &= \left| \int_{[0,s]} f_{d+1} \circ F_0^- d\lambda - \int_{[0,s]} f_{d+1} \circ F_n^- d\lambda \right| \\
&\leq \int_{[0,s]} |f_{d+1} \circ F_0^- - f_{d+1} \circ F_n^-| d\lambda \\
&\leq \int_{[0,1]} |f_{d+1} \circ F_0^- - f_{d+1} \circ F_n^-| d\lambda \\
&\leq \lambda(K_\varepsilon) \cdot \varepsilon + \lambda(K_\varepsilon^C) \cdot 2 \cdot m \leq \varepsilon(1 + 2m),
\end{aligned}$$

holds for all $s \in [0, 1]$. Thus

$$\sup_{s \in [0,1]} |u(s) - u_n(s)| \xrightarrow{n \rightarrow \infty} 0.$$

Now we can apply Theorem 79 to infer

$$\sup_{s \in [0,1]} \left| T_{n-d} \circ g \left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n \right) \right) (s) - \Gamma(s) \right| = \sup_{s \in [0,1]} |\phi_n(u_n)(s) - \phi(u)(s)| \xrightarrow{n \rightarrow \infty} 0.$$

□

Example 80. To understand the example, note the comment before Lemma 14. Let $[a, b] = [0, 1]$ and $t_{10,i} := \frac{i}{10}$, $i \in \{1, \dots, 10\}$. According to Rabovski (2003), p. 20,

$$u_{10} \left(\frac{k}{10k+1} \right) = \int_{[0,1]} f_{d+1} dF_{10, \frac{k}{10k+1}} = 0,$$

because

$$F_{10, \frac{k}{10k+1}}(x) = \min \left(F_{10}(x), \frac{\lfloor \frac{10k}{10k+1} \rfloor}{10} \right) = 0, \quad \text{for all } x \in [0, 1].$$

On the other hand,

$$u_{10} \left(\frac{1}{10} \right) = \int_{[0,1]} f_{d+1} dF_{10, \frac{1}{10}} = \frac{1}{10} \cdot f_{d+1} \left(\frac{1}{10} \right),$$

because

$$F_{10, \frac{1}{10}}(x) = \min \left(F_{10}(x), \frac{1}{10} \right) = \begin{cases} 0 & \text{if } 0 \leq x < \frac{1}{10} \\ \frac{1}{10} & \text{if } \frac{1}{10} \leq x \leq 1 \end{cases}, \quad \text{for all } x \in [0, 1].$$

Thus u_{10} is not continuous because of $\frac{k}{10k+1} \xrightarrow{k \rightarrow \infty} \frac{1}{10}$ (if $f_{d+1}(\frac{1}{10}) \neq 0$).

Proof of Theorem 15 from page 14:

The computation from (33) gives

$$\begin{aligned}
B_n &= \frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot (X_{n,n} \beta + \varepsilon_n) \right) + \frac{n \cdot \beta_{d+1}^*}{\sigma \cdot \sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n \right) \right) \\
&= \underbrace{\frac{1}{\sigma \sqrt{n-d}} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot (X_{n,n} \beta + \varepsilon_n) \right)}_{=: Z_n} + \underbrace{\frac{\beta_{d+1}}{\sigma} \cdot T_{n-d} \circ g \left(M_{n,n}^\top \cdot \left(\frac{1}{n} \cdot \xi_n \right) \right)}_{=: \Gamma_n}
\end{aligned}$$

for model (35). If we define $h, h_n : C[0, 1] \rightarrow C[0, 1]$ by $h(\alpha) := \alpha + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma$ and $h_n(\alpha) := \alpha + \Gamma_n$, then $P^{B_n} = h_n(P^{Z_n})$ holds and the conditions of Theorem 46 are satisfied because of Lemma 14. Thus it follows $h_n(P^{Z_n}) \xrightarrow{w} h(P^{B'})$, which means $B_n \xrightarrow{\mathcal{D}} B' + \frac{\beta_{d+1}}{\sigma} \cdot \Gamma$. \square

Proof of Theorem 19 from page 17:

First we proceed exactly as in Step 2 in the proof of Theorem 79, up to and including (127). Next, we show the linear independence of $g_i, i \in \{1, \dots, d\}$ in $L_2([a, b], F \circ G)$.

$$\begin{aligned}
& g_1, \dots, g_d \text{ are l.i. in } L_2([a, b], F \circ G) \\
& \Leftrightarrow \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i g_i \cdot 1_{[a, b]} \right)^2 d\mu_{F \circ G} > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i g_i \cdot 1_{(a, b]} \right)^2 d\mu_{F \circ G} + \left(\sum_{i=1}^d \alpha_i g_i(a) \right)^2 \cdot F(G(a)) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \stackrel{(73)}{\Leftrightarrow} \int_{[F(G(a)), F(G(b))]} \left(\sum_{i=1}^d \alpha_i g_i \circ (F \circ G)_0^- \right)^2 d\lambda + \left(\sum_{i=1}^d \alpha_i g_i(G_0^-(a)) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow \int_{[F(a), F(b)]} \left(\sum_{i=1}^d \alpha_i g_i \circ G_0^- \circ F_0^- \right)^2 d\lambda + \left(\sum_{i=1}^d \alpha_i f_i(a) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \stackrel{(73)}{\Leftrightarrow} \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i f_i \cdot 1_{(a, b]} \right)^2 d\mu_F + \left(\sum_{i=1}^d \alpha_i f_i(a) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i f_i \cdot 1_{[a, b]} \right)^2 d\mu_F > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow f_1, \dots, f_d \text{ are l.i. in } L_2([a, b], F),
\end{aligned}$$

with $(F \circ G)_0 := (F \circ G)|_{[F(a), G(b)]}$ and $F_0 := F|_{[a, b]}$. According to Bischoff (1998), Theorem 2.2, $D_n \xrightarrow{\mathcal{D}} D$ follows with

$$\begin{aligned}
D(\omega, z) := & B(\omega, z) + \left(\int_{[a, b]} g d\mu_{H_z} \right)^\top \cdot \left(\int_{[a, b]} g g^\top d\mu_H \right)^{-1} \\
& \cdot \left(\int_{[a, b]}^{(R)} B(\omega, H) dg - B(\omega, 1)g(H^-(1)) \right),
\end{aligned}$$

whereby $H := F \circ G$ and $H_z(x) := \min(z, H(x))$. For all $q \in [F_z(a), z]$ a similar relation holds as in Lemma 57.10 (which we also use in the following):

$$\begin{aligned}
((F \circ G)_z)_0^-(q) &= \inf\{x \in [a, b] \mid \min(z, F \circ G(x)) \geq q\} \\
&= \inf\{x \in [a, b] \mid F \circ G(x) \geq q\} \\
&= (F \circ G)_0^-(q) = G_0^- \circ F_0^-(q),
\end{aligned} \tag{84}$$

whereby $((F \circ G)_z)_0 := ((F \circ G)_z)|_{[a,b]} : [a,b] \rightarrow [F_z(a), F_z(b)]$. Thus we get

$$\begin{aligned}
\int_{[a,b]} g \, d\mu_{(F \circ G)_z} &= \int_{(a,b)} g \, d\mu_{(F \circ G)_z} + g(a) \cdot (F \circ G)_z(a) \\
&\stackrel{(73)}{=} \int_{((F \circ G)_z(a), (F \circ G)_z(b))} g \circ ((F \circ G)_z)_0^- \, d\lambda + g(a) \cdot F_z(a) \\
&= \int_{(F_z(a), z)} g \circ (F \circ G)_0^- \, d\lambda + g(G_0^-(a)) \cdot F_z(a) \\
&\stackrel{(84)}{=} \int_{(F_z(a), z)} g \circ G_0^- \circ F_0^- \, d\lambda + f(a) \cdot F_z(a) \\
&= \int_{(F_z(a), z)} f \circ F_0^- \, d\lambda + f(a) \cdot F_z(a) \\
&= \int_{(0, z)} f \circ F_0^- \, d\lambda
\end{aligned} \tag{85}$$

For this note that $G_0^-(a) = a$ and $\int_{(0, F_z(a))} f \circ F_0^- \, d\lambda = f(a) \cdot F_z(a)$ since $F_0^-(q) = a$ for all $q \in [0, F_z(a)]$ and $F_z(0) = 0$. Further

$$\begin{aligned}
\int_{[a,b]} gg^\top \, d\mu_{F \circ G} &\stackrel{(73)}{=} \int_{(F(G(a)), F(G(b)))} (gg^\top) \circ (F \circ G)_0^- \, d\lambda + g(a)g(a)^\top \cdot F(G(a)) \\
&= \int_{(F(a), F(b))} (gg^\top) \circ G_0^- F_0^- \, d\lambda + g(G_0^-(a))g(G_0^-(a))^\top \cdot F(a) \\
&\stackrel{(73)}{=} \int_{(a,b)} ff^\top \, d\mu_F + f(a)f(a)^\top \cdot F(a) \\
&= \int_{[a,b]} ff^\top \, d\mu_F
\end{aligned} \tag{86}$$

and

$$g((F \circ G)^-(1)) = g(G^-(F^-(1))) = f(F^-(1)). \tag{87}$$

is valid. Finally,

$$\begin{aligned}
\int_{[a,b]}^{(R)} B(\omega, F \circ G) \, dg &\stackrel{(*)}{=} \int_{(a,b)} B(\omega, F \circ G) \, d\mu_g \\
&\stackrel{(**)}{=} \int_{[a,b]} B(\omega, F) \, d\mu_g^G \\
&\stackrel{(***)}{=} \int_{[a,b]} B(\omega, F) \, d\mu_f
\end{aligned} \tag{88}$$

is valid. For $(*)$ note that the Riemann-Stieltjes integral is equal to the corresponding Lebesgue-Stieltjes integral (cf. Kirillov and Gvishiani (1982), Theorem 14, p. 29 or Stroock (1994), Theorem 5.1.2). $(**)$ follows directly from the change of variable formula $\int_{\Omega'} \kappa \, d\mu^T = \int_{\Omega} (\kappa \circ T) \, d\mu$ for $\kappa : \Omega' \rightarrow \mathbb{R}$ and $T : \Omega \rightarrow \Omega'$, see, e.g., Stroock (1994), Lemma 5.0.1, where in our case $\kappa(t) := B(\omega, F(t))$ and $T(x) := G(x)$, $x \in \Omega$ holds, with $\Omega = (a, b)$ and $\Omega' = [a, b]$. For this, note that $T(\Omega) \subseteq \Omega'$ actually holds as well. For $(***)$ we show $\mu_g^G = \mu_f$ on $\mathcal{B}([a, b])$.

It suffices to show the equality on $\mathcal{E} := \{E \cap [a, b] \mid E \in \mathcal{E}_0\}$, where $\mathcal{E}_0 := \{[x, y] \mid x \leq y\}$ is a generator of $\mathcal{B}(\mathbb{R})$. The equality now follows from $\mathcal{E} = \{[x, y] \mid a \leq x \leq y \leq b\}$ and

$$\begin{aligned} \mu_g^G([x, y]) &= \mu_g(G^{-1}([x, y])) = \mu_g([G^-(x), G^-(y)]) = g(G^-(y)) - g(G^-(x)) \\ &= f(y^-) - f(x^-) \quad (\text{since } f \text{ is left continuous and } g \text{ is continuous}) \\ &= \mu_f([x, y]), \quad (\text{cf. Theorem 63}). \end{aligned}$$

Note that to define the measures we can canonically extend f (resp. g) by $f(x) := f(a)$ for $x < a$ and $f(x) := f(b)$ for $x > b$ (resp. analogously with g) left continuous (resp. continuous) to \mathbb{R} . Thus $D = B_{f,F}$, because of (85), (86), (87), and (88). \square

Proof of Theorem 21 from page 17:

We proceed as in Theorem 79 and Theorem 10 and their proofs. Note that

$$T_{n-d}(g(a))(x) = ((n-d)x - \lfloor (n-d)x \rfloor) a_{n-d-i_1+i_0-1} + \sum_{i=i_0}^{i_1} a_{n-d-i_1+i} \quad (89)$$

for $x \geq \frac{\lfloor (n-d)\delta \rfloor}{n-d}$ and $T_{n-d}(g(a))(x) = 0$ for $x < \frac{\lfloor (n-d)\delta \rfloor}{n-d}$, where $i_0 := \lfloor (n-d)\delta \rfloor + 1$ and $i_1 := \lfloor (n-d)x \rfloor$. Note also $\delta \in \left[\frac{\lfloor (n-d)\delta \rfloor}{n-d}, \frac{\lfloor (n-d)\delta \rfloor + 1}{n-d} \right)$. Analogous to (81), we define

$$\psi_n : C[0, 1] \rightarrow C[0, 1], \quad \psi_n := T_{n-d} \circ g \circ M_{n,n}^\top \circ V_n,$$

$$\begin{array}{ccc} \mathbb{R}^n & \xrightarrow{M_{n,n}^\top} & \mathbb{R}^{n-d} \\ V_n \uparrow & & \downarrow T_{n-d} \circ g \\ C[0, 1] & \xrightarrow{\psi_n} & C[0, 1] \end{array}$$

according to the digram, where $M_{n,n}$ is defined in (15) and T_{n-d} and g were defined in (19) and (46), respectively. Analogous to (83),

$$T_{n-d} \circ g \circ r_n \stackrel{(18)}{=} T_{n-d} \circ g \circ M_{n,n}^\top \circ \varepsilon_n \stackrel{(82)}{=} T_{n-d} \circ g \circ M_{n,n}^\top \circ V_n \circ T_n \circ \varepsilon_n = \psi_n \circ T_n \circ \varepsilon_n$$

is also valid. We now proceed exactly as in the proof of our main Theorem 79 from page 70. The arguments given there – slightly adapted – remain completely valid here as well. And here, too, we proceed in two steps.

Step 1, the regression functions f_i are first assumed to be continuous. We set $\xi_{n-d} := M_{n,n}^\top \circ V_n(u) \in \mathbb{R}^{n-d}$ and obtain the following computation

$$\begin{aligned} \psi_n(u)(x) &= T_{n-d} \circ g(M_{n,n}^\top \circ V_n(u))(x) = T_{n-d} \circ g(\xi_{n-d})(x) \\ &= \underbrace{((n-d)x - \lfloor (n-d)x \rfloor)}_{=:z} \cdot (\xi_{n-d})_{n-d-i_1+i_0} + \sum_{i=i_0}^{i_1} (\xi_{n-d})_{n-d-i_1+i} \quad \text{by (89)} \\ &= z + \sum_{i=i_0}^{i_1} \frac{1}{\sqrt{c_{n,n-i_1+i}}} \left[u\left(\frac{n-i_1+i}{n}\right) - u\left(\frac{n-i_1+i-1}{n}\right) \right] \\ &\quad - \sum_{i=i_0}^{i_1} \frac{1}{\sqrt{c_{n,n-i_1+i}}} \left[f(t_{n,n-i_1+i})^\top \left(X_{n,n-i_1+i-1}^\top X_{n,n-i_1+i-1} \right)^{-1} X_{n,n-i_1+i-1}^\top \cdot \begin{pmatrix} u(1/n) - u(0) \\ \vdots \\ u\left(\frac{n-i_1+i-1}{n}\right) - u\left(\frac{n-i_1+i-2}{n}\right) \end{pmatrix} \right] \\ &= z - \sum_{i=i_0}^{i_1-1} \left[\frac{1}{\sqrt{c_{n,n-i_1+i+1}}} - \frac{1}{\sqrt{c_{n,n-i_1+i}}} \right] u\left(\frac{n-i_1+i}{n}\right) + \frac{u(1)}{\sqrt{c_{n,n}}} - \frac{u\left(\frac{n-i_1+i_0-1}{n}\right)}{\sqrt{c_{n,n-i_1+i_0-1}}} \\ &\quad + \sum_{i=i_0}^{i_1} \frac{1}{n} \cdot \frac{1}{\sqrt{c_{n,n-i_1+i}}} f(t_{n,n-i_1+i})^\top \left(\frac{1}{n} X_{n,n-i_1+i-1}^\top X_{n,n-i_1+i-1} \right)^{-1} \left[(\Delta X_{n,n-i_1+i-1})^\top \cdot \begin{pmatrix} u(1/n) \\ \vdots \\ u\left(\frac{n-i_1+i-2}{n}\right) \end{pmatrix} + z_i^* \right] \end{aligned}$$

$$\begin{aligned}
& \text{where } z_i^* = f(t_{n,1})u(0) - f(t_{n,n-i_1+i-1})u\left(\frac{n-i_1+i-1}{n}\right) \\
& = z - \sum_{i=i_0}^{i_1-1} \left[\frac{1}{\sqrt{c_{n,n-i_1+i+1}}} - \frac{1}{\sqrt{c_{n,n-i_1+i}}} \right] u\left(\frac{n-i_1+i}{n}\right) + \frac{u(1)}{\sqrt{c_{n,n}}} - \frac{u\left(\frac{n-i_1+i_0-1}{n}\right)}{\sqrt{c_{n,n-i_1+i_0-1}}} + \int_{\frac{n-i_1+i_0-1}{n}}^1 \Theta_n(t) dt \\
& \xrightarrow{n \rightarrow \infty} 0 - 0 + u(1) - u(1-x+\delta) + \int_{1-x+\delta}^1 \Theta(t) dt, \text{ where} \\
& \Theta_n(t) \xrightarrow{n \rightarrow \infty} \Theta(t) := f(h(t))^\top \left(\int_{[0,t]} (ff^\top) \circ h d\lambda \right)^{-1} \left[\int_{[a,F^-(t)]}^{(R)} (u \circ F) df + f(h(0))u(0) - f(h(t))u(t) \right] \\
& \text{and } h := F_0^-. \text{ Exactly as in the proof of Theorem 79 we conclude}
\end{aligned}$$

$$\|\psi(u) - \psi_n(u_n)\| \xrightarrow{n \rightarrow \infty} 0$$

for all $u, u_n \in C[0, 1]$ with $\|u - u_n\| \xrightarrow{n \rightarrow \infty} 0$, where

$$\begin{aligned}
\psi(u)(x) & := u(1) - u(1-x+\delta) \\
& + \int_{1-x+\delta}^1 f(h(t))^\top \left(\int_{[0,t]} (ff^\top) \circ h d\lambda \right)^{-1} \left[\int_{[a,F^-(t)]}^{(R)} (u \circ F) df + f(h(0))u(0) - f(h(t))u(t) \right] dt
\end{aligned}$$

for $x \in [\delta, 1]$ and $\psi(u)(x) = 0$ for $x \in [0, \delta]$.

In the **2nd step** we proceed exactly as in the 2nd step from the proof of Theorem 79 (which can be taken practically unchanged). To complete the proof, we also mimic the procedure in the proof of Theorem 10 (which we can do without any changes), which then concludes this proof. \square

Proof of Theorem 23 from page 18

We can directly adopt the calculation (33) under the alternative (32) for this situation. Likewise, we can adopt Lemma 14 and finally the assertion follows exactly as in the proof of Theorem 15. We only need to replace ϕ with ψ and Γ with Δ in all places. For this, of course, note the proof of Theorem 21, since ψ was defined there. \square

Proof of Theorem 28 from page 20:

In Luschgy (1991), for a process

$$X_t = S(\theta, t) + Z_t, \quad 0 \leq t \leq T, \quad T < \infty,$$

where S is a known (nonrandom) function, Z is a zero-mean Gaussian process with known covariance function $K(s, t) = \text{Cov}(Z_s, Z_t)$ and $\theta \in \Theta$ is an unknown parameter with Θ a right open interval in \mathbb{R} , criteria are given that a uniformly most powerful test exists to test

$$H = \{\theta_0\} \quad \text{vs.} \quad K = \{\theta \in \Theta \mid \theta > \theta_0\}.$$

To be more concrete, the following conditions must be fulfilled.

(A.1) Z has continuous sample paths.

(A.2) $t \mapsto S(\theta, t)$ must be an element in the reproducing kernel Hilbert space generated by K .

(A.3) $\theta \mapsto S(\theta, t)$ is differentiable at θ_0 with derivative $\frac{\partial S(\theta, t)}{\partial \theta}(\theta_0) \neq 0$.

