Technische Universität Dresden
Fachrichtung Mathematik
Institut für Mathematische Stochastik

Weak Convergence of the Empirical Process and the Rescaled Empirical Distribution Function in the Skorokhod Product Space

Diplomarbeit
zur Erlangung des ersten akademischen Grades
Diplommathematiker

vorgelegt von

Daniel Vogel,
geboren am 14. März 1980 in Erfurt

Tag der Einreichung: 1. November 2005
Betreuer: Prof. Dr. rer. nat. habil. D. Ferger

Contents

1 Introduction

This paper brings together two important convergence results in empirical process theory. One is the convergence in law of the uniform empirical process (u.e.p.) to the Brownian bridge. This result is originally due to M. D. Donsker [Don52] who carried out an idea by J. L. Doob [Doo49]. The other one is the following: If G_n is the uniform empirical distribution function (u.e.d.f.), then

$$\beta_n(t) := n\, G_n(\tfrac{t}{n}), \qquad t \geq 0,$$

converges in law to a Poisson process with rate 1, see e.g. [AHE84] or [KLS80]. We want to call the process β_n the *rescaled* (uniform) empirical distribution function (r.u.e.d.f.). Although being nowadays a standard exercise in empirical process theory, the origin of this result has remained, up to this day, unknown to me.
The Brownian bridge, closely linked to the Brownian motion, and the Poisson process are two fundamental stochastic processes, the relevance of which goes far beyond being limit processes in asymptotic statistics. They appear in all kinds of applications of probabiliy theory. The empirical distribution function (e.d.f.) and derived processes (such as the empirical process) are an important fields of study in mathematical statistics, see for example [SW86] or [vdWW96]. These volumes both contain a profound treatment of up-to-date empirical process theory with particular focus on statistical applications.

The aim of this paper is to prove the asymptotic independence of the u.e.p. and the r.u.e.d.f. In order to properly specify the task, some mathematical formulae are necessary. Let $U_1, U_2, ...$ be independent, uniformly on $(0,1)$ distributed random variables, and G the distribution function of U_1. The corresponding empirical distribution function (u.e.d.f.) G_n is given by

$$G_n(t) := \frac{1}{n}\sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq t\}}, \qquad t \in \mathbb{R},\ n \geq 1.$$

We want to call the u.e.p. α_n, it is defined by

$$\begin{aligned}\alpha_n(t) &:= \sqrt{n}(G_n(t) - G(t)) = \sqrt{n}(G_n(t) - t) \\ &= \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\left(\mathbb{1}_{\{U_k \leq t\}} - t\right), \qquad t \in [0,1],\ n \geq 1.\end{aligned}$$

The r.u.e.d.f. β_n, introduced above, can be written as

$$\beta_n(t) := n\, G_n\left(\tfrac{t}{n}\right) = \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq \frac{t}{n}\}}, \qquad t \geq 0,\ n \geq 1.$$

Finally, let B_0 be a Brownian bridge on $[0,1]$ and N a Poisson process with rate 1 on $[0,\infty)$. Then,

(I) $\alpha_n \xrightarrow{\mathscr{L}} B_0$ and

(II) $\beta_n \xrightarrow{\mathscr{L}} N$

are well established. We are going to prove that, if furthermore B_0 and N are stochastically independent of each other, then

(III) $(\alpha_n, \beta_n) \xrightarrow{\mathscr{L}} (B_0, N)$.

The $\mathscr{L}$ stands for "law", and $\xrightarrow{\mathscr{L}}$ denotes convergence *in law* - or convergence *in distribution.*

Before tackling the proof let me say a few words about the implications of (III). It is, in my opinion, quite a remarkable result. The fact that the convergence extends from the individual sequences to the joint sequence is, although not to be taken for granted, hardly surprising. But B_0 and N are independent, while α_n and β_n – *if* derived from the same sequence $\{U_n\}$ – are clearly not. Consider a fixed $t \in [0,1]$. One implication of (III) is that $\alpha_n(t)$ and $\beta_n(t)$ are asymptotically independent. This may seem plausible, since they are deterministic transformations of $G_n(t)$ and $G_n(\frac{t}{n})$, respectively. But (III) states even stronger that the whole processes are asymptotically independent – and that although α_n and β_n are linked via the strongest form of stochastic dependence there is: Knowing one means knowing the other.

When it comes to proving the result, the first question arising is: weak convergence – but in which measurable *spaces*? Apparently these must be different ones for (I) and (III). Customarily one takes a Borel-σ-field, thus it comes down to choosing a topological space, which desirably is metrizable and separable.

Let us have a look at the processes involved. B_0 and α_n live on $[0,1]$. N and β_n have $[0,\infty)$ as their largest "common ground". They could as well be treated on finite time interval, but that would be somewhat unsatisfactory. All process except the Brownian bridge B_0 have discontinuous paths. This strikes out, for example, the nice, separable metric space $(C[0,1], ||\cdot||_\infty)$ – the space of continuous functions on $[0,1]$.

But the trajectories of all processes are right-continuous, and the left-hand limits exist in all points. Such a function is called a *cadlag* function – "continue à droite, limites à gauche". The space of all cadlag functions is usually denoted by $D[0,1]$, or $D[0,\infty)$, depending on the domain of the functions.

An element of $D[0,1]$, or any cadlag function on a compact set, stays bounded, just as a continuous function does. Hence the sup-metric $||\cdot||_\infty$ is a possible metric for $D[0,1]$. It induces the *topology of uniform convergence* or short, the *uniform topology*. However, this metric is unsuitable for $D[0,1]$, due to several reasons. First, $(D[0,1], ||\cdot||_\infty)$ is not separable (see e.g. [JS02], page 325). Second, and more severe, there are measurability problems. The empirical process is not measurable with respect to the uniform topology. In fact, Donsker's original proof of (I) was flawed, because he used this topology. In 1956, A. V. Skorokhod [Sko56] proposed a different, coarser topology, which he called J_1-topology, but which usually is refered to as the *Skorokhod topology*. It is separable, metrizable (the metric is denoted by d in this paper), it solves the measurability issue, and – as some examples in the 3. section demonstrate – it also declares a convergence more natural to functions with jumps.
Nowadays, $D[0,1]$ is by default equipped with this topology and simply refered to as the *Skorokhod space*. The other two important papers on $D[0,1]$ are Kolmogorov [Kol56] and Prokhorov [Pro56]. The analogue on $D[0,\infty)$ is due to C. Stone [Sto63].
Thus, (I) is modeled in the Skorokhod space $D[0,1]$, (II) is modeled in the Skorokhod space $D[0,\infty)$, and (III) in a natural way in the product space $D[0,1]\times D[0,\infty)$. Primarily this means the product of the measurable spaces, but in this case the product of the Borel-σ-fields is the same as the Borel-σ-field on the product topology (cf. Lemma 5.1).

The mathematician always has the generalization in the back of her or his mind. Instead of being uniformly distributed, let the $U_1, U_2, ...$ pertain to an arbitrary distribution function F, and let F_n be the empirical. Then,

$$\alpha_n^F(t) := \sqrt{n}\big(F_n(t) - F(t)\big), \qquad t \in \mathbb{R},$$

and

$$\beta_n^{F,\tau}(t) := n\big(F_n(\tau + \tfrac{t}{n}) - F_n(t)\big), \qquad t \in \mathbb{R},$$

for any fixed $\tau \in \mathbb{R}$, are feasible generalizations of α_n and β_n, respectively. They both have paths in $D(-\infty,\infty)$. These processes also converge in law (to some limit processes yet to be specified), and one conjectures that the limits are independent as well. We want to treat the problem in this general setting.
We will show that, if F is continuous[1] and if some further regularity conditions hold, $(\alpha_n^F, \beta_n^{F,\tau})$ converges in distribution in the Skorokhod product

[1] It seems that the result holds as well if this restriction is removed. See remark on page 76.

space $D(-\infty,\infty) \times D(-\infty,\infty)$. We will specify the limit, show that it is the Cartesian product of two independent processes, and derive (III) as a special case.
This endeavor breaks down into basically two tasks:

(A) Derive a weak convergence criterion in the space $D(-\infty,\infty) \times D(-\infty,\infty)$ and

(B) show that $(\alpha_n^F, \beta_n^{F,\tau})$ satisfies it.

The criterion (A) is Theorem 5.5, (B) is the content of Theorem 5.9.
The standard method of proving weak convergence of stochastic processes is as follows: Prove the weak convergence of the finite-dimensional distributions, and show that the sequence is tight. The key argument here is Prokhorov's theorem [Pro56]. For example, this method is used to show that the partial sum process converges to the Brownian motion in $(C[0,1], ||\cdot||_\infty)$ (Donsker's theorem [Don51]). The principle transfers with little alteration to $(D[0,1], d)$ and $D[0,\infty)$. We will show that it extends as well to $D(-\infty,\infty) \times D(-\infty,\infty)$. It is, however, only feasible, if the finite-dimensional distributions are known, and there are other approaches as well, see e.g. the introduction of [JS02].

This leads us to the content of the thesis. I am going to derive the main result (A) gradually and in a detailed manner. I start by introducing the space $D[0,1]$, develop the basic concept, then move on to $D(-\infty,\infty)$, and finally to $D(-\infty,\infty) \times D(-\infty,\infty)$. It is explained, how each stage builds upon the previous one, and the differences between the spaces and the respective convergence results are pointed out. Along the way we prove (I) and (II) – as a warming up to (III).
The outline of the thesis is as follows: In section 2 (Preliminaries) we gather together important results and definitions. It is for two purposes that these are put at the beginning (and not just as references in the appendix). First, they reveal the basic notions we will deal with, and thus indicate roughly the direction we are headed. Second, it is convenient for the reader to have them recapulated beforehand.
Section 3 is devoted to the space $D[0,1]$. We introduce the Skorokhod metric on $D[0,1]$ (subsection 3.1), and develop a weak convergence criterion that realizes the principle described above, including a moment-type tightness criterion (subsection 3.3), and finally, as an example of its application, show (I) (subsection 3.4).
Section 4 repeats the program for the space $D(-\infty,\infty)$. The Skorokhod topology on $D(-\infty,\infty)$ is defined (subsection 4.1), and an analoguous weak