Furthermore, if there is a measure μ with

$$\int_{[0,T]} K(s,t)\mu(dt) = \frac{\partial S(\theta,s)}{\partial \theta}(\theta_0), \quad s \in [0,1], \quad (90)$$

the test can be given in the computable form (95) whereby in our case $T = 1 - \delta$, $Z_t = B_t^\bullet$, $\theta = \frac{\beta_{d+1}}{\sigma}$, $\theta_0 = 0$ and $S(\theta,t) = \theta \cdot \Delta^\bullet(t)$. Obviously, (A.1), (A.2), and (A.3) are satisfied. For (A.2), see also Paulsen and Raghupathi (2016), Theorem 11.3 and the discussion afterwards. It remains to show (90). To do this, we define a measure μ by the following relation.

$$\mu([0,t]) := \Xi'(t),$$

where Ξ is a differentiable function whose derivative Ξ' is of bounded variation. Obviously, $K(s,t) = \min(s,t)$. Thus,

$$\begin{aligned} \int_{[0,1-\delta]} K(s,t)\mu(dt) &= \int_{[0,1-\delta]} \min(s,t) \mu(dt) \\ &= \int_{[0,s]} t \mu(dt) + \int_{(s,1-\delta]} s \mu(dt) \\ &= \int_{[0,s]} t d(\Xi'(t)) + s \cdot \mu((s,1-\delta]) \\ &= s \cdot \Xi'(s) - \int_{[0,s]} \Xi'(t) dt + s \cdot (\Xi'(1-\delta) - \Xi'(s)) \\ &= \Xi(0) - \Xi(s) + s \cdot \Xi'(1-\delta) \end{aligned} \quad (91)$$

follows. For (90) to hold, (91) would have to lead to $\Delta^\bullet(s)$. In any case

$$\Delta^\bullet(0) = 0. \quad (92)$$

Moreover, for the first derivative of Δ from (51), we conclude

$$\begin{aligned} \Delta'(s) &= f_{d+1} \circ F_0^-(s_\delta) \\ &\quad + f(F_0^-(s_\delta))^\top \left(\int_{[0,s_\delta]} (ff^\top) \circ F_0^- d\lambda \right)^{-1} \left[\int_{[a,F_0^-(s_\delta)]} (u \circ F_0) d\mu_f - f(F_0^-(s_\delta))u(s_\delta) \right] \end{aligned}$$

and therefore for the first derivative of Δ^\bullet

$$\begin{aligned} (\Delta^\bullet)'(1-\delta) &= f_{d+1} \circ F_0^-(\delta) \\ &\quad + f(F_0^-(\delta))^\top \left(\int_{[0,\delta]} (ff^\top) \circ F_0^- d\lambda \right)^{-1} \left[\int_{[a,F_0^-(\delta)]} (u \circ F_0) d\mu_f - f(F_0^-(\delta))u(\delta) \right] \end{aligned}$$

with $s_\delta := 1 - s + \delta$ and $u(s) := \int_0^s f_{d+1} \circ F_0^- d\lambda$. But now $f_{d+1} \circ F_0^-(\delta) = 0$, $u(s) = 0$ for all $s \in [0, \delta]$ and $u \circ F_0(x) = 0$ for all $x \in [a, F_0^-(\delta)]$ follows, because $f_{d+1}(x) = 0$ and $F_0(x) \in [0, F_0(F_0^-(\delta))] = [0, \delta]$ holds for all $x \in [a, F_0^-(\delta)]$. Note $\delta = F(c) = F_0(c)$. Thus

$$(\Delta^\bullet)'(1-\delta) = 0. \quad (93)$$

An analogous relation for Γ cannot be derived under these conditions. For $\Xi(t) := -\Delta^\bullet(t)$,

$$\int_{[0,1-\delta]} K(s,t)\mu(dt) = \Delta^\bullet(s) = \frac{\partial S(\theta,s)}{\partial \theta}(\theta_0)$$

follows because of (91), (92) and (93). That is, (90) is verified. Following Luschgy (1991), for the test problem

$$H_0 : \theta = 0 \quad \text{vs.} \quad H_1 : \theta > 0, \quad (\text{observing } B'^{\bullet} + \theta \cdot \Delta^{\bullet}) \quad (94)$$

we obtain the test statistic $\rho : C[0, 1 - \delta] \rightarrow \mathbb{R}$ with

$$\begin{aligned} \rho(u) &:= \left(\int_{[0, 1 - \delta]} \int_{[0, 1 - \delta]} K(s, t) \mu(dt) \mu(ds) \right)^{-\frac{1}{2}} \int_{[0, 1 - \delta]} u(t) \mu(dt) \\ &= \left(- \int_{[0, 1 - \delta]} \Delta^{\bullet}(s) d(\Delta^{\bullet})'(t) \right)^{-\frac{1}{2}} \int_{[0, 1 - \delta]} u(t) d(\Delta^{\bullet})'(t), \quad u \in C[0, 1 - \delta]. \end{aligned}$$

Thus we get the test $\varphi := 1_{\rho^{-1}((\Phi^{-1}(1-\alpha), \infty))}$ with the decision rule

$$\text{reject } H_0 \text{ given in (94)} \quad :\Leftrightarrow \quad \varphi \left(B'^{\bullet} + \frac{\beta_{d+1}}{\sigma} \cdot \Delta^{\bullet} \right) = 1, \quad (95)$$

where $\Phi(x) := \frac{1}{\sqrt{2\pi}} \int_{-\infty}^x e^{-\frac{1}{2}t^2} dt$. If we set $\varphi_n := \varphi$, then (59) is satisfied. This follows from Theorem 46. To see this, we define ϕ to be the filter generated by $\{\varphi\}$. For the conditions of Theorem 46 to be satisfied, φ must be continuous almost everywhere. In fact, φ is continuous everywhere, because for $0 < \varepsilon < 1$ and $u \in C[0, 1 - \delta]$,

$$|\varphi(u) - \varphi(v)| < \varepsilon \quad \Leftrightarrow \quad [\rho(u) > \Phi^{-1}(1 - \alpha) \Leftrightarrow \rho(v) > \Phi^{-1}(1 - \alpha)]. \quad (96)$$

follows. From the definition of ρ it is clear that the right equivalence from (96) is satisfied for sufficiently small ε_0 and all $v \in C[0, 1 - \delta]$ with $\|u - v\| < \varepsilon_0$. Just as in the proof of Theorem 23, we conclude

$$\left(\frac{1}{\sigma\sqrt{n-d}} \cdot T_{n-d} \circ g \circ (M_{n,n}^{\top} Y_n^*) \right)^{\bullet} \xrightarrow{\mathcal{D}} B'^{\bullet} + \frac{\beta_{d+1}}{\sigma} \cdot \Delta^{\bullet}.$$

(59) thus follows from Theorem 46. □

Proof of Lemma 34 from page 23:

- 1 \Rightarrow 2: Let A be closed, so $A = \bigcap_{n=1}^{\infty} O_n$ with open O_n . By Urysohn's Lemma¹⁰ there exists a continuous $f_n : X \rightarrow [0, 1]$ with $A \subseteq f_n^{-1}(0)$ and $X \setminus O_n \subseteq f_n^{-1}(1)$. Then $f : X \rightarrow [0, 1]$, $f(x) := \sum_{n=1}^{\infty} 2^{-n} f_n(x)$ is continuous with $A = f^{-1}(0)$.
- 2 \Rightarrow 3: Let A, B closed and disjoint. So $A = f^{-1}(0)$ and $B = g^{-1}(0)$ for continuous $f, g : X \rightarrow [0, 1]$. Then $h : X \rightarrow [0, 1]$ defined by $h(x) := \frac{f(x)}{f(x)+g(x)}$ is continuous with $A = h^{-1}(0)$ and $B = h^{-1}(1)$.
- 3 \Rightarrow 1: Let A, B closed and disjoint. Let $f : X \rightarrow [0, 1]$ be continuous with $A = f^{-1}(0)$ and $B = f^{-1}(1)$. Then $U := f^{-1}[0, 1/2)$ and $V := f^{-1}((1/2, 1])$ are open and disjoint with $A \subseteq U$ and $B \subseteq V$ (i.e. X is normal). Moreover, $A = \bigcap_{n=1}^{\infty} O_n$, where $O_n := f^{-1}([0, 1/n))$ is open. □

¹⁰ (X, τ) is normal iff to every two closed and disjoint $A, B \subseteq X$ there exists a continuous mapping $f : X \rightarrow [0, 1]$, with $A \subseteq f^{-1}(0)$ and $B \subseteq f^{-1}(1)$, cf. Bourbaki (1966), Chapter IX, §4.1, Theorem 1.

Proof of Lemma 35 from page 24:

1. a), b) and c) are trivial. To show d), one distinguishes all possible cases (minus symmetry, six cases remain) of how two points can be distributed on the three sets F , $F_{<\varepsilon} := \{z \in F \mid d(z, F) < \varepsilon\} \setminus F$ and $F_{\geq\varepsilon} := \{z \in F \mid d(z, F) \geq \varepsilon\}$. In two of these cases, one has to use the inequality $d(x, F) \leq d(x, y) + d(y, F)$:

For $x, y \in F$ and $x, y \in F_{\geq\varepsilon}$, respectively, $|F_{F,\varepsilon}(x) - F_{F,\varepsilon}(y)| = 0 \leq \frac{1}{\varepsilon}d(x, y)$ follows.

For $x, y \in F_{<\varepsilon}$, $|F_{F,\varepsilon}(x) - F_{F,\varepsilon}(y)| = \frac{1}{\varepsilon}|d(x, F) - d(y, F)| \leq \frac{1}{\varepsilon}d(x, y)$ follows.

For $x \in F$ and $y \in F_{<\varepsilon}$, $|F_{F,\varepsilon}(x) - F_{F,\varepsilon}(y)| = \frac{1}{\varepsilon}d(y, F) \leq \frac{1}{\varepsilon}d(x, y)$ follows.

For $x \in F$ and $y \in F_{\geq\varepsilon}$, $|F_{F,\varepsilon}(x) - F_{F,\varepsilon}(y)| = 1 \leq \frac{1}{\varepsilon}d(x, y)$ follows.

For $x \in F_{<\varepsilon}$ and $y \in F_{\geq\varepsilon}$, $|F_{F,\varepsilon}(x) - F_{F,\varepsilon}(y)| = 1 - \frac{1}{\varepsilon}d(x, F) \leq \frac{1}{\varepsilon}d(x, y)$ follows.

2. Let $(A_n)_{n \in \mathbb{N}}$ be a sequence in τ with $\bigcap_{n \in \mathbb{N}} A_n = F$. Let $B_0 \in \tau$ with $F \subseteq B_0 \subseteq \bar{B}_0 \subseteq A_0$. If B_n is defined, let $B_{n+1} \in \tau$ with $F \subseteq B_{n+1} \subseteq \bar{B}_{n+1} \subseteq A_{n+1} \cap B_n$. By Urysohn's Theorem exists $f_n : E \rightarrow [0, 1]$ with $f_n(\bar{B}_{n+1}) \subseteq \{1\}$ and $f_n(E \setminus B_n) \subseteq \{0\}$. \square

Proof of Lemma 36 from page 24:

Define

$$\mathcal{A} := \{A \in \mathcal{B}(E) \mid \forall \varepsilon > 0 \exists F : \text{closed}, G : \text{open}, \text{ with } F \subseteq A \subseteq G \text{ and } P(G \setminus F) < \varepsilon\}.$$

We will show

1. $\alpha := \{A \subseteq E \mid A \text{ is closed}\} \subseteq \mathcal{A}$ and 2. \mathcal{A} is a σ -algebra

Therefore $\mathcal{B}(E) \subseteq \mathcal{A}$ because α generates $\mathcal{B}(E)$.

1. If $A \in \alpha$ and $\varepsilon > 0$ then

$$P(A) \leq P(B_n) \stackrel{n \text{ sufficiently large}}{\leq} P(A) + \varepsilon$$

$B_{n+1} \subseteq B_n, A = \bigcap_{n \in \mathbb{N}} B_n$

with B_n as in Lemma 35 ($F := A$ and $G := B_n$).

2. We have to show:

(i) $A_1, A_2 \in \mathcal{A} \Rightarrow A_1 \setminus A_2 \in \mathcal{A}$ and (ii) $A_n \in \mathcal{A}, n \in \mathbb{N} \Rightarrow \bigcup_{n \in \mathbb{N}} A_n \in \mathcal{A}$

(obviously $\emptyset \in \mathcal{A}$ and $E \in \mathcal{A}$)

- (i) Let $A_1, A_2 \in \mathcal{A}$ and $\varepsilon > 0$. Then there are F_1, F_2, G_1, G_2 with

$$\begin{aligned} F_1 \subseteq A_1 \subseteq G_1 \text{ and } P(G_1 \setminus F_1) < \varepsilon \\ F_2 \subseteq A_2 \subseteq G_2 \text{ and } P(G_2 \setminus F_2) < \varepsilon. \end{aligned}$$

This implies

$$\underbrace{F_1 \setminus G_2}_{:= F' \text{ closed}} \subseteq A_1 \setminus A_2 \subseteq \underbrace{G_1 \setminus F_2}_{:= G' \text{ open}}$$

with

$$\begin{aligned} P(G' \setminus F') &= P((G_1 \setminus F_2) \setminus (F_1 \setminus G_2)) \\ &= P([(G_1 \setminus (F_1 \cup F_2))] \cup [(G_1 \cap G_2) \setminus F_2]) < 2\varepsilon \end{aligned}$$

- (ii) For each $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ and $\varepsilon > 0$ choose some closed F_n and open G_n with $F_n \subseteq A_n \subseteq G_n$ and $P(G_n \setminus F_n) < 2^{-n-1}\varepsilon$. Let $N \in \mathbb{N}$ large enough with

$$P\left(\bigcup_{n \in \mathbb{N}} F_n\right) \leq P\left(\bigcup_{n \leq N} F_n\right) + \frac{\varepsilon}{2}.$$

We define

$$F := \bigcup_{n \leq N} F_n \quad (\text{closed}) \quad \text{and} \quad G := \bigcup_{n \in \mathbb{N}} G_n \quad (\text{open}).$$

Thus $F \subseteq \bigcup_{n \in \mathbb{N}} A_n \subseteq G$ and

$$P(G \setminus F) \leq P(G \setminus \bigcup_{n \in \mathbb{N}} F_n) + P\left(\bigcup_{n \in \mathbb{N}} F_n \setminus F\right) \leq \sum_{n \in \mathbb{N}} P(G_n \setminus F_n) + \frac{\varepsilon}{2} < \frac{\varepsilon}{2} + \frac{\varepsilon}{2} = \varepsilon. \quad \square$$

Proof of the equivalence of (64) and (65) from page 25:

(65) \Rightarrow (64) follows for $f = 1_A$. (64) \Rightarrow (65): Without loss of generality $f \geq 0$. Let $(e_k)_{k \in \mathbb{N}}$ be a sequence of simple funktions with $e_k \nearrow f$ and $\|f - e_k\|_{\text{sup}} \xrightarrow{n \rightarrow \infty} 0$ (cf. Elstrodt (2018), III, Korollar 4.14). From (64) and the definition of the integral for simple funktions, $\int e_k d\mu_n \xrightarrow{n \rightarrow \infty} \int e_k d\mu$ follows ($\forall k$). This leads to the folowing attempt:

$$\begin{aligned} \left| \int f d\mu_n - \int f d\mu \right| &\leq \left| \int f d\mu_n - \int e_k d\mu_n \right| + \left| \int e_k d\mu_n - \int e_k d\mu \right| + \left| \int e_k d\mu - \int f d\mu \right| \\ &\leq \underbrace{\|f - e_k\|_{\text{sup}}}_{\xrightarrow{k \rightarrow \infty} 0} \cdot \underbrace{\mu_n(\Omega)}_{\xrightarrow{n \rightarrow \infty} \mu(\Omega)} + \underbrace{\left| \int e_k d\mu_n - \int e_k d\mu \right|}_{\xrightarrow{n \rightarrow \infty} 0 \quad \forall k} + \underbrace{\|f - e_k\|_{\text{sup}}}_{\xrightarrow{k \rightarrow \infty} 0} \cdot \underbrace{\mu(\Omega)}_{\xrightarrow{n \rightarrow \infty} \mu(\Omega)} \end{aligned}$$

This Inequality give the idea for the proof. Let $\varepsilon > 0$. Let $N_1 \in \mathbb{N}$ such that $\forall n \geq N_1: \mu_n(\Omega) \leq 1 + \mu(\Omega)$. Let $k \in \mathbb{N}$ with $\|f - e_k\|_{\text{sup}} \cdot (1 + 2\mu(\Omega)) < 2\varepsilon/3$. And let $N_2 \geq N_1$ such that $\forall n \geq N_2: |\int e_k d\mu_n - \int e_k d\mu| < \varepsilon/3$. Then $\forall n \geq N_2: |\int f d\mu_n - \int f d\mu| < \varepsilon$. \square

Proof of Lemma 38 from page 26:

Let F be a closed set and $(f_n)_{n \in \mathbb{N}}$ as in Lemma 35. Then

$$\int_E f_n dP = \int_E f_n dQ \quad (\text{for all } n \in \mathbb{N})$$

because of $\int_E f_n d\psi \xrightarrow{\tau_{\mathbb{R}}} \int_E f_n dP$ and $\int_E f_n d\psi \xrightarrow{\tau_{\mathbb{R}}} \int_E f_n dQ$ and $\int_E f_n d\psi$ is a filter on \mathbb{R} (a Hausdorff Space). Now let $\varepsilon > 0$ arbitrary. Then

$$Q(F) = \int_E 1_F dQ \leq \int_E f_n dQ = \int_E f_n dP \stackrel{\text{for } n \text{ sufficiently large}}{\leq} \int_E 1_F dP + \varepsilon = P(F) + \varepsilon$$

by the dominated convergence theorem. This implies $Q(F) \leq P(F)$ because $\varepsilon > 0$ is arbitrary. Analogously $P(F) \leq Q(F)$ and therefore $P(F) = Q(F)$. But $\{F \subseteq E \mid F \text{ is closed}\}$ is a generator of $\mathcal{B}(E)$ and therefore $P(A) = Q(A)$ for all $A \in \mathcal{B}(E)$. \square

Proof of Lemma 39 from page 26:

1. a) $P_1, P_2 \in \psi$ implies $\inf P_1 \leq \sup P_2$ (otherwise $P_1 \cap P_2 = \emptyset$). Therefore

$$\sup\{\inf P \mid P \in \psi\} \leq \inf\{\sup P \mid P \in \psi\}.$$

- b) $x \in \bigcap_{P \in \psi} \bar{P}$ implies $\inf P \leq x \leq \sup P$, for all $P \in \psi$.

2. “ \Rightarrow ” With $\psi_0 := \{(x - \varepsilon, x + \varepsilon) \mid \varepsilon > 0\}$ follows

$$\begin{aligned} \sup\{\inf P \mid P \in \psi\} &\leq \inf\{\sup P \mid P \in \psi\} \\ &\leq \inf\{\sup P \mid P \in \psi_0\} \\ &= x = \sup\{\inf P \mid P \in \psi_0\} \\ &\leq \sup\{\inf P \mid P \in \psi\} \end{aligned}$$

“ \Leftarrow ” Let $\varepsilon > 0$. $x = \sup\{\inf P \mid P \in \psi\}$ implies $\exists P_1 \in \psi$ with $\inf P_1 \in (x - \varepsilon, x + \varepsilon)$.
 $x = \inf\{\sup P \mid P \in \psi\}$ implies $\exists P_2 \in \psi$ with $\sup P_2 \in (x - \varepsilon, x + \varepsilon)$. Therefore $P_3 := P_1 \cap P_2 \subseteq (x - \varepsilon, x + \varepsilon)$

3. a) “ \leq ” because $\psi_0 \subseteq \psi$, “ \geq ” because for $P \in \psi$ exists $P_0 \in \psi_0$ with $P_0 \subseteq P$ and therefore $\inf P \leq \inf P_0$.

- b) follows analogously. □

Proof of Theorem 40 from page 26:

We use Lemma 35 and the dominated convergence theorem.

- 1 $\xrightarrow[\text{metrizable}]{\text{if } E \text{ is}}$ 2: Trivial.

- 3 \Rightarrow 1: Trivial.

- 4 \Leftrightarrow 5: By complements.

- (1 \vee 2) \Rightarrow 4: Let F closed $\subseteq E$, $(f_n)_{n \in \mathbb{N}}$ as in Lemma 35 (if E is metrizable $f_n := f_{F, \frac{1}{n}}$) and $\varepsilon > 0$.
 Then

$$\begin{aligned} \inf\{\sup Q(F) \mid Q \in \psi\} &= \inf \left\{ \sup \left\{ \int_E 1_F dq \mid q \in Q \right\} \mid Q \in \psi \right\} \\ &\leq \inf \left\{ \sup \left\{ \int_E f_n dq \mid q \in Q \right\} \mid Q \in \psi \right\} \\ &= \int_E f_n dP \quad \text{by 1. or 2. and Lemma 39} \\ &\leq \int_E 1_F dP + \varepsilon = P(F) + \varepsilon \quad \text{for } n \text{ large enough.} \end{aligned}$$

Therefore $\inf\{\sup Q(F) \mid Q \in \psi\} \leq P(F)$, because ε was arbitrary.

$(4 \wedge 5) \Rightarrow 6$: follows from Lemma 39 and

$$\begin{aligned}
\sup\{\inf Q(A^\circ) \mid Q \in \Psi\} &\leq \sup\{\inf Q(A) \mid Q \in \Psi\} \\
&\leq \inf\{\sup Q(A) \mid Q \in \Psi\} \\
&\leq \inf\{\sup Q(\bar{A}) \mid Q \in \Psi\} \\
&\leq P(\bar{A}) \\
&= P(A^\circ) + P(\partial A) = P(A^\circ) \quad (= P(A)) \\
&\leq \sup\{\inf Q(A^\circ) \mid Q \in \Psi\}
\end{aligned}$$

$6 \Rightarrow 3$: Let $f : E \rightarrow \mathbb{R}$ measurable, bounded and nonconstant with $P(D_f) = 0$.

Claim: $\forall A \subseteq \mathbb{R}: \partial f^{-1}(A) \subseteq f^{-1}(\partial A) \cup D_f$.

Proof: Let $x \in \partial f^{-1}(A)$. If $x \notin D_f$ let V be open with $f(x) \in V$. Then \exists open U with $x \in U$ and $f(U) \subseteq V$. Let $y \in U \cap f^{-1}(A)$. Then $f(y) \in f(U) \cap A$ and so $V \cap A \neq \emptyset$. This means $f(x) \in \bar{A}$. If $f(x) \in A^\circ$, then \exists open W with $x \in W$ and $f(W) \subseteq A^\circ$. Therefore $x \in W \subseteq f^{-1}(A^\circ) \subseteq f^{-1}(A)$ in contradiction to $x \in \partial f^{-1}(A)$.

Let $\varepsilon > 0$. $A := \{y \in \mathbb{R} \mid P(f^{-1}(\{y\})) > 0\}$ is at most countable. Therefore $\overline{\mathbb{R} \setminus A} = \mathbb{R}$, so we can choose $y_0, y_1, \dots, y_N \in \mathbb{R} \setminus A$ with

$$a) \ y_0 < -\sup_{x \in E} |f(x)| < y_1 < \dots < y_{N-1} < \sup_{x \in E} |f(x)| < y_N \quad \text{and}$$

$$b) \ y_{i+1} - y_i < \varepsilon \text{ for each } i \in \{0, \dots, N-1\}$$

Now $E = \biguplus_{i=1}^N E_i$ with $E_i := f^{-1}([y_{i-1}, y_i])$ and

$$\begin{aligned}
P(\partial E_i) &\stackrel{\text{Claim}}{\leq} \underbrace{P(f^{-1}(\partial [y_{i-1}, y_i]))}_{=P(f^{-1}(y_{i-1})) + P(f^{-1}(y_i)) = 0} + \underbrace{P(D_f)}_{=0} = 0, \quad i \in \{1, \dots, N\}.
\end{aligned}$$

Thus

$$\begin{aligned}
\inf \left\{ \sup \int_E f dQ \mid Q \in \Psi \right\} &= \inf \left\{ \sup \left\{ \sum_{i=1}^N \int_{E_i} f dq \mid q \in Q \right\} \mid Q \in \Psi \right\} \\
&\stackrel{(*)}{\leq} \sum_{i=1}^N \inf \left\{ \sup \left\{ \int_{E_i} f dq \mid q \in Q \right\} \mid Q \in \Psi \right\} \\
&\leq \sum_{i=1}^N \inf \{ \sup Q(E_i) y_i \mid Q \in \Psi \} \\
&= \sum_{i=1}^N [\inf \{ \sup Q(E_i) \mid Q \in \Psi \}] y_i \\
&= \sum_{i=1}^N y_i P(E_i) \leq \sum_{i=1}^N P(E_i) (y_{i-1} + \varepsilon) \\
&\leq \sum_{i=1}^N \int_{E_i} f dP + \varepsilon = \int_E f dP + \varepsilon, \tag{97}
\end{aligned}$$

by Lemma 39. Proof of (*): Let $t > 0$ arbitrary and define

$$s := \inf \left\{ \sup \left\{ \sum_{i=1}^N \int_{E_i} f d q \mid q \in \mathcal{Q} \right\} \mid \mathcal{Q} \in \psi \right\} \quad \text{and}$$

$$s' := \sum_{i=1}^N \inf \left\{ \sup \left\{ \int_{E_i} f d q \mid q \in \mathcal{Q} \right\} \mid \mathcal{Q} \in \psi \right\}$$

Let $Q_1, \dots, Q_N \in \psi$ with $s' \leq \sum_{i=1}^N \sup \{ \int_{E_i} f d q \mid q \in Q_i \} < s' + t$. Then

$$s \leq \sup \left\{ \sum_{i=1}^N \int_{E_i} f d q \mid q \in \mathcal{Q}' \right\} \leq \sum_{i=1}^N \sup \left\{ \int_{E_i} f d q \mid q \in Q_i \right\} < s' + t$$

with $\mathcal{Q}' := \bigcap_{i=1}^N Q_i$ and therefore $s \leq s'$ because t was arbitrary. This proves (*).

(97) implies

$$\inf \left\{ \sup \int_E f d Q \mid \mathcal{Q} \in \psi \right\} \leq \int_E f d P \quad (98)$$

because ε was arbitrary. By substitute f through $-f$ we get the analogous inequality to (98)

$$\sup \{ \inf \int_E f d Q \mid \mathcal{Q} \in \psi \} \geq \int_E f d P.$$

These two inequalities together with Lemma 39 implie $\int_E f d \psi \xrightarrow{\tau_{\mathbb{R}}} \int_E f d P$ \square

Proof of Theorem 41 from page 27:

We will use Theorem 40.5. Let $G \subseteq E$ be an open set. There exists $A_i \in \mathcal{U}$, $i = 1, 2, \dots$ with $G = \bigcup_{i=1}^{\infty} A_i$, especially $P(\bigcup_{i=1}^n A_i) \xrightarrow[n \rightarrow \infty]{\tau_{\mathbb{R}}} P(G)$. Now let $\varepsilon > 0$. Then there exists $n \geq 1$ with $P(G) - \varepsilon \leq P(\bigcup_{i=1}^n A_i)$.

Claim: $\psi(\bigcup_{i=1}^n A_i) \xrightarrow{\tau_{\mathbb{R}}} P(\bigcup_{i=1}^n A_i)$

Proof of the claim: Let $L := \mathcal{P}(\{1, \dots, n\}) \setminus \{\emptyset\} = \{J_1, \dots, J_s\}$, $f : \mathbb{R}^s \rightarrow \mathbb{R}$, $f(x_1, \dots, x_s) := \sum_{r=1}^s (-1)^{|J_r|+1} x_r$ and

$$\phi := \left\{ R \subseteq \mathbb{R}^s \mid \exists Q_1, \dots, Q_s \in \psi \text{ with } \prod_{r=1}^s Q_r(\bigcap_{i \in J_r} A_i) \subseteq R \right\},$$

the productfilter of $\psi(\bigcap_{i \in J_r} A_i)$, $r = 1, \dots, s$. Thus $\phi \xrightarrow{\tau_{\mathbb{R}^s}} (P(\bigcap_{i \in J_1} A_i), \dots, P(\bigcap_{i \in J_s} A_i))$ by assumption. Since f is continuous,

$$f(\phi) \xrightarrow{\tau_{\mathbb{R}}} f(P(\bigcap_{i \in J_1} A_i), \dots, P(\bigcap_{i \in J_s} A_i)) = \sum_{r=1}^s (-1)^{|J_r|+1} P(\bigcap_{i \in J_r} A_i) = P(\bigcup_{i=1}^n A_i)$$

follows. On the other hand $f(\phi) \subseteq \psi(\bigcup_{i=1}^n A_i)$. To see this let $R \in \phi$. Then there are $Q_1, \dots, Q_s \in \psi$ with $Q_1(\bigcap_{i \in J_1} A_i) \times \dots \times Q_s(\bigcap_{i \in J_s} A_i) \subseteq R$. Thus

$$f(R) \supseteq f(Q(\bigcap_{i \in J_1} A_i) \times \dots \times Q(\bigcap_{i \in J_s} A_i)) \supseteq Q(\bigcup_{i=1}^n A_i), \quad (99)$$

where $Q := \bigcap_{r=1}^s Q_r \in \psi$ and the last inclusion relation in (99) holds by the inclusion-exclusion principle. This proves $f(\phi) \subseteq \psi(\bigcup_{i=1}^n A_i)$ and therefore the claim. With Lemma 39 we conclude

$$P(G) - \varepsilon \leq \sup\{\inf \mathcal{R} \mid \mathcal{R} \in \psi(\bigcup_{i=1}^n A_i)\} = \sup\{\inf Q(\bigcup_{i=1}^n A_i) \mid Q \in \psi\} \leq \sup\{\inf Q(G) \mid Q \in \psi\}$$

and therefore $P(G) \leq \sup\{\inf Q(G) \mid Q \in \psi\}$ because ε was arbitrary. \square

Proof of Theorem 42 from page 27:

“ \Leftarrow ”: We will use Theorem 40.2. Let $f : \mathbb{R} \rightarrow [0, 1]$ be Lipschitz-continuous with constant $L > 0$. We need to show $\int_{\mathbb{R}} f d\psi \rightarrow \int_{\mathbb{R}} f dp$. To do this we use Lemma 39 and show

$$\inf \sup \int_{\mathbb{R}} f d\psi \leq \int_{\mathbb{R}} f dp. \quad (100)$$

By analogous reasoning, or alternatively going over to $1 - f$ and use (100) again, we get $\sup \inf \int_{\mathbb{R}} f d\psi \geq \int_{\mathbb{R}} f dp$ and we are done. To show (100), let $\varepsilon > 0$, $N \in \mathbb{N}$, $a := p(\mathbb{R})$ and $y_0 < \dots < y_N$ with $y_0, \dots, y_N \in c(F(p))$ with $F(p)(y_0) < \varepsilon$, $F(p)(y_N) > a - \varepsilon$ and $y_i - y_{i-1} < \varepsilon$ for all $i \in \{1, \dots, N\}$. Define $f(y_i^*) = \sup_{y \in [y_{i-1}, y_i]} f(y)$ and $f(y_{i*}) = \inf_{y \in [y_{i-1}, y_i]} f(y)$. For any $x \in c(F(p))$, $\inf \sup F(\psi)(x) = F(p)(x) = \sup \inf F(\psi)(x)$ holds by lemma 39. Let $r := \max\{1, \sum_{i=1}^N f(y_i)\}$. For every $x \in c(F(p))$ there is a $Q_x \in \psi$ such that

$$F(p)(x) - \varepsilon/r \leq F(q)(x) \leq F(p)(x) + \varepsilon/r$$

holds for all $q \in Q_x$. Let $Q' := Q \cap Q_{y_0} \cap \dots \cap Q_{y_N} \in \psi$. For each $q \in Q'$,

$$\begin{aligned} \int_{\mathbb{R}} f dq &\leq q((-\infty, y_0]) + q([y_N, \infty)) + \sum_{i=1}^N f(y_i^*) q([y_{i-1}, y_i]) \\ &\leq F(q)(y_0) + a - F(q)(y_N) + \sum_{i=1}^N (f(y_i) + L \cdot \varepsilon) q([y_{i-1}, y_i]) \\ &\leq F(p)(y_0) + a - F(p)(y_N) + (2 + L) \varepsilon + \sum_{i=1}^N f(y_i) (F(q)(y_i) - F(q)(y_{i-1})) \\ &\leq (4 + L) \varepsilon + \sum_{i=1}^N f(y_i) (F(p)(y_i) - F(p)(y_{i-1})) + 2\varepsilon/r \\ &\leq (6 + L) \varepsilon + \sum_{i=1}^N f(y_i) (F(p)(y_i) - F(p)(y_{i-1})) \\ &\leq (6 + L) \varepsilon + \sum_{i=1}^N (f(y_{i*}) + L \cdot \varepsilon) p([y_{i-1}, y_i]) \\ &\leq (6 + 2L) \varepsilon + \sum_{i=1}^N f(y_{i*}) p([y_{i-1}, y_i]) \\ &\leq (6 + 2L) \varepsilon + \int_{\mathbb{R}} f dp \end{aligned}$$

follows. Thus (100) because ε was arbitrary.

“ \Rightarrow ”: Let $x \in c(F(p))$. Then $p(\partial(-\infty, x]) = p(\{x\}) \stackrel{x \in c(F(p))}{=} 0$. Thus

$$F(\psi)(x) = \psi((-\infty, x]) \rightarrow p((-\infty, x]) = F(p)(x)$$

follows by Theorem 40.6. □

Proof of Lemma 45 from page 28:

1. \Rightarrow 2: Let $x_n \rightarrow x$, φ the filter associate with $(x_n)_{n \in \mathbb{N}}$ and ψ the filter associate with $(f_n)_{n \in \mathbb{N}}$.

From $\varphi \xrightarrow{\tau} x$ follows $\psi(\varphi) \xrightarrow{\sigma} f(x)$. This means for all $U \in f(x) \cap \sigma$ exists $P \in \psi$, $Q \in \varphi$ with $P(Q) \subseteq U$. So there is $N \in \mathbb{N}$ with $f_n(x_m) \in U$ for all $n, m \geq N$. For $n \geq N$ follows $f_n(x_n) \in U$. This direction works without a cnb.

2. \Rightarrow 1: Let ψ the filter associate with $(f_n)_{n \in \mathbb{N}}$, $x \in X$ and $\varphi \xrightarrow{\tau} x$. We show $\psi(\varphi) \xrightarrow{\sigma} f(x)$. Let

$(A_n)_{n \in \mathbb{N}}$ a cnb of x . If not $\psi(\varphi) \xrightarrow{\sigma} f(x)$ then $\exists U \in f(x) \cap \sigma \forall P \in \varphi \forall k \in \mathbb{N} \exists n_k \geq k \exists x_k \in P$ with $f_{n_k}(x_k) \notin U$

To A_0 and 0 exists $n_0 \geq 0$ and $x_0 \in A_0$ with $f_{n_0}(x_0) \notin U$.

To A_{n_k+1} and $n_k + 1$ exists $n_{k+1} \geq n_k + 1$ and $x_{k+1} \in A_{n_{k+1}}$ with $f_{n_{k+1}}(x_{k+1}) \notin U$.

So there is a strictly monotone sequence $(n_l)_{l \in \mathbb{N}}$ in \mathbb{N} , a sequence $(x_l)_{l \in \mathbb{N}}$ in X with $x_l \rightarrow x$ and $f_{n_l}(x_l) \notin U$. For $n \in \mathbb{N}$ let $l(n) := \min\{l \in \mathbb{N} \mid n \leq n_l\}$ and $x'_n := x_{l(n)}$. Obviously $x'_n \rightarrow x$ but not $f_{n'}(x'_n) \rightarrow f(x)$. A contradiction! □

Proof of Theorem 46 from page 28:

We use Theorem 40.5. Let G be open in F . Let $\beta = \{B_n \mid n \in \mathbb{N}\}$ be a countable basis of ϕ . Without restriction of generality, $B_{n+1} \subseteq B_n$ holds for all n . Then $h^{-1}(G) \subseteq D \cup \bigcup_{B \in \beta} T_B^\circ$ holds with $T_B := \bigcap_{f \in B} f^{-1}(G)$. [Proof: Let $x \in h^{-1}(G)$. If $x \notin D$, then $\exists B' \in \beta \exists V \in \dot{x} \cap \tau$ with $B'(V) \subseteq G$. Therefore $V \subseteq \bigcap_{f \in B'} f^{-1}(G)$ and so $x \in (\bigcap_{f \in B'} f^{-1}(G))^\circ \subseteq \bigcup_{B \in \beta} T_B^\circ$.] Now let $\varepsilon > 0$ be arbitrary but fixed. From $T_{B_k} \subseteq T_{B_{k+1}}$ follows $P(T_{B_k}^\circ) \xrightarrow{k \rightarrow \infty} P(\bigcup_{B \in \beta} T_B^\circ) \leq 1$. So there $\exists B' \in \beta$ with $P(\bigcup_{B \in \beta} T_B^\circ) \leq P(T_{B'}^\circ) + \varepsilon$. Therefore

$$\begin{aligned} P(h^{-1}(G)) &\leq P\left(\bigcup_{B \in \beta} T_B^\circ\right) \\ &\leq P(T_{B'}^\circ) + \varepsilon \\ &\leq \sup \{ \inf \{ Q(T_{B'}^\circ) \} \mid Q \in \psi \} + \varepsilon, \quad \text{by Theorem 40.5} \\ &\stackrel{(*)}{\leq} \sup \{ \inf \{ Q^{B'}(G) \} \mid Q \in \psi \} + \varepsilon \\ &= \sup \{ \inf \{ R(G) \} \mid R \in \Phi(\psi) \} + \varepsilon, \quad \text{by Lemma 39.} \end{aligned}$$

For $(*)$ note $Q(T_{B'}^\circ) = \{q(\bigcap_{f \in B'} f^{-1}(G)) \mid q \in Q\}$ and $Q^{B'}(G) = \{q(f^{-1}(G)) \mid q \in Q, f \in B'\}$. Thus $P^h(G) \leq \sup \{ \inf \{ R(G) \} \mid R \in \Phi \}$, because $\varepsilon > 0$ was arbitrary. □

Proof of Lemma 47 from page 29

1. Note that for idempotent $B \in \mathbb{R}^{n \times n}$, $\text{im}(B) \oplus \ker(B) = \mathbb{R}^n$ holds and thus $B = \text{pr}_{U,V}$ is satisfied, where $U := \text{im}(B)$ and $V := \ker(B)$. Since $I_n - B$ is idempotent to and because of $\ker(I_n - B) = \text{im}(B)$ and $\text{im}(I_n - B) = \ker(B)$ it follows $I_n - B = \text{pr}_{V,U}$. Since with B also B^\top is idempotent, it follows $B^\top = \text{pr}_{U',V'}$, where $U' := \text{im}(B^\top)$ and $V' := \ker(B^\top)$. Since $\ker(B^\top) = (\text{im}(B))^\perp$ and $\ker(B) = (U')^\perp$ holds, it follows $B^\top = \text{pr}_{V^\perp, U^\perp}$.
2. The first assertion follows from the easily provable equivalence $\text{pr}_{U, U^\perp}(y) = Xw \Leftrightarrow X^\top Xw = X^\top y$. The second assertion follows from the equivalence $\text{rank}(X) = d \Leftrightarrow X^\top X$ is invertible. For the proof of this second equivalence note $\ker(X^\top X) = \ker(X)$ and $\text{im}(X^\top X) = \text{im}(X^\top)$.
3. Let $P := MM^\top$. Then $PP = MM^\top MM^\top = MM^\top = P$. Thus $P = \text{pr}_{K,L}$ and $P^\top = \text{pr}_{L^\perp, K^\perp}$ with $K = \text{im}(P)$ and $L = \ker(P)$. But $P^\top = P$ and so $K^\perp = L$. Furthermore $\dim(U) = \text{rank}(X) \geq k - \text{rank}(M^\top) = \dim(\ker(M^\top))$ and $U \subseteq \ker(M^\top)$ since $M^\top X = 0$. Thus $U = \ker(M^\top) = \ker(MM^\top) = L$ and $K = (K^\perp)^\perp = L^\perp = U^\perp$. \square

Proof of Lemma 55 from page 31:

2. \Rightarrow 1. is clear. Let us show 1. \Rightarrow 2. Let $(t_n)_{n \in \mathbb{N}}$ be a sequence in $D \cap (-\infty, t)$ with $t_n \rightarrow t$. Assume $f(t_n) \not\rightarrow f(t)$. Then there exists $\varepsilon > 0$ and a subsequence $(t_{n_k})_{k \in \mathbb{N}}$ with $|f(t_{n_k}) - f(t)| \geq \varepsilon$ for all $k \in \mathbb{N}$. Since every sequence of real numbers contains a monotone subsequence, we can assume without restriction that $(t_{n_k})_{k \in \mathbb{N}}$ is already monotone. Because of $t_{n_k} \leq t$ and $t_{n_k} \rightarrow t$, $(t_{n_k})_{k \in \mathbb{N}}$ is therefore nondecreasing. Thus $f(t_{n_k}) \rightarrow f(t)$ holds in contradiction to $|f(t_{n_k}) - f(t)| \geq \varepsilon$ for all $k \in \mathbb{N}$. \square

Proof of Lemma 57 from page 32:

1. $q, q' \in Q_F$ with $q \leq q'$ implies $F^-(q) = \inf\{t \in I \mid q \leq F(t)\} \leq \inf\{t \in I \mid q' \leq F(t)\} = F^-(q')$, since $\{t \in I \mid q' \leq F(t)\} \subseteq \{t \in I \mid q \leq F(t)\}$.
2. Let $t := F^-(q)$ and let $(t_n)_{n \in \mathbb{N}}$ be a sequence in $\{t' \in I \mid F(t') \geq q\}$ with $t_n \searrow t$. It follows $F(t) = \lim_{n \rightarrow \infty} F(t_n) \geq q$. Thus $t \in \{t' \in I \mid F(t') \geq q\}$ and $F^-(q) = \min\{t' \in I \mid F(t') \geq q\}$.
3. Follows directly from 2.
4. $F^-(F(t)) = \inf\{t' \in I \mid F(t') \geq F(t)\} \leq t$ since $t \in \{t' \in I \mid F(t') \geq F(t)\}$.
5. $F^-(q) = \inf\{t' \in I \mid F(t') \geq q\} \leq t$ since $t \in \{t' \in I \mid F(t') \geq q\}$.
6. $t \geq F^-(q) \Rightarrow F(t) \geq F(F^-(q)) \stackrel{3.}{\geq} q$
7. Let $t \in I$ with $q = F(t)$. Therefore $F^-(q) \leq t$ and thus $q \leq F(F^-(q)) \stackrel{3.}{\leq} F(t)$.
8. We assume that F is not constant (otherwise $|Q_F| = 1$). Let $q \in Q_F \setminus \{Q_F^-\}$, $t := F^-(q)$ and $(q_n)_{n \in \mathbb{N}}$ be a sequence in Q_F with $q_n \nearrow q$. Since $q \in Q_F \setminus \{Q_F^-\}$ there exists $\delta > 0$ with $q - \delta \in Q_F \setminus \{Q_F^-\}$. Thus there is $a \in I$ with $F(a) \leq q - \delta < q - \frac{\delta}{2}$. This implies $F(a) < q - \frac{\delta}{2}$, $a < F^-(q - \frac{\delta}{2})$, $F^-(q - \frac{\delta}{2}) \in I \setminus \{I^-\}$ and $t \in I \setminus \{I^-\}$. Let $\varepsilon > 0$ with

$t - \varepsilon \in I$. Since $F(t - \varepsilon) < q$, there is an $N \in \mathbb{N}$ with $F(t - \varepsilon) < q_n \leq q$ and $F^-(q_n) \in I \setminus \{I^-\}$ for all $n \geq N$. From 6. and 1. we deduce $t - \varepsilon < F^-(q_n) \leq F^-(q)$ for all $n \geq N$. Thus $F^-(q_n) \rightarrow F^-(q)$, since ε was arbitrary.

9. $F^-(F(t)) = \inf\{t' \in I \mid F(t) \leq F(t')\} = t$ for all $t \in I$, since F is strictly increasing. Let $q = F(t) \in B$. Thus $F^-(q) = t$ and $F(F^-(q)) = F(t) = q$. Hence $(F|_B^-) \circ F = \text{id}$ and $F \circ (F|_B^-) = \text{id}$, i.e. $F|_B^- = F^{-1}$.

10. Applying 5. several times yields $(G \circ F)^-(q) = \inf\{t \in I \mid q \leq G(F(t))\} = \inf\{t \in I \mid G^-(q) \leq F(t)\} = \inf\{t \in I \mid F^-(G^-(q)) \leq t\} = F^-(G^-(q))$. \square

Proof of Lemma 60 from page 32:

1. $P(F^-(U) \leq t) = P(\{\omega \mid F^-(U(\omega)) \leq t\}) \stackrel{57}{=} P(\{\omega \mid U(\omega) \leq F(t)\}) = F(t)$.

2. Let us first show (70). $P(F(X) \leq 1) = 1$ is obvious. So let $r \in [0, 1)$, $M := \{t \mid F(t) \leq r\}$ and $t_0 := \sup M$, if $M \neq \emptyset$. We distinguish the cases $r \in F(\mathbb{R})$ and $r \notin F(\mathbb{R})$.