convergence result is proved (subsection 4.2). In subsection 4.3 we derive the $D(-\infty, \infty)$-version of the tightness criterion from subsection 3.3. A short discussion of other possible tightness criteria is added. In subsection 4.4 we show (II), with β_n and N embedded in $D(-\infty, \infty)$, i.e. they are set to 0 on the negative time axis.
Section 5 then deals with the Skorokhod product space $D(-\infty, \infty) \times D(-\infty, \infty)$. It is explained that this product space can be identified with the space of all cadlag functions $\mathbb{R} \to \mathbb{R}^2$, and how the Skorokhod topology on this space (which is declared analoguously to the one one $D(-\infty, \infty)$) relates to our product topology. Subsection 5.1 contains the weak convergence criterion (Theorem 5.5). In subsection 5.2 we apply this criterion to prove the weak convergence of $(\alpha_n^F, \beta_n^{F,\tau})$ for continuous F (Theorem 5.9). We deduce $(\alpha_n, \beta_n) \xrightarrow{\mathscr{L}} (B_0, N)$, thus solving the initial task. The section ends with a suggestion how to overcome the unsatisfactory restriction of F being continuous.

Some remarks on the literature. The main reference of this paper is Patrick Billingsley's "Convergence of Probability Measures" [Bil99]. This book's first edition dates back to 1969. Although recent results have been added in the 2nd edition, it falls short of covering the entirety of what is known today. Nevertheless it is still the number one reference in this field. Partially due to these historic reasons [Bil99] features a stage-wise development from $C[0, 1]$ to $D[0, 1]$ to $D[0, \infty)$. Other books I have consulted include [Pol84], [EK86], [Whi02] and [JS02]. Most of these volumes consider the space $D[0, \infty)$ right away, or a more general version of it, without paying extra attention to $D[0, 1]$. In particular, [JS02] gives an exhaustive treatment of weak convergence on $D[0, \infty)$, and it certainly qualifies as one of "the industrial-strength treatises now available" (Billingsley in the preface to [Bil99]).
The space $D(-\infty, \infty)$, according to my knowledge, does actually not exist in the literature. This is due to the fact that for most applications of stochastic processes the semi-infinite time interval suffices. While it is quite a substantial step from $D[0, 1]$ to $D[0, \infty)$, $D(-\infty, \infty)$-theory is qualitatively not much different from $D[0, \infty)$-theory. My definition of the Skorokhod metric d_∞ on $D(-\infty, \infty)$ follows Billingsley's [Bil99] construction of a $D[0, \infty)$-metric by means of the Skorokhod metric d on $D[0, 1]$. Billingsley adopts a suggestion of T. Lindvall [Lin73], who in turn follows W. Whitt's approach on $C[0, \infty)$, [Whi70]. Whitt also suggests another metric on $D[0, \infty)$, [Whi71].
Lindvall's paper furthermore contains a weak convergence criterion, which roughly goes like this: A sequence of processes in $D[0, \infty)$ converges if the sequences of every (or at least sufficiently many) finite restriction of the processes do. [ŁR85] provides the analogue for the k-dimensional Skorokhod

space. This tells us that weak convergence on the infinite time interval is basically the same as on finite intervals, but it is otherwise not useful as a starting point for our problem.

Results taken from other sources are always indicated as such. Nevertheless, I would like to say a few extra words about what is my own contribution and what is not. (III) is of course an idea of my supervisor, Professor Dr. D. Ferger. He also indicated the possible extension to a general distribution function.

Section 3 serves as a preface and does only contain well established facts. Most of it can be found in a more compressed form in [Bil99]. Theorem 3.8 is 13.1 in [Bil99]. Example (III) on page 13, the proofs of 3.6 and 3.5 (a) are my ideas. Using the Markov inequality in 3.8 is my supervisor's advice. 3.10 and 3.11 are adopted from [Fer03a].

Although **Section 4** is still primarily a preface, according to my knowledge, none of its content is contained elsewhere in this form. As mentioned above, I did not find the space $D(-\infty, \infty)$ in the literature. The definition of the metric d_∞ is my own. I took a lot of care re-working any results from $D[0, \infty)$, and did not just "assume" their validity in $D(-\infty, \infty)$. In particular, formulation and proof of Theorem 4.9 are my work. Also, compare this Theorem to Theorem 4.1 on page 355 in [JS02]. I did 4.18 and 4.19 by myself.