$r \notin F(\mathbb{R})$: If $r = 0$, $M = \emptyset$ and $P(F(X) \leq r) = P(\{\omega \mid X(\omega) \in M\}) = P(\emptyset) = 0$ follows. If $r > 0$ then $M \neq \emptyset$. Assume $t_0 \in M$. Since $r \notin F(\mathbb{R})$, $M = \{t \mid F(t) < r\}$ follows. Thus $F(t_0) < r$. But F is right continuous, so $F(t_0) = \lim_{x \searrow t_0} F(x) \geq r$ (since $F(x) > r$ for all $x > t_0$). A contradiction. Thus $t_0 \notin M$. Therefore $t_0 = F^-(r)$ and

$$\begin{aligned} P(F(X) \leq r) &= P(\{\omega \mid X(\omega) \in M\}) = P(\{\omega \mid X(\omega) \in (-\infty, t_0)\}) \\ &= P(X < t_0) = \lim_{x \nearrow t_0} F(x) = \lim_{x \nearrow F^-(r)} F(x). \end{aligned}$$

$r \in F(\mathbb{R})$: In this case we do not need to consider $r = 0$ separately. If $t_0 \in M$, $P(F(X) \leq r) = P(\{\omega \mid X(\omega) \in M\}) = F(t_0) = r$ follows, since $r \in F(\mathbb{R})$. If $t_0 \notin M$, $F(t_0) > r$ follows. Let $t_1 < t_0$ with $F(t_1) = r$. Hence $F(t) = r$ for all $t \in [t_1, t_0)$ and thus

$$\begin{aligned} P(F(X) \leq r) &= P(\{\omega \mid X(\omega) \in M\}) = P(\{\omega \mid X(\omega) \in (-\infty, t_0)\}) \\ &= P(X < t_0) = \lim_{x \nearrow t_0} F(x) = r. \end{aligned}$$

Let us show (71). A generator of $(F^-)^{-1}(\mathcal{B}(\mathbb{R}))$ is $\mathcal{E} := \{(F^-)^{-1}((-\infty, r]) \mid r \in \mathbb{R}\}$. By Lemma 57, $(F^-)^{-1}((-\infty, r]) = I_r$ with $I_r = [0, F(r)]$ or $I_r = (0, F(r)]$, depending on whether $0 \in F(\mathbb{R})$ or $0 \notin F(\mathbb{R})$, because $F^-(q) \leq r \Leftrightarrow q \leq F(r)$. With this we get

$$P^U((F^-)^{-1}((-\infty, r])) = P^U(I_r) \stackrel{(70)}{=} F(r) = \lambda(I_r) = \lambda((F^-)^{-1}((-\infty, r])).$$

Thus (71) is also shown. Finally, we come to the proof of (72). First of all,

$$P(\tilde{X} \leq r) = P(F^-(U) \leq r) = P(U \leq F(r)) \stackrel{(70)}{=} F(r) = P(X \leq r),$$

is valid and so is $P^X = P^{\tilde{X}}$. Moreover, $\tilde{X}(\omega) = F^-(F(X(\omega))) \leq X(\omega)$ for all ω . Thus,

$$\begin{aligned} P(\tilde{X} \neq X) &= P(\tilde{X} < X) = P(F^-(F(X)) < X) = P^X(\{t \mid F^-(F(t)) < t\}) \\ &= P^{\tilde{X}}(\{t \mid F^-(F(t)) < t\}) = P(F^-(F(\tilde{X})) < \tilde{X}) \\ &= P(F^-(F(F^-(U))) < F^-(U)) \stackrel{(*)}{=} 0. \end{aligned}$$

For (*), note that $F^-(F(F^-(q))) = F^-(q)$ holds by Lemma 57. Moreover,

$$F(\tilde{X}) = F(F^-(F(X))) = F(X) = U$$

holds because of lemma 57.7. Thus, the second equation in (72) is also shown. \square

Proof of Lemma 62 from page 33:

Let $V : (\Omega, \mathcal{A}, P) \rightarrow (\mathbb{R}, \mathcal{B}(\mathbb{R}))$ be a normally distributed random variable, $V \sim N(0, 1)$ and $\Phi(x)$ be its distribution function. If $F_n \rightsquigarrow F$ then $F_n(V) \rightarrow F(V)$ almost surely, because

$$\Omega_0 := \{\omega \mid F_n(V(\omega)) \not\rightarrow F(V(\omega))\} \subseteq \{\omega \mid V(\omega) \in \mathbb{R} \setminus c(F)\} = V^{-1}(\mathbb{R} \setminus c(F))$$

and $P(\Omega_0) \leq P(V^{-1}(\mathbb{R} \setminus c(F))) = P^V(\mathbb{R} \setminus c(F)) = 0$, since $\mathbb{R} \setminus c(F)$ is at most countable. Let F^- be continuous at u and $\Omega_1 := \{\omega \in \Omega \mid F(V(\omega)) = u\}$. Then $P(\Omega_1) = 0$, since $\Omega_1 \subseteq \{\omega \in \Omega \mid V(\omega) = F^-(u)\}$, because F^- is continuous at u . Thus

$$\begin{aligned} \Phi(F_n^-(u)) &= P(V < F_n^-(u)) = P(F_n(V) < u) \\ &= \int_{\Omega} 1_{\{\omega \in \Omega \mid F_n(V(\omega)) < u\}} dP \xrightarrow{(*)} \int_{\Omega} 1_{\{\omega \in \Omega \mid F(V(\omega)) < u\}} dP \\ &= P(F(V) < u) = P(V < F^-(u)) = \Phi(F^-(u)) \end{aligned}$$

holds. In (*) we can use the dominated convergence theorem, because

$$1_{\{\omega \in \Omega \mid F_n(V(\omega)) < u\}}(\omega) \rightarrow 1_{\{\omega \in \Omega \mid F(V(\omega)) < u\}}(\omega)$$

holds for all $\omega \in \Omega \setminus (\Omega_0 \cup \Omega_1)$ and $P(\Omega_0 \cup \Omega_1) = 0$. By continuity of Φ^{-1} , $F_n^-(u) \rightarrow F^-(u)$ follows for every such u .

For a proof of the converse let U be uniformly distributed on $(0, 1)$. Because of $F_n^- \rightsquigarrow F^-$, just like in the first case, $F_n^-(U) \rightarrow F^-(U)$ almost surely follows and therefore¹¹ $F_n^-(U) \xrightarrow{\mathcal{D}} F^-(U)$. Thus $F_n \rightsquigarrow F$ by Lemma 42 and Lemma 60.1 \square

Proof of Lemma 59 from page 32:

Let $g := h^+$. Obviously $g : [h(a), h(b)] \rightarrow [a, b]$ is non decreasing. Let $r = g(t)$ and $(r_n)_{n \in \mathbb{N}}$ be a sequence in $\{r' \in [a, b] \mid h(r') \leq t\}$ with $r_n \nearrow r$. It follows $h(r) = \lim_{n \rightarrow \infty} h(r_n) \leq t$. Thus $r \in \{r' \in [a, b] \mid h(r') \leq t\}$ and therefore $g(t) = \max\{r' \in [a, b] \mid h(r') \leq t\}$. From this we can directly deduce

$$h(r) > t \Leftrightarrow r > g(t), \quad \text{for all } r \in [a, b], t \in [h(a), h(b)]. \quad (101)$$

Next, we show that g is right continuous. Let $r = g(t)$, $t_n \searrow t \in [h(a), h(b))$ and $\varepsilon > 0$. Then $h(r + \varepsilon) > t$ follows. So there is $N \in \mathbb{N}$ with $h(r + \varepsilon) > t_n \geq t$ for all $n \geq N$. Thus $r + \varepsilon > g(t_n) \geq g(t)$ follows for all $n \geq N$, because of (101). Since $\varepsilon > 0$ is arbitrary, $g(t_n) \xrightarrow{n \rightarrow \infty} g(t)$ follows. Now we show $g^-(r) = h(r)$ for all $r \in [a, b]$. In any case

$$g(h(r)) = \sup\{r' \in [a, b] \mid h(r') \leq h(r)\} \geq r$$

¹¹If $X_n \rightarrow X$ converges almost surely, then $X_n \xrightarrow{\mathcal{D}} X$ holds. This follows directly from the definition together with the change of variable formula and the dominated convergence theorem.

and so $g^-(r) \leq h(r)$. Suppose $t := g^-(r) < h(r)$. Due to (101), $g(t) < r$ follows in contradiction to $g^-(r) = \min\{t' \in [h(a), h(b)] \mid r \leq g(t')\}$ (cf. Lemma 57.2). Thus $g^-(r) = h(r)$.

Let us now consider the case where h is right continuous. Also in this case we set to abbreviation $g := h^- : [h(a), h(b)] \rightarrow [a, b]$. Now we show $g^+ = h$. We have $g(h(r)) = \inf\{r' \in [a, b] \mid h(r') \geq h(r)\} \leq r$. Consequently $g^+(r) = \sup\{t \in [h(a), h(b)] \mid g(t) \leq r\} \geq h(r)$. Assume $t := g^+(r) > h(r)$. Completely analogous to (101), one shows that this assumption is equivalent to $r < g(t)$. This is a contradiction to $g^+(r) = \max\{t' \in [h(a), h(b)] \mid g(t') \leq r\}$. \square

Proof of Lemma 65 from page 33:

1. Trivial.
2. For the proof of “ \leq ” transform a decomposition of $[a, b]$ into decompositions of $[a, c]$ and $[c, b]$ (if necessary add c as new decomposition point) and use the triangle inequality.
For the proof of “ \geq ” let $\varepsilon > 0$ be given. Now approximate $V(f, a, c)$ by some decomposition with a precision of $\varepsilon/2$. Proceed analogously with $V(f, c, b)$. Then it follows $V(f, a, b) \leq V(f, a, c) + \varepsilon/2 + V(f, c, b) + \varepsilon/2$. Since ε is arbitrary, it follows $V(f, a, b) \leq V(f, a, c) + V(f, c, b)$.
3. Follows trivially from the triangle inequality.
4. For “ \Rightarrow ” let $g(x) := \frac{1}{2}(V(f, a, x) + f(x))$ and $h(x) := \frac{1}{2}(V(f, a, x) - f(x))$. For $x' < x$

$$\begin{aligned} g(x) - g(x') &= \frac{1}{2}(V(f, a, x) - V(f, a, x') + f(x) - f(x')) \\ &= \frac{1}{2}(V(f, x', x) - (f(x') - f(x))) \geq 0 \end{aligned}$$

follows by definition of the variation V . Analogously $h(x) - h(x') \geq 0$. Thus $f = g - h$ and g and h are non decreasing.

For “ \Rightarrow ” obviously $V(g, a, b) = g(b) - g(a)$ holds for non-decreasing g (analogously for non-increasing). The assertion then follows from 3.

5. Because of 4. it is sufficient to show the measurability of non-decreasing mappings $f : [a, b] \rightarrow [f(a), f(b)]$. This follows from $f^{-1}([c, f(b)]) = [f^-(c), b]$, cf. Lemma 57 and the fact that $\{[c, f(b)] \mid c \in [f(a), f(b)]\}$ is a generator of Borel sets in $[f(a), f(b)]$.
6. If $x \mapsto V(f, a, x)$ were not left continuous in $c \in (a, b]$, it would have a jump since it is non-decreasing. I.e.

$$\exists \varepsilon > 0 \forall x \in [a, c) : V(f, a, x) < V(f, a, c) - \varepsilon. \quad (102)$$

However, f is left continuous in c , i.e.

$$\exists \delta > 0 \forall x \in (c - \delta, c] : f(c) - f(x) < \varepsilon/2. \quad (103)$$

Let $a = x_1 < \dots < x_n = c$ with $x_{i+1} - x_i < \delta$ for all $i = 1, \dots, n-1$ and

$$V(f, a, c) - \varepsilon/2 < \sum_{i=1}^{n-1} |f(x_{i+1}) - f(x_i)| \leq V(f, a, c). \quad (104)$$

Thus

$$\sum_{i=1}^{n-2} |f(x_{i+1}) - f(x_i)| \leq V(f, a, x_{n-1}) \stackrel{(102)}{<} V(f, a, c) - \varepsilon \quad (105)$$

and therefore

$$\begin{aligned} V(f, a, c) - \varepsilon/2 &\stackrel{(104)}{<} \sum_{i=1}^{n-1} |f(x_{i+1}) - f(x_i)| \leq V(f, a, x_{n-1}) + |f(x_n) - f(x_{n-1})| \\ &\stackrel{(103)}{<} V(f, a, x_{n-1}) + \varepsilon/2 \stackrel{(105)}{<} V(f, a, c) - \varepsilon + \varepsilon/2 = V(f, a, c) - \varepsilon/2. \end{aligned}$$

Obviously, this is a contradiction. Hence $x \mapsto V(f, a, x)$ is left continuous. \square

Proof of Lemma 66 from page 34:

1. Define $F_1 : \mathbb{R} \rightarrow \mathbb{R}$ by $F_1(x) = F(x)$ for $a \leq x \leq b$, $F_1(x) = F(a)$ for $x < a$ and $F_1(x) = F(b)$ for $b < x$. Then $\lambda^{F_0^-} = \mu_{F_1}$ holds on $\mathcal{B}(\mathbb{R})$, because for $x < y$, $x \leq b$ and $a \leq y$,

$$\begin{aligned} \lambda^{F_0^-}((x, y]) &= \lambda((F_0^-)^{-1}((x, y])) \\ &= \lambda(\{q \in [F(a), F(b)] \mid x < F_0^-(q) \leq y\}) \\ &= (*) \end{aligned}$$

$$\begin{aligned} \text{if } a \leq x, y \leq b: \quad (*) &= \lambda(\{q \in [F(a), F(b)] \mid F_0(x) < q \leq F_0(y)\}) \quad \text{cf. Lemma 57.5} \\ &= F_1(y^+) - F_1(x^+) \quad \text{since } F \text{ and } F_1 \text{ are right continuous} \\ &= \mu_{F_1}((x, y]) \end{aligned}$$

$$\begin{aligned} \text{if } x < a, y \leq b: \quad (*) &= \lambda(\{q \in [F(a), F(b)] \mid a \leq F_0^-(q) \leq y\}) \\ &= \lambda(\{q \in [F(a), F(b)] \mid F_0(a) \leq q \leq F_0(y)\}) \\ &= \mu_{F_1}((x, y]) \end{aligned}$$

$$\begin{aligned} \text{if } a \leq x, b < y: \quad (*) &= \lambda(\{q \in [F(a), F(b)] \mid x < F_0^-(q) \leq b\}) \\ &= \lambda(\{q \in [F(a), F(b)] \mid F_0(x) < q \leq F_0(b)\}) \\ &= \mu_{F_1}((x, y]) \end{aligned}$$

$$\begin{aligned} \text{if } x < a, b < y: \quad (*) &= \lambda(\{q \in [F(a), F(b)] \mid a \leq F_0^-(q) \leq b\}) \\ &= \lambda(\{q \in [F(a), F(b)] \mid F_0(a) \leq q \leq F_0(b)\}) \\ &= \mu_{F_1}((x, y]) \end{aligned}$$

follows. If $y \leq x$ or $b < x$ or $y < a$, then $\lambda^{F_0^-}((x, y]) = 0 = \mu_{F_1}((x, y])$. By means of the change-of-variables formula $\int_{\Omega'} \alpha d\mu^T = \int_{\Omega} (\alpha \circ T) d\mu$ for $\alpha : \Omega' \rightarrow \mathbb{R}$ and $T : \Omega \rightarrow \Omega'$, see e. g. Stroock (1994), Lemma 5.0.1, we conclude (note $T := F_0^- : [F(a), F(b)] \rightarrow \mathbb{R}$ and $\alpha := g$)

$$\int_{[F(a), F(b)]} g \circ F_0^- d\lambda = \int_{\mathbb{R}} g d\lambda^{F_0^-} = \int_{\mathbb{R}} g d\mu_{F_1} = \int_{\mathbb{R}} 1_{(a, b]} g d\mu_{F_1} = \int_{\mathbb{R}} 1_{(a, b]} g d\mu_F.$$

2. First note $(F^-)_{|(0,1)} : (0, 1) \rightarrow \mathbb{R}$. Then one shows $\lambda^{(F^-)_{|(0,1)}} = \mu_F$ similarly as above.

$$\begin{aligned} \lambda^{(F^-)_{|(0,1)}}((x, y]) &= \lambda(\{q \in (0, 1) \mid x < F^-(q) \leq y\}) \\ &= \lambda(\{q \in (0, 1) \mid F(x) < q \leq y\}) \\ &= F(y^+) - F(x^+) = \mu_F((x, y]), \end{aligned}$$

for all $x \leq y$. Thus $\lambda^{(F^-)|_{(0,1)}} = \mu_F$ holds on $\mathcal{B}(\mathbb{R})$. Now

$$\begin{aligned} \int_{(F(a), F(b))} g \circ (F^-)|_{(0,1)} d\lambda &= \int_{(0,1)} 1_{(F(a), F(b))} \cdot (g \circ (F^-)|_{(0,1)}) d\lambda \\ &= \int_{(0,1)} (1_{(a,b]} \circ (F^-)|_{(0,1)}) \cdot (g \circ (F^-)|_{(0,1)}) d\lambda \\ &= \int_{\mathbb{R}} 1_{(a,b]} \cdot g d\lambda^{(F^-)|_{(0,1)}} = \int_{\mathbb{R}} 1_{(a,b]} \cdot g d\mu_F \end{aligned}$$

follows again with the change-of-variables formula and Lemma 57.5. \square

Proof of Lemma 67 from page 34:

1. Let $D := D' \cup ((0,1) \cap \mathbb{Q})$. Then D is countably infinite, includes all discontinuities of all \tilde{f}_i , $i \in J$, and D is also dense in $[0,1]$. Now let $D = \{q_k \mid k \in \mathbb{N}^{\geq 1}\}$ be a fixed bijective enumeration of D . Each $r \in [0,1]$ is assigned a (possibly degenerate) interval $I_r \subset [0,1]$,

$$I_r := \left[\sum_{q_k < r} 2^{-k}, \sum_{q_k \leq r} 2^{-k} \right].$$

(Where $\sum_{q_k < r}$ means that the sum is formed over all $k \in \mathbb{N}^{\geq 1}$ to which $q_k < r$ applies. Analogue with “ \leq ”.) I_r is an interval of positive length if $r \in D$. For $r \in [0,1] \setminus D$, $I_r = \{\sum_{q_k < r} 2^{-k}\}$ is degenerated to a single point, since in this case $\sum_{q_k < r} 2^{-k} = \sum_{q_k \leq r} 2^{-k}$. Furthermore,

$$r < r' \Leftrightarrow I_r^+ < I_{r'}^- \quad \text{for all } r, r' \in [0,1]. \quad (106)$$

This follows immediately from the fact that $D \cap (r, r') \neq \emptyset$ holds for all $r < r'$. Between every two $I_r, I_{r'}$ with $r < r'$, there are infinitely many $I_{r''}$ with $r'' \in D$ and infinitely many with $r'' \in [0,1] \setminus D$. Nevertheless, $Z := \{I_r \mid r \in [0,1]\}$ is a decomposition of $[0,1]$.

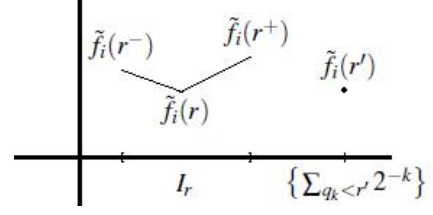
Proof: The only thing to show is that for each $z \in [0,1] \setminus \bigcup_{r \in D} I_r$ there is $x \in [0,1] \setminus D$, with $z = \sum_{q_k < x} 2^{-k}$. The uniqueness then follows from (106). So let $z \in [0,1] \setminus \bigcup_{r \in D} I_r$.

1st case $0 \in D'$, so $0 \in D$. Let $M := \{\sum_{q_k \leq r} 2^{-k} \mid r \in D \wedge \sum_{q_k \leq r} 2^{-k} < z\}$. Then $M \neq \emptyset$ because $z \notin I_0 = [0, 2^{-k}]$, where $0 = q_k$. Thus $z_0 := \sup M$ exists. Obviously $z_0 \leq z$. Since all intervals in Z are pairwise disjoint, $z_0 < z$ cannot be valid because of $\sum_{r \in D} \text{length}(I_r) = 1$. Therefore $z_0 = z$. Let $(\sum_{q_k \leq r_i} 2^{-k})_{i \in \mathbb{N}}$ be a strictly monotone increasing sequence of M with $\lim_{i \rightarrow \infty} \sum_{q_k \leq r_i} 2^{-k} = z$. Then $(r_i)_{i \in \mathbb{N}}$ is a strictly monotone increasing sequence of D . Let $x := \lim_{i \rightarrow \infty} r_i \in (0,1]$. Thus $\sum_{q_k < x} 2^{-k} = \lim_{i \rightarrow \infty} \sum_{q_k \leq r_i} 2^{-k} = z$. By choice, $x \in [0,1] \setminus D$ must hold.

2nd case $0 \notin D'$, so $0 \notin D$. If $z = 0$ holds, then $z = \sum_{q_k < x} 2^{-k}$ where $x := 0 \in [0,1] \setminus D$. If $z > 0$, we define M as above. For all $N \in \mathbb{N}^{\geq 1}$, let $x_N := \frac{1}{2} \min(q_1, \dots, q_N)$. Thus $S_N := \sum_{q_k \leq x_N} 2^{-k} \leq \sum_{k=N+1}^{\infty} 2^{-k} = \sum_{k=0}^{\infty} 2^{-N-1-k} = 2^{-N-1} \sum_{k=0}^{\infty} 2^{-k} = 2^{-N}$. Choose $N \in \mathbb{N}^{\geq 1}$ and $r \in D$ with $r < x_N$ and $S_N < z$. Then $\sum_{q_k \leq r} 2^{-k} \in M$, i.e. $M \neq \emptyset$. The rest now follows exactly as in the 1st case.

This means (illustratively) that each $r \in D$ is “blown up” to a small interval I_r and each element $r \in [0,1] \setminus D$ is left as a single element. Now we come to the definition of the f_i , $i \in J$.

On the degenerate intervals $I_{r'} = \{\sum_{q_k < r'} 2^{-k}\}$, $r' \in [0, 1] \setminus D$, f_i has the value $\tilde{f}_i(r')$. On the intervals of positive length $I_r = [\sum_{q_k < r} 2^{-k}, \sum_{q_k \leq r} 2^{-k}]$, $r \in D$, f_i has value $\tilde{f}_i(r^-)$ on I_r^- with $f_i(0) := \tilde{f}_i(0)$ if $0 = I_0^-$, $\tilde{f}_i(r^+)$ on I_r^+ with $f_i(1) := \tilde{f}_i(1)$ if $1 = I_1^+$ and $\tilde{f}_i(r)$ in the middle between them and everything linearly



connected (see sketch). We will show f_i is right continuous (left continuous is shown analogously). From the definition of f_i it follows immediately that f_i is right continuous in every $x \in I_r \setminus \{I_r^+\}$ for $r \in D$. It remains to show f_i is right continuous in the points I_r^+ with $r \in D$ and in all $x \in I_{r'} = \{\sum_{q_k < r'} 2^{-k}\}$, $r' \in [0, 1] \setminus D$. Each $x \in [0, 1]$ lies in exactly one $I_{r(x)}$. Thus

$$\min\{\tilde{f}_i(r(x)^-), \tilde{f}_i(r(x)), \tilde{f}_i(r(x)^+)\} \leq f_i(x) \leq \max\{\tilde{f}_i(r(x)^-), \tilde{f}_i(r(x)), \tilde{f}_i(r(x)^+)\} \quad (107)$$

follows, due to the construction of f_i . If $x_n \searrow x$ holds in $[0, 1]$, then $r(x_n) \searrow r(x)$ follows, since $\sum_{r(x) < q_k < r(x_n)} 2^{-k} = I_{r(x_n)}^- - I_{r(x)}^+ \rightarrow 0$. If $r(x) \in [0, 1] \setminus D$, then

$$\tilde{f}_i(r(x_n)^-) \rightarrow \tilde{f}_i(r(x)), \tilde{f}_i(r(x_n)) \rightarrow \tilde{f}_i(r(x)) \text{ and } \tilde{f}_i(r(x_n)^+) \rightarrow \tilde{f}_i(r(x))$$

follows, since \tilde{f}_i is continuous in $r(x)$. Thus $f_i(x_n) \rightarrow f_i(x)$, due to $\tilde{f}_i(r(x)) = f_i(x)$ and (107). If $r(x) \in D$, i.e. $x = I_{r(x)}^+$, then

$$\tilde{f}_i(r(x_n)^-) \rightarrow \tilde{f}_i(r(x)^+), \tilde{f}_i(r(x_n)) \rightarrow \tilde{f}_i(r(x)^+) \text{ and } \tilde{f}_i(r(x_n)^+) \rightarrow \tilde{f}_i(r(x)^+)$$

by definition of the right limit $\tilde{f}_i(r(x)^+)$. So $f_i(x_n) \rightarrow f_i(x)$, due to $\tilde{f}_i(r(x)^+) = f_i(x)$ and (107). Now h is defined by

$$h(r) := \frac{1}{2} \left(\sum_{q_k < r} 2^{-k} + \sum_{q_k \leq r} 2^{-k} \right).$$

Then h is strictly monotonically increasing (obviously) with $\tilde{f}_i = f_i \circ h$ for all $i \in J$ (this is easy to prove; it is best to systematically go through all possible cases).