Section 5. The basic idea of 5.5 is due to my supervisor. The proof, as it is here, is a slightly changed version. I added the formulation 5.8. All examples in section 5 are ideas of mine. The two essential tools in the proof of Proposition 5.10 are ideas of my supervisor, i.e. using (80) and the Taylor expansion of the exponential function – and of course using characteristic functions. My supervisor pointed out to me the limit of $\beta_n^{F,\tau}$, see also [Fer03b]. Apart from that, the generalization from (α_n, β_n) to $(\alpha_n^F, \beta_n^{F,\tau})$ is my work.

Appendix. Lemmas A.1 (III), A.3, A.4 and A.5 are my ideas, Lemma A.7 uses ideas from the proof of (12.33) in [Bil99].

The thesis contains six illustrations, which were created using MS Excel.

2 Preliminaries

This section gathers together basic notions and results we will use.

Two Remarks on Notation.

(I) $X_n \xrightarrow{\mathscr{L}} X$ denotes the convergence in distribution of X_n to X, which is equivalent to the weak convergence of the corresponding distributions. I thereby justify the notation $P_n \xrightarrow{\mathscr{L}} P$ for measures P_n and P, meaning that P_n weakly converges to P.

(II) Whenever the members of a finite set M are declared by writing $M = \{x_1, \ldots, x_n\}$, and M is a subset of an ordered space, usually $\mathbb{N}$ or $\mathbb{R}$, then $x_1, \ldots, x_n$ are supposed to be ordered according to their indices.

2.1 Some Measure Theory

Definition 2.1 (π-system) A system of sets is called a π-*system*, if it is closed under the formation of finite intersections. (See also [Bil99], page 9.)

Definition 2.2 (Separating class) Let $(\Omega, \mathscr{A})$ be a measurable space. $\mathscr{S} \subset \mathscr{A}$ is called a *separating class* for $\mathscr{A}$, if any two probability measures that agree on $\mathscr{S}$ also agree on the whole of $\mathscr{A}$. (See also [Bil99], page 9.)

Proposition 2.3 (Uniqueness theorem for probability measures) *If $\mathscr{S}$ is a π-system, generating the σ-field $\mathscr{A}$, then $\mathscr{S}$ is a separating class for $\mathscr{A}$.*

Proof. See e.g. [Bil95], Theorem 3.3, page 42. □

Proposition 2.4 *Let $(\Omega, \mathscr{A})$ be an arbitrary measurable space. Furthermore let S be a metric space, and $\mathscr{B}(S)$ its Borel-σ-field. If h_n and h are functions from Ω to S with $h_n \longrightarrow h$ pointwise, and h_n is $(\mathscr{A}, \mathscr{B}(S))$-measurable for all $n \in \mathbb{N}$, then h is $(\mathscr{A}, \mathscr{B}(S))$-measurable, too.*

Proof. See [Bil99], Appendix [M10], page 243. □

Proposition 2.5 *Let $(h_i)_{i\in I}$ be a family of functions h_i from a set S into measurable spaces $(\Omega_i, \mathscr{A}_i)$. Furthermore, let h be a function from a measurable space $(\Omega, \mathscr{A})$ into S. The function h is $(\mathscr{A}, \sigma\{h_i|i \in I\})$-measurable if and only if each mapping $h_i \circ h$ is $(\mathscr{A}, \mathscr{A}_i)$-measurable.*

Proof. See [Bau92], Theorem 7.4, page 42.

$\square$

Proposition 2.6 *Let $(\Omega_1, \mathscr{A}_1)$, $(\Omega_2, \mathscr{A}_2)$ be measurable spaces and $\mathscr{S}$ a generating class for $\mathscr{A}_2$. The mapping $h : \Omega_1 \to \Omega_2$ is measurable if and only if $h^{-1}(S) \in \mathscr{A}_1$ for all $S \in \mathscr{S}$.*

Proof. See [Bau92], Theorem 7.2, page 41.

$\square$

Proposition 2.7 *For $i = 1, ..., n$ let $(\Omega_i, \mathscr{A}_i)$ be a measurable space and $\mathscr{F}_i$ a generating class of the σ-field $\mathscr{A}_i$. Suppose each $\mathscr{F}_i$ contains a sequence of sets $\{F_k^{(i)}\}$ such that $F_k^{(i)} \uparrow \Omega_i$ (i.e. $F_1^{(i)} \subset F_2^{(i)} \subset ...$ and $\bigcup F_n^{(i)} = \Omega_i$). Then*

$$\mathscr{F}_1 \times ... \times \mathscr{F}_n = \{F_1 \times ... \times F_n | F_i \in \mathscr{F}_i, i = 1, ..., n\}$$

generates the product σ-field. $\mathscr{A}_1 \otimes ... \otimes \mathscr{A}_n$.

Proof. See [Bau92], Theorem 22.1, page 151.

$\square$

Remark. The assumptions are met if each $\mathscr{F}_i$ contains Ω_i, $i = 1, ..., n$.