2. We only consider the case where all \tilde{f}_i are left continuous. Note that this implies $1 \notin D$. In this case we define h by¹²

$$h(r) := \sum_{q_k < r} 2^{-k}.$$

Then $h(r_1) < h(r_2)$ for all $r_1 < r_2$, $h(0) = 0$, $h(1) = 1$, h is left continuous and

$$I_r = [h(r^-), h(r^+)] = [h(r), h(r^+)].$$

for all $r \in (0, 1]$. Since every \tilde{f}_i , $i \in J$, is left continuous, $f_i(h(r)) = f_i(h(r^-)) = \tilde{f}_i(r^-) = \tilde{f}_i(r)$ for all $r \in (0, 1]$ and $f_i(h(0)) = f_i(0) = \tilde{f}_i(0)$ holds, i.e. $f_i \circ h = \tilde{f}_i$. It remains to be shown that h is the quantile function of an corresponding distribution function $F : [0, 1] \rightarrow [0, 1]$. We define $F(t) := h^+(t) = \sup\{r \in [0, 1] \mid h(r) \leq t\}$, cf. Remark 58. Then $F(0) = 0$, since $h(0) = 0$ and $h(r) > 0$ for all $r > 0$. Moreover $F(1) = 1$, F is non-decreasing, right continuous and satisfies $F^- = h$, by Lemma 59. If F were not continuous, F would have jumps and $F^- = h$ would be

¹²If all \tilde{f} are right continuous, h could instead be defined by $h(r) := \sum_{q_k \leq r} 2^{-k}$. Then h would be right continuous and would satisfy $\tilde{f}_i = f_i \circ h$ for all $i \in J$ and all other corresponding properties.

constant on some intervals of positive length, in contradiction to h being strictly monotonically increasing.

3. We assume that all \tilde{f}_i are continuous in 0. Then (75) holds. For if $F(t) = \sup\{r' \in [0, 1] \mid h(r') \leq t\} = 0$ holds for $t > 0$, then

$$h(x_N) = \sum_{q_k < x_N} 2^{-k} \leq \sum_{k=N+1}^{\infty} 2^{-k} = \sum_{k=0}^{\infty} 2^{-N-1-k} = 2^{-N-1} \sum_{k=0}^{\infty} 2^{-k} = 2^{-N},$$

with $x_N := \min(q_1, \dots, q_N) > 0$. For this, note that $0 \notin D$ holds, since all \tilde{f}_i are continuous in 0. For sufficiently large N with $2^{-N} < t$, a contradiction would result. The assumption $F(t) = \sup\{r' \in [0, 1] \mid h(r') \leq t\} = 1$ for some $t < 1$ leads directly to a contradiction.

4. If all \tilde{f}_i are of bounded variation, there are non-decreasing functions \tilde{g}_{i1} and \tilde{g}_{i2} with $\tilde{f}_i = \tilde{g}_{i1} - \tilde{g}_{i2}$. The same construction as above with \tilde{f}_i , but now carried out with \tilde{g}_{i1} and \tilde{g}_{i2} , leads because of $\tilde{f}_i(r^-) = \tilde{g}_{i1}(r^-) - \tilde{g}_{i2}(r^-)$, $\tilde{f}_i(r^+) = \tilde{g}_{i1}(r^+) - \tilde{g}_{i2}(r^+)$ and the linear construction on I_r to non-decreasing functions g_{i1} and g_{i2} , with $f_i = g_{i1} - g_{i2}$. Consequently, all f_i are also of bounded variation. \square

Proof of Remark 68 from page 35:

We assume f has a left limit in each $x \in [0, 1]$. Let $D' := \{x \in [0, 1] \mid f \text{ is discontinuous in } x\}$. If $x \in D'$, then

$$\exists n \in \mathbb{N}^{\geq 1} \forall \delta > 0 \exists x' \in [0, 1] \text{ such that } |x - x'| < \delta \wedge |f(x) - f(x')| \geq \frac{1}{n}.$$

Thus $D' = \bigcup_{n \in \mathbb{N}^{\geq 1}} D'_n$, where,

$$D'_n := \{x \in [0, 1] \mid \forall \delta > 0 \exists x' \in [0, 1] \text{ with } |x - x'| < \delta \wedge |f(x) - f(x')| \geq 1/n\}.$$

Suppose D' is uncountable. Then there exists $n \in \mathbb{N}^{\geq 1}$ such that D'_n is uncountable too. We will show

$$\exists x_0 \in D'_n \forall \varepsilon > 0 : (x_0 - \varepsilon, x_0) \cap D'_n \neq \emptyset. \quad (108)$$

If (108) were false, then for every $x \in D'_n$ there would be an $\varepsilon_x > 0$ with $(x - \varepsilon_x, x) \cap D'_n = \emptyset$. Then $(x - \varepsilon_x, x)$, $x \in D'_n$ would be a family of pairwise disjoint nonempty open intervals and in particular D'_n would thus be countable - a contradiction! Consequently, (108) is true. By definition of D'_n and because of (108), for every $k \in \mathbb{N}^{\geq 1}$ exists $x_k, x'_k \in (x_0 - \frac{1}{k}, x_0) \cap D'_n$ with $|f(x_k) - f(x'_k)| \geq \frac{1}{n}$. The sequence $(x_1, x'_1, x_2, x'_2, \dots)$ converges from the left to x_0 in contradiction to the fact that $(f(x_1), f(x'_1), f(x_2), f(x'_2), \dots)$ does not converge at all. \square

Proof of Lemma 69 from page 35:

For each $i \in J$ let $\tilde{f}_{i,1}$ and $\tilde{f}_{i,2}$ non-decreasing with $\tilde{f}_i = \tilde{f}_{i,1} - \tilde{f}_{i,2}$. We first consider the case where $\tilde{f}_i(0) = 0$ holds for all $i \in J$. Thus $\tilde{f}_{i,1}$ and $\tilde{f}_{i,2}$ can be chosen a priori such that $\tilde{f}_{i,1}(0) = 0$ and $\tilde{f}_{i,2}(0) = 0$ for all $i \in J$ too. Define $h : [0, 1] \rightarrow [0, 1]$ by

$$h(r) := \frac{r}{L} + \frac{1}{L} \cdot \sum_{i=1}^d (\tilde{f}_{i,1}(r) + \tilde{f}_{i,2}(r)) \quad \text{with } L := 1 + \sum_{i=1}^d (\tilde{f}_{i,1}(1) + \tilde{f}_{i,2}(1)). \quad (109)$$

Obviously h is strictly monotonically increasing with $h(0) = 0$ and $h(1) = 1$. We define mappings $g_i : \text{Image}(h) \rightarrow \mathbb{R}$ by $g_i(h(x)) := \tilde{f}_i(x)$ for every $i \in J$. Since h is strictly monotonically increasing, all g_i are well-defined with $\tilde{f}_i = g_i \circ h$. All g_i are Lipschitz continuous. To show this, we assume $x < y$ for $x, y \in \text{Image}(h)$. Then there are $s, t \in [0, 1]$ with $x = h(s)$ and $y = h(t)$. Therefore $s < t$ and

$$\begin{aligned} |g_i(y) - g_i(x)| &= |g_i(h(t)) - g_i(h(s))| = |\tilde{f}_i(t) - \tilde{f}_i(s)| \\ &\leq t - s + \sum_{i=1}^d |\tilde{f}_{i,1}(t) - \tilde{f}_{i,2}(t) - \tilde{f}_{i,1}(s) + \tilde{f}_{i,2}(s)| \\ &\leq t - s + \sum_{i=1}^d |\tilde{f}_{i,1}(t) - \tilde{f}_{i,1}(s)| + \sum_{i=1}^d |\tilde{f}_{i,2}(t) - \tilde{f}_{i,2}(s)| \\ &= t + \sum_{i=1}^d (\tilde{f}_{i,1}(t) + \tilde{f}_{i,2}(t)) - \left(s + \sum_{i=1}^d (\tilde{f}_{i,1}(s) + \tilde{f}_{i,2}(s)) \right) \\ &= L(h(t) - h(s)) = L|y - x| \end{aligned}$$

Using McShane's Lemma¹³ or the Kirszbraun Theorem, we extend each g_i , $i \in J$ to a Lipschitz continuous mapping $f_i : [0, 1] \rightarrow \mathbb{R}$. Thus $\tilde{f}_i = f_i \circ h$, for all $i \in J$. If all \tilde{f}_i are left continuous (or right continuous), then h is also left continuous (or right continuous). We now conclude, as in the proof of Lemma 67, that $F : [0, 1] \rightarrow [0, 1]$ defined by $F(t) := \sup\{r \in [0, 1] \mid h(r) \leq t\}$ is a non-decreasing continuous function (in case that all \tilde{f}_i are left continuous) with $F(0) = 0$, $F(1) = 1$ and $h = F^-$. Finally, if all \tilde{f}_i are continuous in 0, then (75) follows, as in the proof of Lemma 67, from $h(0) = 0$ and the fact that h is continuous in 0.

We still have to prove the general case with $\tilde{f}_i(0) \in \mathbb{R}$ arbitrarily. For this we define $\hat{f}_i(t) := \tilde{f}_i(t) - \tilde{f}_i(0)$, $i \in J$. Then $\hat{f}_i(0) = 0$ holds and we get $\hat{f}_i = f_i \circ h$ according to what we proved before. Then we set $h_i(t) := f_i(t) + \tilde{f}_i(0)$. And it follows $\tilde{f}_i = \hat{f}_i + \tilde{f}_i(0) = h_i \circ h$ for all $i \in J$. Of course, all h_i , $i \in J$ are also continuous. Thus the lemma is completely proved. \square

Remark 81. h from (109) in the proof of Lemma 69 could also be defined by

$$h(x) := \frac{x}{L} + \frac{1}{L} \cdot \sum_{i=1}^d V(\tilde{f}_i, 0, x) \quad \text{with} \quad L := 1 + \sum_{i=1}^d V(\tilde{f}_i, 0, 1) > 0. \quad (110)$$

¹³This is a very useful theorem. See e.g. Heinonen (2001), Theorem 6.2. Since the proof is short and relatively simple, we give it here. We extend each g_i to a Lipschitz continuous mapping $f_i : [0, 1] \rightarrow \mathbb{R}$ by $f_i(x) := \inf_{y \in A} (g_i(y) + L \cdot |x - y|)$, where $A := \text{Image}(h)$. Since

$$\begin{aligned} g(y) + L \cdot |a - y| - g(a) &= g(y) - g(a) + L \cdot |a - y| \leq 2L \cdot |a - y| \quad \text{and} \\ g(a) - (g(y) + L \cdot |a - y|) &= g(a) - g(y) - L \cdot |a - y| \leq 0 \end{aligned}$$

$f_i(a) = g_i(a)$ for all $a \in A$ follows. Let us now show that all f_i , $i \in J$ are Lipschitz continuous. Let $x, x' \in [0, 1]$ and let $\varepsilon > 0$. Choose $y_0 \in A$ with $g(y_0) + L \cdot |x' - y_0| - \varepsilon < f(x')$. Because of

$$\begin{aligned} f(x) - f(x') &= \inf_{y \in A} (g_i(y) + L \cdot |x - y|) - \inf_{y' \in A} (g_i(y') + L \cdot |x' - y'|) \\ &< \inf_{y \in A} (g_i(y) + L \cdot |x - y|) - (g_i(y_0) + L \cdot |x' - y_0| - \varepsilon) \\ &\leq (g_i(y_0) + L \cdot |x - y_0|) - (g_i(y_0) + L \cdot |x' - y_0|) + \varepsilon \\ &= L(|x - y_0| + |x' - y_0|) + \varepsilon \leq L \cdot |x - x'| + \varepsilon \end{aligned}$$

and (analogously) $f(x') - f(x) < L \cdot |x' - x| + \varepsilon$, $|f(x) - f(x')| \leq L \cdot |x - x'|$ follows, since $\varepsilon > 0$ was arbitrary.

Then, h is left continuous by Lemma 65.6 and strictly monotonically increasing with $h(0) = 0$ and $h(1) = 1$. We define $g_i : \text{Image}(h) \rightarrow \mathbb{R}$ by $g_i(h(x)) := \tilde{f}_i(x)$ for every $i \in J$. Since h is strictly monotonically increasing, all g_i are well-defined with $\tilde{f}_i = g_i \circ h$. The Lipschitz continuity of g_i then follows from

$$\begin{aligned} |g_i(y) - g_i(x)| &= |g_i(h(t)) - g_i(h(s))| = |\tilde{f}_i(t) - \tilde{f}_i(s)| \\ &\leq t - s + V(\tilde{f}_i, s, t) \leq t - s + \sum_{i=1}^d V(\tilde{f}_i, s, t) \\ &= t - s + \sum_{i=1}^d (V(\tilde{f}_i, 0, t) - V(\tilde{f}_i, 0, s)) = L(h(t) - h(s)) = L|y - x| \end{aligned}$$

for $x < y$ with $x = h(s)$, $y = h(t)$ and $s < t$. The rest follows as in the proof of Lemma 69.

Proof of Lemma 70 from page 35:

Since $(t_{n, \lfloor sn \rfloor})_{n \in \mathbb{N}}$ is bounded, it has cluster points. Assume there is a cluster point x' with $x' < x := F^-(s)$. We choose a subsequence $(t_{n_k, \lfloor sn_k \rfloor})_{k \in \mathbb{N}}$ of $(t_{n, \lfloor sn \rfloor})_{n \in \mathbb{N}}$ with $t_{n_k, \lfloor sn_k \rfloor} \xrightarrow{k \rightarrow \infty} x'$. Since every real sequence contains a monotone subsequence, we can assume without loss of generality that the subsequence is already monotone. Then $F(x') < s \leq F(x)$ holds by Lemma 57. Let $\varepsilon := s - F(x') > 0$.

1st case $(t_{n_k, \lfloor sn_k \rfloor})_{k \in \mathbb{N}}$ is non decreasing. Because of $F_{n_k}(t_{n_k, \lfloor sn_k \rfloor}) \geq \frac{\lfloor sn_k \rfloor}{n_k} \xrightarrow{k \rightarrow \infty} s$ there is a k_0 with $F_{n_k}(t_{n_k, \lfloor sn_k \rfloor}) \geq s - \varepsilon/3$ for all $k \geq k_0$. Due to $\sup_{x \in [a, b]} |F_n(x) - F(x)| \xrightarrow{n \rightarrow \infty} 0$ there is a $k_1 \geq k_0$ with $|F_{n_k}(x) - F(x)| < \varepsilon/3$ for all $k \geq k_1$ and all $x \in [a, b]$. Thus $F(t_{n_k, \lfloor sn_k \rfloor}) \geq s - 2\varepsilon/3$ for all $k \geq k_1$. This is a contradiction to $t_{n_k, \lfloor sn_k \rfloor} \leq x'$ and $F(t_{n_k, \lfloor sn_k \rfloor}) \leq F(x') = s - \varepsilon$.

2nd case $(t_{n_k, \lfloor sn_k \rfloor})_{k \in \mathbb{N}}$ is non increasing. As just shown $F(t_{n_k, \lfloor sn_k \rfloor}) \geq \frac{\lfloor sn_k \rfloor}{n_k} \geq s - 2\varepsilon/3$ for all sufficiently large k , contradicting $F(t_{n_k, \lfloor sn_k \rfloor}) \xrightarrow{k \rightarrow \infty} F(x') = s - \varepsilon$, since F is right continuous.

To show the second assertion, we assume that F^- is continuous in s . Let x_0 be the smallest cluster point of $(t_{n, \lfloor sn \rfloor})_{n \in \mathbb{N}}$. Let $\varepsilon > 0$ be arbitrary. Thus there exists $n_0 \in \mathbb{N}$ with $x_0 - \varepsilon < t_{n, \lfloor sn \rfloor}$ for all $n \geq n_0$. Consequently

$$\forall n \geq n_0 : F^-(s) - \varepsilon \leq x_0 - \varepsilon < t_{n, \lfloor sn \rfloor}. \quad (111)$$

Because of $\sup_{x \in [a, b]} |F_n(x) - F(x)| \xrightarrow{n \rightarrow \infty} 0$ it follows $F_n^-(s) \rightarrow F^-(s)$ by Lemma 62, i.e.

$$\exists n_1 \geq n_0 \forall n \geq n_1 : F^-(s) \geq F_n^-(s) - \varepsilon \geq F_n^-(\lfloor sn \rfloor / n) - \varepsilon = t_{n, \lfloor sn \rfloor} - \varepsilon. \quad (112)$$

Thus $t_{n, \lfloor sn \rfloor} \xrightarrow{n \rightarrow \infty} F^-(s)$ by (111) and (112). \square

Proof of Lemma 72 from page 36

Our proof based on the proof of Lemma A.1 in Bischoff (1998). Let $D := \text{support}(F) \subseteq [a, b]$. First we show

$$\int_{[a, F^-(s)]}^{(R)} (u \circ F) \, df \quad \text{exists.} \quad (113)$$

Proof of (113): First, note that the integral $\int_{[a, F^-(s)]}^{(R)} (u \circ F) \, df$ exists if and only if both integrals $\int_{[a, F^-(s)]}^{(R)} (u \circ F) \, dg$ as well as $\int_{[a, F^-(s)]}^{(R)} (u \circ F) \, dh$ exist, where $f = g - h$ is the canonical Jordan decomposition (cf. Lojasiewicz (1988), Theorem 1.5.3 and Lemma 65 in this thesis). The rest follows with Theorem 5.1.2 from Stroock (1994) because $u \circ F$ has at most countably many discontinuity points and consequently the set of discontinuity points is a null set with respect to the measures μ_g and μ_h generated by f , resp. g and h (because f is continuous by premise, so g and h are continuous too, see Lemma 65).

We will now show that the difference $\left| \sum_{i=2}^{\lfloor sn \rfloor} (f(t_{n,i}) - f(t_{n,i-1})) u\left(\frac{i-1}{n}\right) - \int_{[a, F^-(s)]}^{(R)} (u \circ F) \, df \right|$ becomes arbitrarily small for sufficiently large n . To this end, let $\varepsilon > 0$. Since $f \in C[a, b]$, f is uniformly continuous on $[a, b]$, i.e. there exists δ_0 with $0 < \delta_0 \leq \varepsilon$ and

$$|x - x'| < \delta_0 \quad \Rightarrow \quad |f(x) - f(x')| < \varepsilon \quad (114)$$

for all $x, x' \in [a, b]$. Since (113), there exists δ_1 with $0 < \delta_1 \leq \delta_0$ and

$$\left| \sum_{i=1}^k (f(v_i) - f(v_{i-1})) u(F(v_{i-1})) - \int_{[a, F^-(s)]}^{(R)} (u \circ F) \, df \right| < \varepsilon \quad (115)$$

for all $k \in \mathbb{N}^{\geq 1}$ and $a = v_0 \leq v_1 \leq \dots \leq v_{k-1} \leq v_k = F^-(s)$ with $v_i - v_{i-1} < \delta_1$ for all $i \in \{1, \dots, k\}$. Since the variation (cf. Definition 64) $x \mapsto V(f, a, x)$ of f is uniformly continuous on $[a, b]$ (because f is continuous, cf. Lemma 65.6) there exists δ with $0 < \delta \leq \delta_0$ and

$$x' - \delta < x \leq x' \quad \Rightarrow \quad V(f, x, x') < \varepsilon \quad (116)$$

for all $x, x' \in [a, b]$. Because of Lemma 70, there is $n_0 \in \mathbb{N}$ with

$$|t_{n, \lfloor sn \rfloor} - F^-(s)| < \delta/2 \quad \text{for all } n \geq n_0. \quad (117)$$

Since u is also uniformly continuous on $[0, 1]$ and because of $\sup_{x \in [a, b]} |F_n(x) - F(x)| \xrightarrow{n \rightarrow \infty} 0$, there exists an $n_1 \geq n_0$ with

$$|u(F_n(x)) - u(F(x))| \leq \varepsilon \quad \text{for all } n \geq n_1 \text{ and all } x \in [a, b]. \quad (118)$$

Because of $\sup_{x \in [a, b]} |F_n(x) - F(x)| \rightarrow 0$ and (117) there exists an $n_2 \geq n_1$ such that

$$\forall n \geq n_2 \forall d \in D \cap [a, F^-(s)] \exists j \in \{1, \dots, \lfloor sn \rfloor\} : |d - t_{n,j}| < \delta/2.$$

Thus

$$\begin{aligned} t_{n,1} - a \geq \delta &\quad \Rightarrow \quad \left(a + \frac{\delta}{2}, t_{n,1} - \frac{\delta}{2} \right) \cap D = \emptyset, \\ t_{n,i} - t_{n,i-1} \geq \delta &\quad \Rightarrow \quad \left(t_{n,i-1} + \frac{\delta}{2}, t_{n,i} - \frac{\delta}{2} \right) \cap D = \emptyset, \\ F^-(s) - t_{n,i} \geq \delta &\quad \Rightarrow \quad \left(t_{n,i} + \frac{\delta}{2}, F^-(s) - \frac{\delta}{2} \right) \cap D = \emptyset, \\ t_{n,i} - F^-(s) \geq \delta &\quad \Rightarrow \quad \left(F^-(s) + \frac{\delta}{2}, t_{n,i} - \frac{\delta}{2} \right) \cap D = \emptyset \end{aligned}$$

holds for all $n \geq n_2$ and $i \leq \lfloor sn \rfloor$. Consequently $s_1, \dots, s_r \in [a, F^-(s)] \setminus D$ can be found with $s_1 < \dots < s_r$ such that

$$v_{n,1} - a < \delta \quad \text{and} \quad v_{n,i} - v_{n,i-1} < \delta \quad \text{for all } i \in \{1, \dots, \ell(n, s)\} \quad (119)$$

where $v_{n,1} < \dots < v_{n,\ell(n,s)}$ with $\{v_{n,1}, \dots, v_{n,\ell(n,s)}\} = \{t_{n,1}, \dots, t_{n,\lfloor sn \rfloor}\} \cup \{s_1, \dots, s_r\}$. Thus

$$A(n, s) := \sum_{i=2}^{\lfloor sn \rfloor} (f(t_{n,i}) - f(t_{n,i-1})) u \left(\frac{i-1}{n} \right) = \sum_{i=2}^{\ell(n,s)} (f(v_{n,i}) - f(v_{n,i-1})) u(F(v_{n,i-1})).$$

Note for this that for $t_{n,i} = \dots = t_{n,j}$ the corresponding summands on the left-hand side are $= 0$ and for $t_{n,i} < s_{n,p} < \dots < s_{n,q} < t_{n,j}$ a finite telescoping series results on the right-hand side. Let $\ell'(n, s) := \max\{i \in \{1, \dots, \ell(n, s)\} \mid v_{n,i} \leq F^-(s)\}$ and

$$I(n, s) := \sum_{i=2}^{\ell'(n,s)} (f(v_{n,i}) - f(v_{n,i-1})) u(F(v_{n,i-1})).$$