Proposition 2.8 *Let T_k, $k \in \mathbb{N}$, be separable topological spaces. Then*

$$\mathscr{B}(\prod_{k\in\mathbb{N}} T_k) = \bigotimes_{k\in\mathbb{N}} \mathscr{B}(T_k),$$

or in words, the Borel-σ-field of the product of the T_k, $k \in \mathbb{N}$, is the product σ-field of the Borel-σ-fields of each of the T_k.

Proof. See [Els02], Theorem 5.10, page 115.

$\blacksquare$

2.2 Prokhorov's Theorem

Definition 2.9 (Tightness of a measure) Let S be a metric space. A probability measure P on $(S, \mathscr{B}(S))$ is *tight*, if for every $\varepsilon > 0$ there exists a compact set K such that $P(K) > 1 - \varepsilon$.

Proposition 2.10 *If S is separable and complete, then every probability measure on*
$(S, \mathscr{B}(S))$ is tight.

Proof. see [Bil99], page 8, Theorem 1.3. $\square$

Definition 2.11 (Tightness of a sequence) Let S be a metric space. A family $\mathscr{P}$ of probability measures on $(S, \mathscr{B}(S))$ is *tight*, if for every $\varepsilon > 0$ there exists a compact set K such that $P(K) > 1 - \varepsilon$ for every $P \in \mathscr{P}$. A family of random variables is *tight* if the family of their respective distributions is tight. (See e.g. [Bil99], page 59.)

Definition 2.12 (Relative compactness) Let S be a metric space. A family $\mathscr{P}$ of probability measures on $(S, \mathscr{B}(S))$ is *relatively compact*, if every sequence in $\mathscr{P}$ contains a weakly convergent subsequence. (See e.g. [Bil99], page 57.)

Remark. The annex "relative" refers to the fact that the limiting measure of such a converging subsequence does not need to be an element of $\mathscr{P}$.

Corollary 2.13 *A weakly converging sequence $\{P_n\}$ of probability measures on $(S, \mathscr{B}(S))$ (with S as above) is relatively compact.*

Proof. Immediate corollary of the definition of relative compactness. $\square$

Theorem 2.14 (Prokhorov) *Let S be a metric space and $\mathscr{P}$ a family of probability measures on $(S, \mathscr{B}(S))$. Then*

(1) If $\mathscr{P}$ is tight, then it is relatively compact.

(2) If $\mathscr{P}$ is relatively compact, and if S is complete and separable, then $\mathscr{P}$ is tight.

This means, for a family of probability measures on the Borel-σ-field of a separable, complete metric space, tightness and relative compactness are equivalent.

In this paper we will only make use of (1).

2.3 Some Probability Theory

Theorem 2.15 (Continuous mapping theorem) *Let S_1 and S_2 be two metric spaces, and X, X_n, $n \in \mathbb{N}$, random variables in $(S_1, \mathscr{B}(S_1))$ with $X_n \xrightarrow{\mathscr{L}} X$. If $h : S_1 \to S_2$ is $(\mathscr{B}(S_1), \mathscr{B}(S_2))$-measurable and $\mathscr{L}(X)$-a.e. continuous, then $h(X_n) \xrightarrow{\mathscr{L}} h(X)$.*

Proof. See for example [Fer02b], Theorem 5.3 or [Wel02], Proposition 3.4, Remark 3.1 or [Bil99], Theorem 2.7, page 21.

□

Theorem 2.16 (Uniqueness of characteristic functions) *Let X be a random variable in $\mathbb{R}^k$. The distribution of X is uniquely determined by its characteristic function $\varphi_X : \mathbb{R}^k \to \mathbb{C}$:*

$$\varphi_X(t) = \mathbb{E}e^{i\langle t,X\rangle}, \qquad t \in \mathbb{R}^k.$$

Proof. Cf. e.g. [Bil95], page 382.

□

Theorem 2.17 *Let X, X_n, $n \in \mathbb{N}$, be random variables in $\mathbb{R}^k$, and φ, φ_n their respective characteristic functions. X_n converges in law to X if and only if $\varphi_n(t) \to \varphi(t)$, $t \in \mathbb{R}^k$, pointwise.*

Proof. See e.g. [Bil95], Theorems 26.3, page 349, and 29.4, page 383.

□

3 Weak Convergence in $D[0,1]$

This section is altogether of purely introductory nature and only contains well established facts. It explains the main principles and lays the foundations for all later considerations. The content is mostly taken from [Bil99], sections 12, 13 and 14.
We introduce the Skorokhod space $D[0,1]$ and establish a criterion for weak convergence in this space, Theorem 3.9. In the last subsection 3.4 we show, as example of the application of Theorem 3.9, that the uniform empirical process converges in law to the Brownian bridge.

3.1 The Skorokhod Space $D[0,1]$

Definition 3.1 Let $D = D[0,1]$ be the space of all real functions x on $[0,1]$ that are right-continuous and have left-hand limits, that is

(i) for $0 < t \leq 1$, $x(t-)$ exists, and

(ii) for $0 \leq t < 1$, $x(t+)$ exists and equals $x(t)$.

Functions havings these properties are called *cadlag* functions ("continue à droite, limites à gauche").
The elements of D have even nicer properties than above's definition claims. A function $x \in D$ has at most countably many discontinuity points and is bounded in the $\|\cdot\|_\infty$-metric (cf. [Bil99], page 122).
D is usually equipped with a different than the uniform metric, the *Skorokhod metric*, which is defined as follows. Let Λ denote the class of all homeomorphisms $\lambda : [0,1] \to [0,1]$ with $\lambda(0) = 0$ and $\lambda(1) = 1$. Recall that a homeomorphism is an invertible, continuous function whose is inverse is also continuous. An equivalent way of characterizing the elements of Λ is to say, λ is a strictly increasing, continuous mapping from $[0,1]$ onto itself. Note here the small significant word *onto*, stating that λ is surjective.
Now, for $x, y \in D$,

$$\begin{aligned} d(x,y) &:= \inf_{\lambda\in\Lambda}\{\|\lambda - \mathrm{id}\|_\infty \vee \|x \circ \lambda - y\|_\infty\} \qquad (1) \\ &= \inf_{\lambda\in\Lambda}\{\sup_{t\in[0,1]} |\lambda(t) - t| \vee \sup_{t\in[0,1]} |x(\lambda(t)) - y(t)|\}. \end{aligned}$$

Although it might not look like it at first glance, it is actually not very hard to get convinced that d is a metric. Symmetry and the triangle inequality follow pretty straight forward from the group properties of Λ. For example, note that, if λ is an element of Λ, so is λ^{-1}, and it holds

$$\sup_{t\in[0,1]} |\, x(t) - y(\lambda(t))| = \sup_{s\in[0,1]} |\, x(\lambda^{-1}(s)) - y(s)|. \tag{2}$$

This provides the already the symmetry. We are not going to prove that d is metric, but refer to [Bil99], page 124.
(D, d) is a separable metric space (cf. [Bil99], page 128). However, it is not complete (for an example see [Bil99], Example 12.2), but this is of no relevance to us. First of all it is "fixable", in the sense that one can declare a complete metric equivalent to d. Second, we will not need completeness, d - as defined above - is fully sufficient for our purposes.
The topology induced on D by the metric d is called *Skorokhod topology.* The term "Skorokhod metric" then specifies a metric only up to the fact, that it induces the Skorokhod topology. Whenever we use it, we mean d. Since the identy on $[0, 1]$ is an element of Λ it is clear that d is always smaller than or equal to the sup-metric. Hence the Skorokhod topology is coarser than the topology of uniform convergence, i.e. more sequences converge. In order to give an idea about convergence in the Skorokhod topology, consider the following three examples. All functions are defined on $[0, 1]$, and $\mathbb{0}$ denotes the function identical zero.

	Convergence with respect to	$\|\| \cdot - \cdot \|\|_\infty$	d	L_2-metric
(I)	$(1 - \frac{1}{n})\mathbb{1}_{[0,\frac{1}{2})} \to \mathbb{1}_{[0,\frac{1}{2})}$	yes	yes	yes
(II)	$\mathbb{1}_{[0,\frac{1}{2}-\frac{1}{n})} \to \mathbb{1}_{[0,\frac{1}{2})}$	no	yes	yes
(III)	$\mathbb{1}_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2})} \to \mathbb{0}$	no	no	yes

The two not-so-obvious claims, i.e. middle column, last two lines, are treated in Lemmas A.2 and A.3 in the appendix. It should be mentioned that the Skorokhod topology on $C[0, 1]$, i.e. the topology induced by d restricted to $C[0, 1]$, coincides with the uniform convergence topology there. (cf. [Bil99], page 124). Keep in mind, that the Skorokhod space D is "completely metrizable", i.e metrizable with a complete metric.
Let $\mathscr{D}$ be the Borel-σ-field in (D, d). If we talk about D we mean the measurable space $(D, \mathscr{D})$, unless it is stated otherwise.

3.2 Some Basic Tools

We introduce two notions that will repeatedly appear throughout the paper: the projection and the grid.

For any $t \in [0,1]$ the projection $\pi_t : D \to \mathbb{R}$ assigns to $x \in D$ its value at the point t, i.e.

$$\pi_t(x) := x(t). \tag{3}$$

Throughout this paper, $T = \{t_1, ..., t_k\}$ shall denote a finite subset of $[0,1]$, having k elements $0 \leq t_1 < ... < t_k \leq 1$. Then define

$$\pi_T : D \longrightarrow \mathbb{R}^k : x \mapsto (x(t_1), ..., x(t_k)) = (\pi_{t_1}(x), ..., \pi_{t_k}(x)). \tag{4}$$

Lemma 3.2

(1) π_0, π_1 *are continuous.*

(2) If $0 < t < 1$, *then* π_t *is continuous at* x *if and only if* x *is continuous at* t.

Proof. See [Bil99], pages 133 and 134.