Let $u_0 := \sup_{r \in [0,1]} |u(r)|$. Thus

$$\begin{aligned} \left| I(n, s) - \int_{[a, F^-(s)]}^{(R)} (u \circ F) \, df \right| &< \varepsilon + |f(v_{n,1}) - f(a)| \cdot u_0 + |f(F^-(s)) - f(v_{n,\ell'(n,s)})| \cdot u_0 \\ &< \varepsilon(1 + 2u_0) \end{aligned} \quad (120)$$

for all $n \geq n_2$ by (114), (115) and (119) and

$$\begin{aligned} |A(n, s) - I(n, s)| &\leq \sum_{i=2}^{\ell'(n,s)} |f(v_{n,i}) - f(v_{n,i-1})| \cdot |u(F(v_{n,i-1})) - u(F(v_{n,i-1}))| \\ &\quad + \sum_{i \in J} |f(v_{n,i}) - f(v_{n,i-1})| \cdot 2 \cdot u_0 \\ &\leq V(f, a, b) \cdot \varepsilon + V(f, F^-(s) - \delta/2, F^-(s) + \delta/2) \cdot 2u_0 \\ &< \varepsilon(V(f, a, b) + 2u_0) \end{aligned} \quad (121)$$

for all $n \geq n_2$ by (116), (118) and (119), where $J := \{i \in \mathbb{N} \mid \ell'(n, s) < i \leq \ell(n, s)\}$. Finally

$$\left| A(n, s) - \int_{[a, F^-(s)]}^{(R)} (u \circ F) \, df \right| < \varepsilon(1 + V(f, a, b) + 4u_0)$$

for all $n \geq n_2$. Thus (76). □

Proof of Lemma 73 from page 36:

As $h_{i,j}(s) := \int_{(0,s)} (f_i f_j) \circ F_0^- \, d\lambda$ is continuous with respect to $s \in [F(c), F(b)]$, the minimal eigenvalue $\chi_{\min}(H(s))$ and the maximal eigenvalue $\chi_{\max}(H(s))$ of $H(s)$ are continuous as well¹⁴. Since $[F(c), F(b)]$ is compact $s_{\min}, s_{\max} \in [F(c), F(b)]$ exist with

$$\chi_{\min} := \chi_{\min}(H(s_{\min})) \leq \chi_{\min}(H(s)) \leq \chi_{\max}(H(s)) \leq \chi_{\max}(H(s_{\max})) =: \chi_{\max}$$

¹⁴All eigenvalues $\lambda_1(s)$ to $\lambda_d(s)$ of the matrix $H(s)$ are continuous functions $\lambda_i: [F(c), F(b)] \rightarrow \mathbb{R}$, if $s \mapsto H(s)$ is continuous (and that is the case here). This is proven, for example, in Bhatia (1997), Corollary VI.1.6. Thus $\chi_{\min}(H(s)) = \min(\lambda_1(s), \dots, \lambda_d(s))$ and $\chi_{\max}(H(s)) = \max(\lambda_1(s), \dots, \lambda_d(s))$ are also continuous functions.

for all $s \in [F(c), F(b)]$. Moreover, if (A2) holds,

$$\begin{aligned}
& f_1 \cdot 1_{[a,c]}, \dots, f_d \cdot 1_{[a,c]} \text{ are l.i. in } L_2(F) \\
\Leftrightarrow & \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i f_i \cdot 1_{[a,c]} \right)^2 d\mu_F > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
\Leftrightarrow & \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i f_i \cdot 1_{(a,c]} \right)^2 d\mu_F + \left(\sum_{i=1}^d \alpha_i f_i(a) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
\stackrel{(74)}{\Leftrightarrow} & \int_{(F(a), F(c))} \left(\sum_{i=1}^d \alpha_i f_i \circ (F^-)|_{(0,1)} \right)^2 d\lambda + \left(\sum_{i=1}^d \alpha_i f_i(a) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
\Leftrightarrow & \int_{(F(a), F(c))} \left(\sum_{i=1}^d \alpha_i f_i \circ (F^-)|_{(0,1)} \right)^2 d\lambda + \int_{(0, F(a))} \left(\sum_{i=1}^d \alpha_i f_i \circ (F^-)|_{(0,1)} \right)^2 d\lambda > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
\Leftrightarrow & \int_{(0, F(c))} \left(\sum_{i=1}^d \alpha_i f_i \circ (F^-)|_{(0,1)} \right)^2 d\lambda > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\}
\end{aligned}$$

If (A1) holds, one can omit the additional summand $(\sum_{i=1}^d \alpha_i f_i(a))^2 \cdot F(a)$ in the above equivalence sequence and the chain shortens accordingly, but one comes to the same conclusion

$$\int_{(0,s)} \left(\sum_{i=1}^d \alpha_i f_i \circ (F^-)|_{(0,1)} \right)^2 d\lambda > 0 \text{ for all } \alpha \in \mathbb{R}^d \setminus \{0\} \text{ and all } s \in [F(c), F(b)].$$

$$\text{Since } \alpha^\top H(s) \alpha = \sum_{i,j=1}^d \int_{(0,s)} (\alpha_i \alpha_j f_i f_j) \circ (F^-)|_{(0,1)} d\lambda = \int_{(0,s)} \left(\sum_{i=1}^d \alpha_i f_i \circ (F^-)|_{(0,1)} \right)^2 d\lambda,$$

$H(s)$ is positive definite for each $s \in [F(c), F(b)]$ and $\chi_{\min}(H(s_{\min})) > 0$. Thus (77). \square

Proof of Lemma 74 from page 37:

$g(F_n^-(s)) \xrightarrow{n \rightarrow \infty} g(F^-(s))$, for λ -a.a. $s \in [0, 1]$, since Lemma 62 and g is continuous λ -a.e. F_n^- , F^- and g are λ -measurable, since they are λ -a.e. continuous. Thus, it follows by Egorov's theorem, cf. Kallenberg (2002), Lemma 1.36 or Bauer (2001), §20, exercise 7. \square

Proof of Lemma 75 from page 37:

Let $1 \leq i, j \leq d$, and $\varepsilon > 0$. Note that $f_i f_j$ is of bounded variation, since both f_i and f_j are (this follows from Lemma 65.4). By Lemma 74 it holds true

$$\begin{aligned}
\exists K_\varepsilon = K_\varepsilon^{(i,j)} \subseteq (0, 1) \exists n_0 \in \mathbb{N} \forall n \geq n_0 : \\
\lambda(K_\varepsilon) > 1 - \varepsilon \quad \text{and} \quad \sup_{s \in K_\varepsilon} |f_i(F_n^-(s)) f_j(F_n^-(s)) - f_i(F^-(s)) f_j(F^-(s))| \leq \varepsilon.
\end{aligned}$$

Moreover, let $m_{i,j} := \sup_{t \in [a,b]} |f_i(t)f_j(t)|$ and $n_1 > \max(n_0, \frac{K+1}{F(c)}, \frac{K+1}{\varepsilon})$. Thus $\frac{K}{n} \leq t - \frac{\lfloor tn \rfloor - K}{n} \leq \frac{K+1}{n}$ for all $n \geq n_1$ and for all $n \geq n_1$

$$\begin{aligned}
&= \left| \int_{(0,t)} (f_i f_j) \circ (F^-)_{|(0,1)} d\lambda - \int_{(0, \frac{\lfloor tn \rfloor - K}{n})} (f_i f_j) \circ F_n^- d\lambda \right| \\
&\leq \left| \int_{(0, \frac{\lfloor tn \rfloor - K}{n})} (f_i f_j) \circ (F^-)_{|(0,1)} d\lambda - \int_{(0, \frac{\lfloor tn \rfloor - K}{n})} (f_i f_j) \circ F_n^- d\lambda \right| + \left(t - \frac{\lfloor tn \rfloor - K}{n} \right) \cdot m_{i,j} \\
&\leq \left| \int_{(0, \frac{\lfloor tn \rfloor - K}{n})} |(f_i f_j) \circ (F^-)_{|(0,1)} - (f_i f_j) \circ F_n^-| d\lambda \right| + \varepsilon \cdot m_{i,j} \\
&\leq \int_{(0,1)} |(f_i f_j) \circ (F^-)_{|(0,1)} - (f_i f_j) \circ F_n^-| d\lambda + \varepsilon \cdot m_{i,j} \\
&\leq \lambda(K_\varepsilon) \cdot \varepsilon + 2\lambda(K_\varepsilon^C) \cdot m_{i,j} + \varepsilon \cdot m_{i,j} \leq \varepsilon(1 + 3m_{i,j}) \tag{122}
\end{aligned}$$

follows (note that all integrals are well defined). So, we have

$$\sup_{t \in (0,1)} \left| \frac{1}{n} \sum_{k=1}^{\lfloor tn \rfloor - K} f_i(t_{n,k}) f_j(t_{n,k}) - H_{i,j}(t) \right| \rightarrow 0, \quad H_{i,j}(t) := \int_{(0,t)} (f_i f_j) \circ (F^-)_{|(0,1)} d\lambda \tag{123}$$

for all $1 \leq i, j \leq d$, since $\frac{1}{n} \sum_{k=1}^{\lfloor tn \rfloor - K} f_i(t_{n,k}) f_j(t_{n,k}) = \int_{[0, \frac{\lfloor tn \rfloor - K}{n}]} (f_i f_j) \circ F_n^- d\lambda$. Thus $\forall \varepsilon > 0 \exists n_1 \in \mathbb{N} \forall n \geq n_1 \forall t \in (0, 1)$: $-\varepsilon \mathbf{1}_d \mathbf{1}_d^\top \leq \frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K} - H(t) \leq \varepsilon \mathbf{1}_d \mathbf{1}_d^\top$, where $\mathbf{1}_d = (1, \dots, 1)^\top$, $H(t) = (H_{i,j}(t))_{i,j=1}^d$ and “ \leq ” holds componentwise. Let $\zeta_n(t)$ be a normalized eigenvector belonging to the smallest eigenvalue of $\frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K}$. Then $\forall \varepsilon > 0 \exists n_1 \in \mathbb{N} \forall n \geq n_1 \forall t \in [F(c), F(b)]$: $(A \leq B \text{ componentwise, implies } x^\top A x \leq x^\top B x)$

$$\begin{aligned}
&-\varepsilon (\mathbf{1}_d^\top \zeta_n(t))^\top (\mathbf{1}_d^\top \zeta_n(t)) \leq \chi_{\min} \left(\frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K} \right) - \zeta_n(t)^\top H(t) \zeta_n(t) \\
\Rightarrow &-\varepsilon d + \zeta_n(t)^\top H(t) \zeta_n(t) \stackrel{(1)}{\leq} -\varepsilon \left(\sum_i (\zeta_n(t))_i \right)^2 + \zeta_n(t)^\top H(t) \zeta_n(t) \leq \chi_{\min} \left(\frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K} \right) \\
\Rightarrow &-\varepsilon d + \chi_{\min} \stackrel{(2)}{\leq} -\varepsilon d + \zeta_n(t)^\top H(t) \zeta_n(t) \leq \chi_{\min} \left(\frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K} \right), \tag{124}
\end{aligned}$$

where χ_{\min} is defined in Lemma 73 and in (1) we used $|\mathbf{1}_d^\top \cdot x|^2 \leq \|\mathbf{1}_d\|^2 \cdot \|x\|^2$ and in (2) we used the following argumentation: Let $\lambda_1 \leq \dots \leq \lambda_n$ be the eigenvalues of $H(t)$ and $\{x_1, \dots, x_d\}$ an orthonormal basis of corresponding eigenvectors. Let $x = \sum \alpha_i x_i$ with $|x| = 1$. Then

$$x^\top H(t) x = \left(\sum \alpha_i x_i \right)^\top H(t) \left(\sum \alpha_i x_i \right) = \left(\sum \alpha_i x_i \right)^\top \left(\sum \alpha_i \lambda_i x_i \right) = \sum \alpha_i^2 \lambda_i \geq \lambda_1 \sum \alpha_i^2 = \lambda_1 x^\top x = \lambda_1.$$

The assertion for $\chi_{\min} \left(\frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K} \right)$ thus follows from (124), since $\varepsilon > 0$ is arbitrarily small. The result for $\chi_{\max} \left(\frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K} \right)$ can be shown analogously. \square

Proof of Lemma 77 from page 37:

(78) follows from (123). For the proof of (79) let $V := \text{GL}(d, \mathbb{R}) := \{M \in \mathbb{R}^{d^2} \mid \det(M) \neq 0\}$. V is open in \mathbb{R}^{d^2} . Let $i: V \rightarrow V$, $i(M) := M^{-1}$. Since $\det: \mathbb{R}^{d^2} \rightarrow \mathbb{R}$ is continuous and $M^{-1} =$

$\frac{1}{\det(M)} \text{Adj}(M)$ holds, i is continuous. Let A_n, B be defined by $A_n(t) := \frac{1}{n} X_{n, \lfloor tn \rfloor - K}^\top X_{n, \lfloor tn \rfloor - K}$ and $B(t) := \int_{[0, t]} (ff^\top) \circ F^- d\lambda$ for $t \in [F(c), 1]$. Obviously B is continuous, $W := B([F(c), 1]) \subseteq V$, since $\alpha^\top \left(\int_0^t (ff^\top) \circ F^- d\lambda \right) \alpha = \int_0^t \left(\sum_{i,j=1}^d \alpha_i \alpha_j (f_i f_j) \circ F^- \right) d\lambda = \int_0^t \left(\sum_{i=1}^d \alpha_i f_i \circ F^- \right)^2 d\lambda \geq \int_{F(a)}^{F(c)} \left(\sum_{i=1}^d \alpha_i f_i \circ F^- \right)^2 d\lambda = \int_a^c \left(\sum_{i=1}^d \alpha_i f_i \right)^2 dF > 0$ for all $t \geq F(c)$, $\alpha \in \mathbb{R}^d \setminus \{0\}$ and W is compact in V (and compact in \mathbb{R}^{d^2}). For any $t \in [F(c), 1]$, let $\varepsilon_t > 0$ with $K(B(t), 2\varepsilon_t) \subseteq V$, where $K(M, r) := \{M' \in \mathbb{R}^{d^2} \mid \|M - M'\| < r\}$. Since W is compact, there exists t_1, \dots, t_r with $W \subseteq \bigcup_{k=1}^r K(B(t_k), \frac{\varepsilon_k}{2})$. Hence $W \subseteq \bigcup_{k=1}^r K(B(t_k), \frac{\varepsilon_k}{2}) \subseteq \bigcup_{k=1}^r K(B(t_k), \varepsilon_k) =: P \subseteq \bar{P} \subseteq \bigcup_{k=1}^r K(B(t_k), 2\varepsilon_k) \subseteq V$. Since \bar{P} is closed and bounded, \bar{P} is also compact. Consequently, $i: \bar{P} \rightarrow V$ is uniformly continuous (Heine-Cantor theorem). Let $\varepsilon > 0$ be arbitrary. Then there is $\delta > 0$ with $\delta < \min\{\varepsilon_k/2 \mid 1 \leq k \leq r\}$ such that $\|M_1 - M_2\| < \delta \Rightarrow \|i(M_1) - i(M_2)\| < \varepsilon$ holds for all $M_1, M_2 \in \bar{P}$. By (78) $\exists n_0 \in \mathbb{N} \forall n \geq n_0 \forall t \in [F(c), 1]: \|A_n(t) - B(t)\| < \delta$. Thus $A_n(t) \in \bar{P}$ follows for any $n \geq n_0$ and $t \in [F(c), 1]$ and hence also $\|i(A_n(t)) - i(B(t))\| < \varepsilon$. Since $\varepsilon > 0$ was arbitrary, $\sup_{t \in [F(c), 1]} \|i(A_n(t)) - i(B(t))\| \xrightarrow{n \rightarrow \infty} 0$ follows. \square

Proof of Lemma 78 from page 37:

$$\begin{aligned} 0 \leq \|a_{n, \lfloor tn \rfloor - K}\|^2 &= \frac{1}{n} f(t_{n, \lfloor tn \rfloor - K})^\top \left(\frac{1}{n} X_{n, \lfloor tn \rfloor - 1 - K}^\top X_{n, \lfloor tn \rfloor - 1 - K} \right)^{-1} f(t_{n, \lfloor tn \rfloor - K}) \\ &\leq \frac{1}{n} \cdot \frac{\|f(t_{n, \lfloor tn \rfloor - K})\|^2}{\chi_{\min}(n, t)} \leq \frac{1}{n} \cdot \frac{\sum_{i=1}^d \max_{t \in [0, 1]} f_i(t)^2}{\chi_{\min}}, \end{aligned}$$

where $\chi_{\min}(n, t) = \chi_{\min} \left(\left(\frac{1}{n} X_{n, \lfloor tn \rfloor - 1 - K}^\top X_{n, \lfloor tn \rfloor - 1 - K} \right) \right)$, χ_{\min} is defined in Lemma 75 and the last inequality holds true by Lemma 75. Thus the assertion follows. \square

Proof of Theorem 79 from page 38:

We divide the proof into two steps. For the proof note that all integrals are Lebesgue-Stieltjes ones, except those denoted by $\int^{(R)}$ which are obtained as the limit of Riemann-Stieltjes sums. Note that $a \leq \inf(\text{support}(F))$, since $(t_{n,1}, \dots, t_{n,n})$ is an asymptotic F -design.

Step 1, all f_i are assumed to be continuous. Let $\xi_{n-d} := M_{n,n}^\top \circ V_n(u) \in \mathbb{R}^{n-d}$. Thus for $x \geq \delta$, $\phi_n(u)(x) = T_{n-d} \circ g(M_{n,n}^\top \circ V_n(u))(x) = T_{n-d} \circ g(\xi_{n-d})(x)$

$$\begin{aligned} &= T_{n-d} \circ g \left(\left(\frac{1}{\sqrt{1 + \|a_{n,d+i}\|^2}} \left[u\left(\frac{d+i}{n}\right) - u\left(\frac{d+i-1}{n}\right) - a_{n,d+i}^\top \begin{pmatrix} u\left(\frac{1}{n}\right) - u(0) \\ \vdots \\ u\left(\frac{d+i-1}{n}\right) - u\left(\frac{d+i-2}{n}\right) \end{pmatrix} \right] \right)_{i=1}^{n-d} \right) (x) \\ &= \underbrace{\left((n-d)x - \lfloor (n-d)x \rfloor \right)}_{=:z} \cdot (\xi_{n-d})_{\lfloor (n-d)x \rfloor + 1} + \sum_{i=\lfloor (n-d)\delta \rfloor + 1}^{\lfloor (n-d)x \rfloor} (\xi_{n-d})_i \quad \text{by (24)} \\ &= z + \sum_{i=i_0}^{i_1} \frac{1}{\sqrt{c_{n,d+i}}} \left[u\left(\frac{d+i}{n}\right) - u\left(\frac{d+i-1}{n}\right) \right] \\ &\quad - \sum_{i=i_0}^{i_1} \frac{1}{\sqrt{c_{n,d+i}}} \left[f(t_{n,d+i})^\top \left(X_{n,d+i-1}^\top X_{n,d+i-1} \right)^{-1} X_{n,d+i-1}^\top \cdot \begin{pmatrix} u(1/n) - u(0) \\ \vdots \\ u\left(\frac{d+i-1}{n}\right) - u\left(\frac{d+i-2}{n}\right) \end{pmatrix} \right] \end{aligned}$$

$$\begin{aligned}
& \left(\text{with } c_{n,d+i} := 1 + \|a_{n,d+i}\|^2, \quad i_0 := \lfloor (n-d)\delta \rfloor + 1 \text{ and } i_1 := \lfloor (n-d)x \rfloor \right) \\
& = z - \sum_{i=i_0}^{i_1-1} \left[\frac{1}{\sqrt{c_{n,d+i+1}}} - \frac{1}{\sqrt{c_{n,d+i}}} \right] u\left(\frac{d+i}{n}\right) + \frac{u\left(\frac{d+i_1}{n}\right) - u\left(\frac{d+i_0-1}{n}\right)}{\sqrt{c_{n,d+i_1}} - \sqrt{c_{n,d+i_0}}} \\
& \quad + \sum_{i=i_0}^{i_1} \frac{1}{n} \cdot \frac{1}{\sqrt{c_{n,d+i}}} f(t_{n,d+i})^\top \left(\frac{1}{n} X_{n,d+i-1}^\top X_{n,d+i-1} \right)^{-1} \left[(\Delta X_{n,d+i-1})^\top \cdot \begin{pmatrix} u(1/n) \\ \vdots \\ u\left(\frac{d+i-2}{n}\right) \end{pmatrix} + z_i^* \right] \\
& \left(\text{where } (\Delta X_{n,d+i-1})^\top := \begin{pmatrix} f_1(t_{n,2}) - f_1(t_{n,1}) & \cdots & f_1(t_{n,d+i-1}) - f_1(t_{n,d+i-2}) \\ \vdots & \ddots & \vdots \\ f_d(t_{n,2}) - f_d(t_{n,1}) & \cdots & f_d(t_{n,d+i-1}) - f_d(t_{n,d+i-2}) \end{pmatrix} \in \mathbb{R}^{d \times (d+i-2)} \right) \\
& \left(\text{and } z_i^* := f(t_{n,1})u(0) - f(t_{n,d+i-1})u\left(\frac{d+i-1}{n}\right) \right) \\
& = z - \int_{[\delta,1]} u(t) dG_n + \frac{u\left(\frac{d+i_1}{n}\right) - u\left(\frac{d+i_0-1}{n}\right)}{\sqrt{c_{n,d+i_1}} - \sqrt{c_{n,d+i_0}}} + \int_{\frac{d+i_0-1}{n}}^{\frac{d+i_1}{n}} \Theta_n(t) dt \\
& \xrightarrow{n \rightarrow \infty} 0 - 0 + u(x) - u(\delta) + \int_{\delta}^x \Theta(t) dt,
\end{aligned}$$

by Lebesgue's dominated convergence theorem, since $\Theta_n(t) \rightarrow \Theta(t)$ holds for all t outside a Lebesgue null set. Below follow the details of the last two steps.

- $z = \frac{(n-d)x - k + 1}{\sqrt{1 + \|a_{n,d+k}\|^2}} \left[u\left(\frac{d+k}{n}\right) - u\left(\frac{d+k-1}{n}\right) + \frac{1}{n} \cdot A(n,x) \right]$ with $k := \lfloor (n-d)x \rfloor + 1$ and
$$\begin{aligned}
A(n,x) & := -n \cdot a_{n,d+k}^\top \begin{pmatrix} u\left(\frac{1}{n}\right) - u(0) \\ \vdots \\ u\left(\frac{d+k-1}{n}\right) - u\left(\frac{d+k-2}{n}\right) \end{pmatrix} \\
& = f(t_{n,d+k})^\top \left(\frac{1}{n} X_{n,d+k-1}^\top X_{n,d+k-1} \right)^{-1} \left[(\Delta X_{n,d+k-1})^\top \begin{pmatrix} u(1/n) \\ \vdots \\ u\left(\frac{d+k-2}{n}\right) \end{pmatrix} + f(t_{n,1})u(0) - f(t_{n,d+k-1})u\left(\frac{d+k-1}{n}\right) \right].
\end{aligned}$$

By Lemmas 75 and 78 and since all f_k are continuous and of bounded variation there is $s \in \mathbb{R}$ with

$$\begin{aligned}
|A(n,x)| & \leq \sup |f| \cdot \left\| \left(\frac{1}{n} X_{n,d+k-1}^\top X_{n,d+k-1} \right)^{-1} \right\|_2 \\
& \quad \cdot \left[\left(\sum_{k=1}^d \sum_{j=2}^{d+k-1} |f_k(t_{n,j}) - f_k(t_{n,j-1})| \right) \cdot \sup |u| + 2 \sup |f| \cdot \sup |u| \right] \leq s
\end{aligned}$$

for all x and all sufficiently large n (note of course that u is also continuous). Thus $z \xrightarrow{n \rightarrow \infty} 0$.

- $\int_{[\delta,1]} u(t) dG_n = \sum_{i=i_0}^{i_1-1} \left[\frac{1}{\sqrt{c_{n,d+i+1}}} - \frac{1}{\sqrt{c_{n,d+i}}} \right] u\left(\frac{d+i}{n}\right)$, where the discrete signed measure G_n is given by the measure generating function $g_n : [\delta, 1] \rightarrow \mathbb{R}$, defined by $g_n(t) := \begin{cases} \frac{1}{\sqrt{c_{n,d+i_0}}} & \text{for } t \in [\delta, \frac{d+i_0}{n}) \\ \frac{1}{\sqrt{c_{n,\lfloor tn \rfloor + 1}}} & \text{for } t \in [\frac{d+i_0}{n}, \frac{d+i_1}{n}) \\ \frac{1}{\sqrt{c_{n,d+i_1}}} & \text{for } t \in (\frac{d+i_1}{n}, 1] \end{cases}$

Note that $\delta < \frac{d+i_0}{n}$ and g_n is a step function and thus of bounded variation. Since $g_n(t)$ converges to 1 uniformly for all $t \in [\delta, 1]$ and $x \in [0, 1]$ by Lemma 78, G_n converges weakly to the zero-measure (i.e. the measure given by the measure generating function $1_{[\delta,1]}$, see Bogachev (2018), Theorem 1.4.7) thus $\int_{[\delta,1]} u(t) dG_n \xrightarrow{n \rightarrow \infty} 0$ independently of x .

- $\Theta_n(t) := \frac{f(t_{n,\lfloor tn \rfloor + 1})^\top}{\sqrt{c_{n,\lfloor tn \rfloor + 1}}} \left(\frac{1}{n} X_{n,\lfloor tn \rfloor}^\top X_{n,\lfloor tn \rfloor} \right)^{-1} \cdot \left[(\Delta X_{n,\lfloor tn \rfloor})^\top \cdot \begin{pmatrix} u(1/n) \\ \vdots \\ u\left(\frac{\lfloor tn \rfloor - 1}{n}\right) \end{pmatrix} + f(t_{n,1})u(0) - f(t_{n,\lfloor tn \rfloor})u\left(\frac{\lfloor tn \rfloor}{n}\right) \right]$

for $t \in \left[\frac{d+i_0-1}{n}, \frac{d+i_1}{n} \right)$ and $\Theta_n(t) := 0$ for $t \in [0, 1] \setminus \left[\frac{d+i_0-1}{n}, \frac{d+i_1}{n} \right)$. Because of Lemmas 75 and 78 and because all f_k are bounded and of bounded variation and

$$|\Theta_n(t)| \leq \left| \frac{1}{\sqrt{c_{n, \lfloor tn \rfloor + 1}}} \right| \cdot \sup |f| \cdot \left\| \left(\frac{1}{n} X_{n, \lfloor tn \rfloor}^\top X_{n, \lfloor tn \rfloor} \right)^{-1} \right\|_2 \cdot \left[\left(\sum_{k=1}^d \sum_{j=2}^{\lfloor tn \rfloor} |f_k(t_{n,j}) - f_k(t_{n,j-1})| \right) \cdot \sup |u| + 2 \sup |f| \cdot \sup |u| \right]$$

for all $t \in \left[\frac{d+i_0-1}{n}, \frac{d+i_1}{n} \right)$, there is $s \in \mathbb{R}$ with $|\Theta_n(t)| \leq s$ for all $t \in [0, 1]$. Furthermore

$$\Theta_n(t) \xrightarrow{n \rightarrow \infty} \Theta(t) := f(h(t))^\top \left(\int_{[0,t]} (ff^\top) \circ h d\lambda \right)^{-1} \left[\int_{[a,h(t)]}^{(R)} (u \circ F) df + f(h(0))u(0) - f(h(t))u(t) \right]$$

holds for all $t \in [\delta, 1]$ where $h := F^-$ is continuous (by Lemma 77 and Lemma 72). Also in this case the convergence is independent of x , in the sense that

$$\sup_{x \in [\delta, 1]} \left| \int_{\frac{d+i_0-1}{n}}^{\frac{d+i_1}{n}} \Theta_n(t) dt - \int_{\delta}^x \Theta(t) dt \right| \xrightarrow{n \rightarrow \infty} 0 \quad (125)$$

holds. To any $\varepsilon > 0$, according to Egorov's theorem, there exists an $A_\varepsilon \subseteq [\delta, 1]$ with $\sup_{t \in A_\varepsilon} |\Theta_n(t) - \Theta(t)| \xrightarrow{n \rightarrow \infty} 0$ and $\lambda([\delta, 1] \setminus A_\varepsilon) < \varepsilon$. To show (125), we now proceed in the same way as for (122).

Hence $\sup_{x \in [\delta, 1]} |\phi(u)(x) - \phi_n(u)(x)| \xrightarrow{n \rightarrow \infty} 0$ holds true. Note that ϕ and ϕ_n are linear and $\|\phi\| < \infty$. Thus we conclude $\|\phi(u) - \phi_n(u_n)\| \leq \|\phi(u) - \phi_n(u)\| + \|\phi_n\| \cdot \|u - u_n\| \xrightarrow{n \rightarrow \infty} 0$ and the assertions of the Theorem are proved in case of step 1.

Step 2, the regression functions are left continuous. Let $\alpha : [0, 1] \rightarrow [a, b]$, $\alpha(x) := (b-a)x + a$. By Lemma 67 there exists continuous $\tilde{g}_i : [0, 1] \rightarrow \mathbb{R}$, $i \in J$, all of bounded variation and a continuous distribution function $\tilde{F} : [0, 1] \rightarrow [0, 1]$ such that $f_i \circ \alpha = \tilde{g}_i \circ \tilde{F}^-$ for all $i \in J$ and $\tilde{F}(0) = 0$, $\tilde{F}(1) = 1$. Thus we conclude with $\beta : [a, b] \rightarrow [0, 1]$, $\beta(x) := \frac{x-a}{b-a}$ and by means of Theorem 57.10 (note $\alpha^{-1} = \alpha^-$, $\beta^{-1} = \beta^-$ and $\beta = \alpha^{-1}$),

$$f_i = \tilde{g}_i \circ \tilde{F}^- \circ \alpha^{-1} = \tilde{g}_i \circ \beta \circ \beta^{-1} \circ \tilde{F}^- \circ \alpha^{-1} = (\tilde{g}_i \circ \beta) \circ (\alpha \circ \tilde{F} \circ \beta)^- = g_i \circ G_0^-,$$

where $g_i := \tilde{g}_i \circ \beta : [a, b] \rightarrow \mathbb{R}$, $i \in J$ are continuous and of bounded variation and $G_0 := \alpha \circ \tilde{F} \circ \beta : [a, b] \rightarrow [a, b]$ is non-decreasing and continuous (note that $G_0(a) = a$ and $G_0(b) = b$). We define $G : \mathbb{R} \rightarrow [a, b]$ by $G(x) := G_0(x)$ for $x \in [a, b]$, $G(x) := G(a)$ for $x < a$ and $G(x) = G(b)$ for $x > b$. Since all f_i are continuous in a , $f_i \circ \alpha$ is continuous in 0 and therefore $\tilde{F}(t) > 0$ for all $t \in (0, 1]$ and $\tilde{F}(t) < 1$ for all $t \in [0, 1)$ by Lemma 67. Thus, by definition of G ,

$$G(t) > a \text{ for all } t > a \text{ and } G(t) < b \text{ for all } t < b. \quad (126)$$

Due to $f_k(t_{n,i}) = g_k(G_0^-(t_{n,i}))$ for all $i \in \{1, \dots, d\}$, we consider the design $(s_{n,1}, \dots, s_{n,n})$ in $[a, b]$, where $s_{n,i} := G_0^-(t_{n,i})$. Next, we prove that $(s_{n,1}, \dots, s_{n,n})$ is an asymptotic H -design, where $H := F \circ G$. Let $H_n : \mathbb{R} \rightarrow [0, 1]$ be the empirical distribution function corresponding to $(s_{n,1}, \dots, s_{n,n})$ with $H_n(s) = 0$ for all $s < a$ and $H_n(s) = 1$ for all $s \geq b$ and let F_n be the empirical distribution function corresponding to $(t_{n,1}, \dots, t_{n,n})$, both according to (26). Then $H_n(s) = F_n(G(s))$ holds for all $s \in [a, b]$, because for all $s \in [a, b]$ and all $t_{n,i}$, $G_0^-(t_{n,i}) \leq s \Leftrightarrow t_{n,i} \leq G(s)$ holds. Thus it follows

$$\sup_{s \in [a, b]} |H_n(s) - H(s)| = \sup_{s \in [a, b]} |F_n(G(s)) - F(G(s))| \xrightarrow{n \rightarrow \infty} 0, \quad (127)$$

because $(t_{n,1}, \dots, t_{n,n})$ is an asymptotic F -design by assumption and the assertion is proven. To apply the previously proved step 1, it remains to show the linear independence of $g_i \cdot 1_{[a,c']}$, $i \in \{1, \dots, d\}$ in $L_2([a,b], F \circ G)$ for some $c' > a$. Let $c' := G_0^-(c)$. Note that $a \leq \inf(\text{support}(F))$.

$$\begin{aligned}
& g_1 \cdot 1_{[a,c']}, \dots, g_d \cdot 1_{[a,c']} \text{ are l.i. in } L_2([a,b], F \circ G) \\
& \Leftrightarrow \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i g_i \cdot 1_{[a,c']} \right)^2 d\mu_{F \circ G} > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i g_i \cdot 1_{(a,c']} \right)^2 d\mu_{F \circ G} + \left(\sum_{i=1}^d \alpha_i g_i(a) \right)^2 \cdot F(G(a)) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \stackrel{(73)}{\Leftrightarrow} \int_{[F(G(a)), F(G(c'))]} \left(\sum_{i=1}^d \alpha_i g_i \circ (F \circ G)_0^- \right)^2 d\lambda + \left(\sum_{i=1}^d \alpha_i g_i(G_0^-(a)) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow \int_{[F(a), F(G(c'))]} \left(\sum_{i=1}^d \alpha_i g_i \circ G_0^- \circ F_0^- \right)^2 d\lambda + \left(\sum_{i=1}^d \alpha_i f_i(a) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \stackrel{(73)}{\Leftrightarrow} \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i f_i \cdot 1_{(a, G(c'))} \right)^2 d\mu_F + \left(\sum_{i=1}^d \alpha_i f_i(a) \right)^2 \cdot F(a) > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow \int_{\mathbb{R}} \left(\sum_{i=1}^d \alpha_i f_i \cdot 1_{[a, G(c'))} \right)^2 d\mu_F > 0, \quad \forall \alpha \in \mathbb{R}^d \setminus \{0\} \\
& \Leftrightarrow f_1 \cdot 1_{[a, G(c')]}, \dots, f_d \cdot 1_{[a, G(c')]} \text{ are l.i. in } L_2([a,b], F),
\end{aligned}$$

with $(F \circ G)_0 := (F \circ G)|_{[F(a), G(b)]}$ and $F_0 := F|_{[a,b]}$. Note that $G_0^-(a) = a$. Since $G(c') \geq c > a$, the last equivalence is satisfied by assumption. Now we apply the previously proven step 1 and thus immediately obtain $\phi_n(u) \rightarrow \phi(u)$ regarding sup-norm, where

$$\begin{aligned}
& \phi(u)(x) := u(x) - u(H_0(c')) \\
& + \int_{H_0(c')}^x g(H_0^-(t))^\top \left(\int_{[0,t]} (g g^\top) \circ H_0^- d\lambda \right)^{-1} \left[\int_{[a, H_0^-(t)]}^{(R)} (u \circ H_0) dg + g(H_0^-(0))u(0) - g(H_0^-(t))u(t) \right] dt,
\end{aligned}$$

with $H_0 := H|_{[a,b]} = F_0 \circ G_0$. Finally we use $g \circ H_0^- = f \circ F_0^-$, $H_0(c') = F_0(G_0(G_0^-(c))) = F_0(c) = \delta$ (because G_0 is continuous with $G_0(a) = a$, $G_0(b) = b$, thus $c \in [a,b] = G_0([a,b])$), cf. Lemma 57.7) and

$$\int_{[a, H_0^-(t)]}^{(R)} (u \circ H_0) dg = \int_{(a, F_0^-(t))} (u \circ F_0) d\mu_f, \quad (128)$$

where μ_f is the Lebesgue-Stieltjes measure generated by f . It remains to prove (128). To do so, we use the change-of-variables formula $\int_{\Omega'} \kappa d\mu^T = \int_{\Omega} (\kappa \circ T) d\mu$ for $\kappa : \Omega' \rightarrow \mathbb{R}$ and $T : \Omega \rightarrow \Omega'$, see e.g. Stroock (1994), Lemma 5.0.1, where in our case $\kappa := u \circ F_0$ and $T(x) := G_0(x)$, $x \in \Omega$ holds with $\Omega = (a, G_0^-(F_0^-(t)))$ and $\Omega' = (a, G_0(G_0^-(F_0^-(t))))$. First, we note

that $G_0(G_0^-(F_0^-(t))) = F_0^-(t)$ holds (same argument as above with $G_0(G_0^-(c)) = c$). Next, we show that $T(\Omega) = \Omega'$ holds. For this, let $a < x < G_0^-(F_0^-(t))$. Then $G_0(x) > a$ by (126) and $G_0(x) < F_0^-(t)$ by Lemma 57.5. Thus $T(\Omega) = \Omega' = (a, F_0^-(t))$ holds. Now we show $\mu_g^{G_0} = \mu_f$ on $\mathcal{B}((a, F_0^-(t)))$. For this it is enough to show the equality on $\mathcal{E} := \{E \cap (a, F_0^-(t)) \mid E \in \mathcal{E}_0\}$, where $\mathcal{E}_0 := \{[x, y] \mid x \leq y\}$ is a generator of $\mathcal{B}(\mathbb{R})$. Since

$$\mathcal{E} = \{[x, y] \mid a < x \leq y \leq F_0^-(t)\} \cup \{(a, y) \mid a < y \leq F_0^-(t)\},$$

$$\begin{aligned} \mu_g^{G_0}([x, y]) &= \mu_g(G_0^{-1}([x, y])) = \mu_g([G_0^-(x), G_0^-(y)]) = g(G_0^-(y)) - g(G_0^-(x)) \\ &= f(y^-) - f(x^-) \quad (\text{since } f \text{ is left continuous}) \\ &= \mu_f([x, y]) \quad (\text{cf. Theorem 63}) \end{aligned}$$

and

$$\mu_g^{G_0}((a, y)) = \lim_{n \rightarrow \infty} \mu_g^{G_0}([a + 1/n, y]) = \lim_{n \rightarrow \infty} \mu_f([a + 1/n, y]) = \mu_f((a, y)),$$

$\mu_g^{G_0} = \mu_f$ holds on $\mathcal{B}((a, F_0^-(t)))$. Thus

$$\begin{aligned} \int_{[a, H_0^-(t)]}^{(R)} (u \circ H_0) \, dg &\stackrel{(*)}{=} \int_{(a, G_0^-(F_0^-(t)))} (u \circ F_0 \circ G_0) \, d\mu_g \quad (\text{note that } g \text{ is continuous}) \\ &= \int_{(a, G_0(G_0^-(F_0^-(t))))} (u \circ F_0) \, d\mu_g^{G_0} = \int_{(a, F_0^-(t))} (u \circ F_0) \, d\mu_f. \end{aligned}$$

For (*) note that the Riemann-Stieltjes integral is equal to the Lebesgue-Stieltjes integral (cf. Kirillov and Gvishiani (1982), Theorem 14, p. 29 or Stroock (1994), Theorem 5.1.2). \square

Part D. Appendices

1 List of figures and tables and R scripts generating them

R script to generate Figure 1 from page 10.

```
1 delta = 0.001
2 q0 = 30 # number of simulations
3 d = 3 # regression functions f_1, f_2, ..., f_d
4 # in this example f_1(t)=1, f_2(t)=t, f_3(t)=t^2 and t_ni=i/n
5 n = 10000 # d<n
6 betaH0 = c(1,1,1)
7 betaH1 = c(1,1,1,1)
8 for (q in 1:q0) {
9   epsilon = c(runif(1)-0.5) # error
10  for (i in 2:n) {
11    epsilon = c(epsilon,runif(1)-0.5)
12  }
13  vH0 = c((sign(1/n -0.25)+3)/4,1/n, (1/n)^2)
14  XH0 = rbind(vH0) # design matrix under H0
15  for (i in 2:n) {
16    if (i/n <= 0.5) {
17      vH0 = c((sign(i/n -0.25)+3)/4,i/n, (i/n)^2) # v_i = (f_1(t_ni),...,f_d(t_ni))
18    } else {
19      vH0 = c((sign(i/n -0.25)+3)/4,i/n, (i/n -3/2)^2)
20    }
21    XH0 = rbind(XH0,vH0)
22  }
23  vH1 = c((sign(1/n -0.25)+3)/4,1/n, (1/n)^2, (1/n)^3)
24  XH1 = rbind(vH1) # design matrix under H1
25  for (i in 2:n) {
26    if (i/n <= 0.5) {
27      vH0 = c((sign(i/n -0.25)+3)/4,i/n, (i/n)^2, (i/n)^3) # v_i = (f_1(t_ni),...,f_d(t_ni))
28    } else {
29      vH0 = c((sign(i/n -0.25)+3)/4,i/n, (i/n -3/2)^2, (i/n)^3)
30    }
31    XH0 = rbind(XH0,vH0)
32  }
33  YH0 = XH0 %*% betaH0 + epsilon # observations under H0
34  YH1 = XH1 %*% betaH1 + epsilon # observations under H1
35  # the following function "recresid" requires R package "strucchange"
36  # library(strucchange)
37  rr = recresid(XH0,YH0) # B_n under H_0 resp. H_1
38  z1 = NULL;
39  for (i in 1:(n-d)) {
40    z1[i] = i/(n-d)
41  }
42  z2 = NULL;
43  i1 = floor((n-d)*delta)
44  for (i in 1:i1) {
45    z2[i] = 0
46  }
47  for (i in (i1+1):(n-d)) {
48    z2[i] = sum(rr[i1+1:i])
49  }
50  z2 = 2*sqrt(3)*z2/sqrt(n-d)
```

```

51 plot(z1,z2,type="l",ylim =c( -3 ,3),main = "Recursive residual partial sum
    process",xlab="t (time)")
52 par(new=T)
53 }
54 # text(x=1,y=1.5,paste("n=100"))

```

R script to generate Figure 2 from page 12.

```

1 delta = 0.001
2 x = 2.573253
3 y = -2.573253
4 z = 2.8042267
5 numSim = 30 # number of simulations
6 N = 10000 # number of endpoints of the decomposition, including T
7 T = 1 # end point of the interval [0,T]
8 D = T/N # time unit (time step)
9 B = NULL # Initialization of the vector B
10 B[1]=0; B[2]=0; B[3]=0; B[4]=0; B[5]=0; B[6]=0; B[7]=0; B[8]=0; B[9]=0; B[10]=0;
11 for (i in 11:(N+1)) {
12 B[i] = B[i-1] + rnorm(1)*sqrt(D)
13 }
14 t = seq(0,T,length=length(B))
15 plot(t,B,type ="l",main = "Brownian Motion",xlab="t (time)",ylim =c( -3 ,3))
16 for (i in 1:(numSim - 1)) {
17 par(new=T)
18 B = NULL; B[1]=0 # Initialization of the vector B
19 for (i in 2:(N+1)) {
20 B[i] = B[i-1] + rnorm(1)*sqrt(D)
21 }
22 plot(t,B,type ="l",ylim =c( -3 ,3))
23 }
24 abline(h=c(x,y,z,-z), col=c("green","blue","red","red"))
25 legend(0, 3, legend=c("+-2.8042267", "-2.573253", "2.573253"),
26 col=c("red", "blue", "green"), lty=1, cex=1)

```

R script to compute the bounds in Remark 12 from page 12.

```

1 alpha = 0.05
2 z = 2.24140277
3 x = qnorm(p=1-alpha/2,mean=0,sd=1)
4 y = qnorm(p=alpha/2,mean=0,sd=1)
5 summe = 0
6 m = 10000 # summation from -m to m
7 for (n in (-m):m) {
8 A1 = pnorm((4*n+1)*z,mean=0,sd=1)
9 A2 = pnorm((4*n-1)*z,mean=0,sd=1)
10 A3 = pnorm(-(4*n-3)*z,mean=0,sd=1)
11 A4 = pnorm(-(4*n-1)*z,mean=0,sd=1)
12 summe = summe + A1 - A2 - A3 + A4
13 }
14 x;y;l-summe

```

R script to generate Figure 3 from page 13.

See R-script from page 75 for Figure 1.

R script to generate Figure 4 from page 14.

See R-script from page 75 for Figure 1.