□

Suppose now that $\sigma = \{s_0, ..., s_k\}$ satisfies $0 = s_0 < s_1 < ... < s_k = 1$. We call such a set a *grid on* $[0,1]$, or simply a *grid*. The *mesh* of a grid is given by

$$\delta(\sigma) := \max_{i=1,...,k} \{s_i - s_{i-1}\}.$$

We define a function $A_\sigma : D \to D$:

$$A_\sigma x(t) := \begin{cases} x(s_{i-1}), & t \in [s_{i-1}, s_i), \quad (i = 1, ..., k), \\ x(1), & t = 1. \end{cases}$$

$A_\sigma x$ is piecewise constant, takes the value $x(s_{i-1})$ on the intervall $[s_{i-1}, s_i)$, and agrees with x at $t = 1$.

Lemma 3.3 *Let* $\{\sigma_m\}$ *be a sequence of grids with* $\delta(\sigma_m) \to 0$ *as* $m \to \infty$. *(This of course implies* $|\sigma_m| \to \infty$.*) Then*

$$d(A_{\sigma_m} x, x) \longrightarrow 0. \tag{5}$$

Proof. See [Bil99], Lemma 3, page 127.

□

3.3 A Criterion for Weak Convergence in $D[0,1]$

In $C[0,1]$, the space of all real continuous functions on $[0,1]$ (equipped with the sup-metric), we know that a sequence of random variables converges if it is tight and all finite-dimensional distributions converge. For the proof of this result, it is essential that the projections π_t are continuous. It is an application of the Continuous mapping theorem, for which it suffices that the function is almost everywhere continuous, but the projections π_t are even continuous on $C[0,1]$.
Now we want to carry over this principle to $D[0,1]$, but here we face the dilemma that the π_t are in general not continuous. But with the elements of D being quite well behaved (at most countably many jumps), there are still sufficiently many $t \in [0,1]$ for which π_t is almost everywhere continuous. Of this fact we convince us first (Proposition 3.4), and then show that this indeed suffices (Proposition 3.7).
Consider a probability measure P on $(D, \mathscr{D})$. By T_P we denote the set of all $t \in [0,1]$, for which the projection π_t is P-almost everywhere continuous, i.e.

$$T_P := \big\{ t \in [0,1] \big| P(\{x \in D | \pi_t \text{ is continuous at } x\}) = 1 \big\}. \tag{6}$$

The set T_P has a convenient property.

Proposition 3.4 *T_P contains 0 and 1, and its complement in $[0,1]$ is at most countable.*

Proof. The proof makes use of Lemma 3.2, cf. [Bil99], page 138. □

This already completes the first step towards our end, to prove Theorem 3.8. We know that the set T_P is in a certain sense good-natured. We show next that this is actually good enough. For any subset T_0 of $[0,1]$ define

$$\mathscr{F}(T_0) := \big\{ \pi_T^{-1}(A) | \, T \subset T_0 \text{ finite}, \, A \in \mathscr{B}(\mathbb{R}^{|T|}) \big\}. \tag{7}$$

We prove in Proposition 3.7 that, if T_0 is dense in $[0,1]$ and contains 0 and 1, $\mathscr{F}(T_0)$ is a separating class for $\mathscr{D}$. This will be done by applying Proposition 2.3, hence we need to show that $\mathscr{F}(T_0)$ generates $\mathscr{D}$ (Proposition 3.5), and that it is a π-system (Lemma 3.6).

Proposition 3.5 *Let T_0 be a dense subset of $[0,1]$, containing 0 and 1. Then*

$$\sigma\{\pi_t | t \in T_0\} = \sigma(\mathscr{F}(T_0)) = \mathscr{D}. \tag{8}$$

Proof. The proof consists of three parts.

(a) $\sigma(\mathscr{F}(T_0)) \subset \sigma\{\pi_t | t \in T_0\}$

(b) $\sigma\{\pi_t | t \in T_0\} \subset \mathscr{D}$

(c) $\mathscr{D} \subset \sigma(\mathscr{F}(T_0))$

That $\sigma(\mathscr{F}(T_0))$ equals $\sigma\{\pi_t | t \in T_0\}$ is relatively easy to see. One could as well consider it as a known fact and skip (a). But since (b) and (c) are much easier to prove directly than the respective inverse inclusion, we close the circle by giving a reminder how (a) can be proved.

Proof of (a). Write $\sigma(\mathscr{F}(T_0))$ as $\sigma\{\pi_T | T \subset T_0 \text{ finite}\}$ and note that $\sigma\{\pi_T | T \subset T_0 \text{ finite}\} \subset \sigma\{\pi_t | t \in T_0\}$ is equivalent to saying π_T is $(\sigma\{\pi_t | t \in T_0\}, \mathscr{B}(\mathbb{R}^k))$-measurable. The latter follows by a suited application of Proposition 2.5. Using the notation from Proposition 2.5, take $I = \{1, ..., k\}$, $S = \mathbb{R}^k$, $(\Omega_i, \mathscr{A}_i) = (\mathbb{R}, \mathscr{B}(\mathbb{R}))$, $h_i = \pi_i : \mathbb{R}^k \to \mathbb{R} : (y_1, ..., y_k) \mapsto y_i$, $i = 1, ..., k$, furthermore $(\Omega, \mathscr{A}) = (D, \sigma\{\pi_t | t \in T_0\})$ and $h = \pi_T : D \to \mathbb{R}^k$. Keep in mind that $\mathscr{B}(\mathbb{R}^k) = \sigma\{\pi_i | i = 1, ..., k\}$.