R script to generate Figure 5 from page 15.

```
1 delta = 0.001
2 q0 = 1 # number of sample paths
3 d = 1
4 # in this example f_1(t)=1 and t_ni=i/n
5 n = 10000 # d<n
6 betaH0 = c(0.5)
7 # betaH1 = c(0.5,1)
8 betaH1 = c(0.5,sqrt(n-d)/n)
9 for (q in 1:q0) {
10   epsilon = c(runif(1)-0.5) # error
11   for (i in 2:n) {
12     epsilon = c(epsilon,runif(1)-0.5)
13   }
14   vH0 = c(1)
15   XH0 = rbind(vH0) # design matrix under H0
16   for (i in 2:n) {
17     vH0 = c(1)
18     XH0 = rbind(XH0,vH0)
19   }
20   vH1 = c(1,1/n)
21   XH1 = rbind(vH1) # design matrix under H1
22   for (i in 2:n) {
23     vH1 = c(1,i/n) # v_i = (f_1(t_ni),...,f_d(t_ni))
24     XH1 = rbind(XH1,vH1)
25   }
26   YH0 = XH0 %*% betaH0 + epsilon # observations under H0
27   YH1 = XH1 %*% betaH1 + epsilon # observations under H1
28   # the following function "recresid" requires R package "strucchange"
29   rr0 = recresid(XH0,YH0) # recursive residuals under H_0
30   rr1 = recresid(XH0,YH1) # recursive residuals under H_1
31   z1 = NULL;
32   for (i in 1:(n-d)) {
33     z1[i] = i/(n-d)
34   }
35   z20 = NULL;
36   i1 = floor((n-d)*delta)
37   for (i in 1:i1) {
38     z20[i] = 0
39   }
40   for (i in (i1+1):(n-d)) {
41     z20[i] = sum(rr0[(i1+1):i])
42   }
43   z20 = 2*sqrt(3)*z20/sqrt(n-d)
44   z21 = NULL;
45   i1 = floor((n-d)*delta)
46   for (i in 1:i1) {
47     z21[i] = 0
48   }
49   for (i in (i1+1):(n-d)) {
50     z21[i] = sum(rr1[(i1+1):i])
51   }
52   z21 = 2*sqrt(3)*z21/sqrt(n-d)
53   plot(z1,z21-z20,type="l",ylim=c(-3,3),main="Recursive residual partial sum
54 process under H0 and H1",xlab="t (time)")
55   par(new=T)
56   plot(z1,z20,type="l",col="blue",ylim=c(-3,3),xlab="t (time)")
57   par(new=T)
58   plot(z1,z21,type="l",col="red",ylim=c(-3,3),xlab="t (time)")
59   if (q==q0) {
60     break
61   }
62   par(new=T)
```

```

63 }
64 N = 10000 # number of points of the decomposition, including T
65 T = 1 # end point of the interval [0,T]
66 D = T/N # time unit (time step)
67 shift = NULL
68 for (i in 1:(delta*N)) {
69   shift[i] = 0
70 }
71 for (i in (delta*N+1):(N+1)) {
72   shift[i] = 0.25*sqrt(12)*(((i-1)/N)^2 - delta^2)
73 }
74 for (q in 1:q0) {
75   B = NULL # Initialization of the vector B
76   for (i in 1:(delta*N)) {
77     B[i] = 0
78   }
79   for (i in (delta*N+1):(N+1)) {
80     B[i] = B[i-1] + rnorm(1)*sqrt(D)
81   }
82   t = seq(0,T,length=length(B))
83   B0 = B + shift
84   plot(t,B0,type="l",col="red",main="Brownian motion with and without trend",xlab="t
      (time)",ylim=c(-3,3))
85   par(new=T)
86   plot(t,B,type="l",col="blue",ylim=c(-3,3),xlab="t (time)")
87   par(new=T)
88   plot(t,shift,type="l",ylim=c(-3,3),xlab="t (time)")
89   if (q==q0) {
90     break
91   }
92   par(new=T)
93 }
94 text(x=0.98,y=2,paste("H0/H1/difference")) # Labeling of the paths

```

R script to generate Figure 6 from page 18.

See R-script from page 77 for Figure 4, noting the following minor changes. Line 41 must be replaced by “`z20[i] = sum(rr0[(n-d):(n-d-i+1+1)])`”, line 50 must be replaced by “`z21[i] = sum(rr1[(n-d):(n-d-i+1+1)])`” and line 72 must be replaced by “`shift[i] = 0.25*sqrt(12)*((i-1)/N - delta)*(2 + delta - (i-1)/N)`”.

R script to generate Figure 7 from page 30.

```

1 N = 100000 # number of points of the decomposition, including T
2 T = 1 # end point of the interval [0,T]
3 D = T/N # time unit (time step)
4 B = NULL; B[1]=0 # Initialization of the vector B
5 for (i in 2:(N+1)) {
6   B[i] = B[i-1] + rnorm(1)*sqrt(D)
7 }
8 t = seq(0,T,length=length(B))
9 plot(t,B,type="l",main="Brownian Motion",xlab="t (time)",ylim=c(-1,1))

```

Figure from page 62

Created by the author with Microsoft Paint.

2 New results

(“Thesen zur Dissertation”)

Unless stated otherwise, the following statements and their proofs are new and cannot be found in the literature so far.

- Computation of the limit process under the null hypothesis in Theorem 10 and Theorem 21 and under a local alternative in Lemma 14, Theorem 15 and Theorem 23, where the technical basis for the computation under the null hypothesis and also under the local alternative is given in Theorem 79.
- The tests in Theorem 11, Theorem 22, and Theorem 28.
- Generalization of some classical results on weak convergence of measures to filter of measures: Lemma 38, Theorem 40, Theorem 41, Theorem 42 and Theorem 46
- Lemma 67 on the factorization of functions with existing left and right limits.
- Lemma 70 and Lemma 72 on asymptotic F -designs and Example 71 motivating the need for certain assumptions.
- Lemma 77 on the uniform convergence of certain matrices.
- In direct collaboration with Wolfgang Bischoff:
Lemma 73, Lemma 74, Lemma 75 and Lemma 78 to estimate the eigenvalues of certain matrices. The authors contribution is to adapt the original lemmas (which were communicated to the author in private correspondence) to the situation given here and modify the proofs accordingly.
- In direct collaboration with Thomas Heindl¹⁵:
Theorem 19 about the least squares residual partial sum limit process was developed jointly by Heindl and Evers.
Lemma 60.1 about the probability integral transformation was first discovered and proved by Thomas Heindl. The authors contribution is an alternative and slightly shorter proof.
Lemma 66, as a more or less direct application of the change of variable formula, seems to be known at least in a similar form in the literature, cf. e.g. Shorack (2017), Exercise 3.3 or Bogachev (2007), Example 3.6.2, but was jointly developed and proved in this concrete formulation by Heindl and Evers.
Lemma 69 on the simultaneous factorization of finitely many functions of bounded variation is also from Heindl (2022). The authors contribution consists in two improved and shortened proofs for this Lemma, both based on the measure-theoretic approach given in Heindl (2022).

¹⁵It should be noted here that our two dissertations were written at the same time and we regularly discussed new results. Some references to Heindl (2022) might therefore be (slightly) inaccurate, since the final results and their proofs might still have been changed by Heindl without the author’s awareness.

3 List of abbreviations and symbols

We follow the generally known notation as far as possible. In particular, this applies to all basic operations on sets and numbers. To accommodate the reader, some more special notations are listed below. In some cases, a reference to the first occurrence is given. In the case of symbols with multiple meanings, their meaning is always clear from the respective context.

Abbreviations:

a.a., a.e.	almost all, almost everywhere
iff	if and only if
l.i.	linear independent
r.v.	random variable

Set theory:

$\mathbb{N}, \mathbb{Z}, \mathbb{Q}, \mathbb{R}$	sets of natural, integer, rational, real numbers ($0 \in \mathbb{N}$)
$\lfloor x \rfloor, \lceil x \rceil$	greatest integer less than or equal to x , least integer greater than or equal to x
$\mathbb{N}^{\geq 1}$	set of positiv natural numbers
$\mathcal{P}(X)$	powerset of X
\bullet_x	filter generated by x , cf. Definition 29
B^A	set of all functions $f : A \rightarrow B$, A, B sets
id	identity function $\text{id}(x) = x$
$f^{-1}(B)$	preimage or inverse image of a set B under a mapping f
$f \circ g$	composition of the mappings f and g
$f _A$	restriction of the mapping f to $A \cap \text{dom}(f)$, where $\text{dom}(f)$ is the domain of f
1_A	indicator function, $1_A(x) = 1$ for $x \in A$ and $1_A(x) = 0$ for $x \notin A$
$\mathcal{F}(X), \mathcal{F}_U(X)$	set of all filter/ultrafilter on X
$[\alpha], \alpha \subseteq \mathcal{P}(X)$	filter generated by α , if $A_1 \cap \dots \cap A_n \neq \emptyset$ for all $A_1, \dots, A_n \in \alpha$ $[\alpha] := \{P \subseteq X \mid \exists \emptyset \neq \alpha' \subseteq \alpha, \alpha' : \text{finite, with } \bigcap_{A \in \alpha'} A \subseteq P\}$

Topology:

(X, τ)	topological space (set X with topology τ)
$\bar{A}, A^\circ, \partial A$	closure, interior and boundary of a subset A of a topological space
(E, d)	metric space (set E with metric d)
$d(A, B), d(x, A)$	distance from A to B with respect to some metric space (X, d) , $A, B \subseteq X$, $d(A, B) := \inf\{d(a, b) \mid (a, b) \in A \times B\}$ and $d(x, A) := d(\{x\}, A)$
$K(x, \varepsilon), K_d(x, \varepsilon)$	open ball centered at x with radius ε , $K(x, \varepsilon) = K_d(x, \varepsilon) = \{y \in E \mid d(x, y) < \varepsilon\}$
$\tau_{\mathbb{R}}, \tau_{\mathbb{R}^n}$	Topology on \mathbb{R} and \mathbb{R}^n , respectively, induced by Euclidean metric
$\phi \xrightarrow{\tau} a, a_n \xrightarrow{\tau} a$	convergence with respect to a topology τ , cf. Definition 30
$a_n \rightarrow a, a_n \xrightarrow{n \rightarrow \infty} a$	convergence of a sequence (in the given context)
$f(r^-), f(r^+)$	left and right limit $f(r^-) = \lim_{r_n \nearrow r} f(r_n)$ and $f(r^+) = \lim_{r_n \searrow r} f(r_n)$, cf. Def. 54
I^-, I^+	left resp. right endpoint of a (bounded or unbounded) interval $I \subseteq \mathbb{R}$
$C(X, Y)$	set of all continuous functions $f : X \rightarrow Y$, for topological spaces X, Y
$C[a, b]$	set of all continuous functions $f : [a, b] \rightarrow \mathbb{R}$
$B(X, E)$	set of all bounded functions $f : X \rightarrow E$, X set, E metric space
$BV[a, b]$	set of all functions $f : [a, b] \rightarrow \mathbb{R}$ of bounded variation
$V(f, a, b)$	variation of f on $[a, b]$, cf. Definition 64
$c(f)$	set of points of continuity of the function f
T_n	partial sum operator, cf. Definition 3

















Linear algebra:





















$\mathbb{K}^{m \times n}$	vector space of $m \times n$ matrices with entries from the field \mathbb{K}
$\ker(\alpha)$	kernel of the linear map α
$\text{im}(\alpha)$	image of the map α
$\dim(U)$	dimension of the linear space U
$\text{rank}(A)$	rank of the matrix A
I_n	n dimensional identity matrix $I_n = (e_{i,j})_{i,j=1}^n$, $e_{i,j} = 1$ if $i = j$, $e_{i,j} = 0$ if $i \neq j$
A^\top	transpose of the matrix A
A^{-1}	inverse of the matrix A
$\det(A)$	determinant of the matrix A
U^\perp	orthogonal complement of a subspace U with respect to a scalar product
$\text{pr}_{U,V}$	projection on the first summand of a direct sum $\text{pr}_{U,V} : U \oplus V \rightarrow U$, $u + v \mapsto u$
pr_U	$\text{pr}_U := \text{pr}_{U,U^\perp}$
r_n	$(r_{n,d+1}, \dots, r_{n,n})^\top$ vector of recursive residuals, cf. (15)
$M_{n,k}$	cf. (15)
$\langle \cdot, \cdot \rangle, \ \cdot\ $	general scalar product and norm on some linear space
$\ a\ _2, \ A\ _2$	Euclidean norm on \mathbb{R}^n , $\ a\ _2 = \sqrt{\sum_{i=1}^n a_i^2}$ and matrix norm $\ A\ _2 := \sup_{x \neq 0} \frac{\ Ax\ _2}{\ x\ _2}$





















Measure Theory:






















$(\Omega, \mathcal{A}, \mu)$	measure space (set Ω with sigma-algebra \mathcal{A} and measure μ)
μ^f or $f(\mu)$	image measure (also known as pushforward measure), $\mu^f(A) := \mu(f^{-1}(A))$
$\sigma(\alpha)$	sigma-algebra generated by α , cf. (63)
$\mathcal{B}(E)$	Borel sigma-algebra of E , generated by open sets in E
λ	Lebesgue measure, unless defined otherwise
$M(E)$	$M(E) := \{p : \mathcal{B}(E) \rightarrow \mathbb{R} \mid p \text{ is a finite measure}\}$, cf. p. 25
$\int_E f d\psi$	if ψ is a filter, see Definition 37 and the passage directly above it
$\psi \xrightarrow{w} p, p_n \xrightarrow{w} p$	weak convergence of measures, cf. Definition 37
$\phi \xrightarrow{\mathcal{D}} X, X_n \xrightarrow{\mathcal{D}} X$	weak convergence of r. v. (convergence in distribution), cf. Definition 43
$B, (B_t)_{t \in [0,1]}$	Brownian motion, cf. Definition 49
$B', (B'_t)_{t \in [0,1]}$	shifted Brownian motion, $B'_t(\omega) = 0$ for all $t \in [0, \delta)$ and $B'_t(\omega) = B_{t-\delta}(\omega)$ for $t \in [\delta, 1]$, cf. Theorem 10
X^\bullet	Definition 26
F^-, F^+	generalized inverse of monotone function, cf. Definition 56 and Remark 58
$V(f, a, b)$	variation of f on $[a, b]$, cf. Definition 64
$E(X)$	expected value of a random variable X
$\text{Var}(X)$	variance of a random variable X
$\text{Cov}(X, Y)$	covariance of two random variables X, Y
$B_n, B_{n,\delta}, B_n^{(b)}, B_{n,\delta}^{(b)}$	recursive residual partial sum process, resp. reversed residual partial sum process, cf. (21) resp. (45)
$\text{support}(P)$	support of a measure (resp. a measure defining function), cf. Definition 6
$L_2([a, b], F)$	space of square-integrable functions on $[a, b]$ with respect to the measure defining function F
$\int_I^{(R)} f d\mathbf{g}$	integrals marked by (R) are obtained as the limit of Riemann-Stieltjes sums all other integrals are understood as general measure integrals

Bibliography

-  ANDERSON, T.W. (1963), Asymptotic Theory for Principal Component Analysis, Ann. Math. Statist. 34 (1) 122 - 148
-  BALTAGI, B.H. (2011), Econometrics, Springer
-  BARTLETT, M.S. (1951), An inverse matrix adjustment arising in discriminant analysis, Ann. Math. Statist. 22, pp. 107-111, 1951
-  BAUER, H. (2001), Measure and Integration Theory, Walter de Gruyter
-  BEATTIE, R., BUTZMANN, H.-P. (2002), Convergence Structures and Applications to Functional Analysis, Springer
-  BHATIA, R. (1997), Matrix Analysis, Springer
-  BILLINGSLEY, P. (1968), Convergence of probability measures, John Wiley and Sons
-  BILLINGSLEY, P. (1999), Convergence of probability measures, Wiley
-  BINZ, E. (1975), Continuous Convergence on $C(X)$, Springer
-  BISCHOFF, W. (1998), A functional central limit theorem for regression models, Ann. Statist., Vol. 26, No. 4, pp. 1398-1410
-  BISCHOFF, W. (2014), Brownsche Bewegung und funktionale zentrale Grenzwertsätze, Lecture notes
-  BISCHOFF, W. (2016), On Designs for Recursive Least Squares Residuals to Detect Alternatives. In: KUNERT J., MÜLLER C., ATKINSON A. (eds) (2016), mODa 11 - Advances in Model-Oriented Design and Analysis, Springer
-  BISCHOFF, W., HASHORVA, E., HÜSLER, J., MILLER, F. (2005), Analysis of a change-point regression problem in quality control by partial sums processes and Kolmogorov type tests, Metrika, 62(1), 85-98
-  BISCHOFF, W., MILLER, F (2000), Asymptotically Optimal Tests and Optimal Designs for Testing the Mean in Regression Models with Applications to Change-Point Problems, Annals of the Institute of Statistical Mathematics, 52(4), 658-679
-  BOGACHEV, V.I. (2007), Measure Theory, Vol. 1 and 2, Springer
-  BOGACHEV, V.I. (2018), Weak Convergence of Measures, American Mathematical Society

-  BORODIN, A.N., SALMINEN, P. (2002), Handbook of Brownian Motion - Facts and Formulae, Springer
-  BOROVKOV, A.A. (2013), Probability Theory, Springer
-  BOURBAKI, N. (1966), General Topology, Part 1 and 2, Addison-Wesley
-  BROWN, R.L., DURBIN, J. AND EVANS, J.M. (1975), Techniques for Testing the Constancy of Regression Relationships over Time, J. Roy. Stat. Soc., Series B, Vol. 37, No. 2, pp. 149-192
-  CARTAN, H. (1937), Theorie des Filtrés, C.R. Acad. Sci. Paris 205, 595-598
-  COMFORT, W. W., NEGREPONTIS, S. (1974), The theory of ultrafilters, Springer
-  CSÖRGO, M., HORVATH, L. (1997), Limit Theorems in Change-Point Analysis, Wiley
-  DALEY, D.J., VERE-JONES, D. (2003), An Introduction to the Theory of Point Processes, Volume I: Elementary Theory and Methods, Springer
-  DAVIDSON, J. (1994), Stochastic Limit Theory, Oxford University Press
-  DOLECKI, S., MYNARD, F. (2016) Convergence Foundations Of Topology, World Scientific
-  EISENBERG, B., SHIXIN, G (1983), Uniform convergence of distribution functions, Proceedings of the American Mathematical Society 88:145-146
-  ELSTRODT, J. (2018), Maß- und Integrationstheorie, Springer
-  FAREBROTHER, R.W. (1978), An historical note on recursive residuals, Journal of the Royal Statistical Society B 40, pp. 373-375
-  HACKL, P. (2012), Einführung in die Ökonometrie, Pearson Studium
-  HARVEY, A.C. (1989), Forecasting Structural Time Series Models and the Kalman Filter, Cambridge University Press
-  HARVILLE, D.A. (2008), Matrix Algebra From a Statistician's Perspective, Springer
-  HART, K.P., NAGATA, J., VAUGHN, J.E. (EDS.) (2004), Encyclopedia of General Topology, Elsevier
-  HAWKINS, D.M. (1980), Identification of Outliers, Chapman & Hall
-  HAWKINS, D.M., OLWELL, D.H. (1998), Cumulative Sum Charts and Charting for Quality Improvement, Springer
-  HEINDL, T. (2022), Cumulated sum processes of residuals for goodness-of-fit tests in linear regression models, Dissertation, Katholische Universität Eichstätt - Ingolstadt

-  HEINONEN, J. (2001), Lectures on Analysis on Metric Spaces, Springer
-  HERRMANN, F. (2005), Integration und Volatilität bei Emerging Markets, Springer
-  HIDA, T. (1980) Brownian Motion, Springer
-  HINDMAN, N., STRAUSS, D. (2012), Algebra in the Stone-Cech compactification, Walter de Gruyter
-  JECH, T. (2003), Set Theory, Springer
-  KALLENBERG, O. (2002), Foundations of Modern Probability, Springer
-  KANNAN, R., KRUEGER, C.K. (1996), Advanced Analysis - on the Real Line, Springer
-  KENNEDY, P. (2008), A Guide to Econometrics, Blackwell Publishing
-  KIRILLOV, A.A., GVISHIANI, A.D. (1982), Theorems and Problems in Functional Analysis, Springer
-  KLENKE, A. (2020), Wahrscheinlichkeitstheorie, Springer-Verlag
-  KRAMER, W., SONNBERGER, H. (1986), The Linear Regression Model Under Test, Physica-Verlag
-  KUNEN, K., VAUGHAN, J.E. (EDS) (1984), Handbook of Set-theoretic Topology, North-Holland
-  LOJASIEWICZ, S. (1988), An Introduction to the Theory of Real Functions, Wiley
-  LUNDELL, A.T., WEINGRAM, S. (1969), The Topology of CW Complexes, Van Nostrand Reinhold
-  LUSCHGY, H. (1991) Testing one-sided hypotheses for the mean of a gaussian process, *Metrika*, 38(1), 179-194
-  MACNEILL, I. (1978a), Properties of sequences of partial sums of polynomial regression residuals with applications to tests for change of regression at unknown times, *Ann. Statist.* 6 (2), pp. 422 - 433
-  MACNEILL, I. (1978b), Limit Processes for Sequences of Partial Sums of Regression Residuals, *Annals of Probability*, 6, No. 4, 695-698
-  MYNARD, F., PEARL, E. (EDS.) (2009), Beyond Topology, Contemporary Mathematics, vol. 486, American Mathematical Society
-  NEL, L. (2016), Continuity Theory, Springer
-  OTTO, S. (2019) Three Essays on Structural Stability of Time Series Models, Dissertation, Universität zu Köln

-  PARTZSCH, L., SCHILLING, R.L. (2012), Brownian Motion, Walter de Gruyter
-  PAULSEN, V.I., RAGHUPATHI, M. (2016), An Introduction to the Theory of Reproducing Kernel Hilbert Spaces, Cambridge University Press
-  PLACKETT, R.L. (1950), Some theorems in least squares, *Biometrika* 37, pp. 149-157
-  PLOBERGER, W., KRÄMER, W. (1990), The local power of the cusum and cusum of squares tests, *Econometric Theory*, 6, pp. 335-347
-  PREUSS, G. (2002), Foundations of Topology - An Approach to Convenient Topology, Kluwer Academic Publisher
-  RABOVSKI, O. (2003), Asymptotische Tests basierend auf rekursiven Residuen von Regressionsmodellen, Diplomarbeit
-  R CORE TEAM (2021), R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>
-  RUBIN, H. (unpublished), Topological properties of measures on topological spaces
-  SEN, P.K. (1982), Invariance Principles for Recursive Residuals, *Ann. Stat.* Vol. 10, No. 1, pp. 307-312
-  SEN, P.K. (1985), Theory and Applications of Sequential Nonparametrics, Society for Industrial and Applied Mathematics
-  SAKHANENKO, A., KOVALEVSKII, A., SHELEPOVA, A. (2021), Remarks on invariance principle for one-parametric recursive residuals, *Sib. Elektron. Mat. Izv.*, 18:2, 1058-1074
-  SHORACK, G.R. (2017), Probability for Statisticians, Springer
-  STROOCK, D.W. (1994), A Concise Introduction to the Theory of Integration, Birkhäuser
-  TAO, T. (2011), An Introduction to Measure Theory, American Mathematical Society
-  TOPSOE, F. (1970), Topology and Measure, Springer
-  VAART, A.W. VAN DER (1998), Asymptotic Statistics, Cambridge University Press
-  WALKER, R.C. (1974), The Stone-Cech Compactification, Springer
-  WELLS, C. (1996), The Kalman Filter in Finance, Springer
-  YOUNG, P.C. (2011), Recursive Estimation and Time-Series Analysis, Springer
-  ZEILEIS A, LEISCH F, HORNIK K, KLEIBER C (2002), strucchange: An R Package for Testing for Structural Change in Linear Regression Models, *Journal of Statistical Software*, 7(2), 1-38 <https://CRAN.R-project.org/package=strucchange>
-  ZELENYUK, Y.G. (2011), Ultrafilters and Topologies on Groups, Walter de Gruyter