Proof of (b). We show that π_t is $(\mathscr{D}, \mathscr{B}(\mathbb{R}))$-measurable for any $t \in [0,1]$, and hence for all $t \in T_0$. From Lemma 3.2 we know that π_0 and π_1 are continuous, so they are measurable. Recall that $\mathscr{D}$ is the Borel-σ-field on D. Now let $0 < t < 1$. We apply Proposition 2.4, i.e. we specify a sequence of measurable functions converging pointwise to π_t. The following one will do:

$$h_m(x) := m \int_t^{t+\frac{1}{m}} x(s)\, ds$$

Because of the right-continuity of x, $h_m(x) \to \pi_t(s)$ as $m \to \infty$. It remains to show that the h_m are indeed measurable. We do so by proving their continuity. Let $x_n \to x$ in (D, d). Then $x_n(s) \to x(s)$ for all continuity points s of x, i.e. for λ^1-almost all points in $[0,1]$. Furthermore, (x_n) is bounded, since it converges. So by Lebegue's Dominated convergence theorem, $h_m(x_n) \to h_m(x)$ as $n \to \infty$. Hence, h_m is continuous for all $m \in \mathbb{N}$.

Proof of (c). We show, the identy map id : $(D, \sigma(\mathscr{F}(T_0))) \to (D, \mathscr{D})$ is measurable. Again, we do so by applying Proposition 2.4. For $m \in \mathbb{N}$ choose a grid $\sigma_m = \{s_0, ..., s_{k(m)}\}$, where $\delta(\sigma_m) < \frac{1}{m}$ and all $s_0, ..., s_{k(m)}$ are in T_0. This is possible due to the density of T_0 and the fact, that it contains 0 and 1. The mapping A_{σ_m} defined in the previous subsection is $(\sigma(\mathscr{F}(T_0)), \mathscr{D})$-measurable. This can be seen as follows. For $y = (y_0, ..., y_{k(m)})$ in $\mathbb{R}^{k(m)+1}$, let $V_m y$ be the element in D taking constant value

y_{i-1} on $[s_{i-1}, s_i)$, $1 \le i \le k(m)$, and taking the value $y_{k(m)}$ at $t = 1$. Clearly, $V_m : \mathbb{R}^{k(m)+1} \to D$ is continuous (relative to the Euclidean and Skorokhod topologies), and hence is measurable with respect to $(\mathscr{B}(\mathbb{R}^{k(m)+1}), \mathscr{D})$. Since π_{σ_m} is $(\sigma(\mathscr{F}(T_0)), \mathscr{B}(\mathbb{R}^{k(m)+1}))$-measurable, $A_{\sigma_m} = V_{\sigma_m} \circ \pi_{\sigma_m}$ is measurable as a composition of measurable mappings. Finally, $\lim_{m\to\infty} A_{\sigma_m} x = x$ by Lemma 3.3. By Proposition 2.4 the identity map on D is $(\sigma(\mathscr{F}(T_0))), \mathscr{D})$-measurable.
∎

Remark. For (8) to hold, it is not necessary that T_0 contains 0, since the elements of D are right-continuous at 0. This merely simplifies the proof. All sets T_0 that will be of interest to us, do contain 0. However, (8) is wrong, if T_0 does not contain 1.

Lemma 3.6 *For an arbitrary subset T_0 of $[0, 1]$, $\mathscr{F}(T_0)$ is a π-system.*

Proof. We have to show, that for two arbitrary elements of $\mathscr{F}(T_0)$, say $\pi_T^{-1}(A)$ and $\pi_S^{-1}(B)$, with $T = \{t_1, \dots, t_k\}$, $A \in \mathscr{B}(\mathbb{R}^k)$ and $S = \{s_1, \dots, s_l\}$, $B \in \mathscr{B}(\mathbb{R}^l)$, there exist a natural number m, a set $U = \{u_1, \dots, u_m\} \subset T_0$ and a measurable set $C \subset \mathbb{R}^m$, such that

$$\pi_T^{-1}(A) \cap \pi_S^{-1}(B) = \pi_U^{-1}(C)$$

The natural choices of U and C must be $U := T \cup S$, and C the intersection of the appropriate projections of A and B into $\mathbb{R}^m$. When we want to write this down properly, finding a good notation is tricky, since we need to assure that we "project into the right coordinates". I suggest the following: $M := \{1, \dots, m\}$ is the index set of U. Then let $I_T = \{i_1, \dots, i_k\} \subset M$ be the index set of those points in U which are also in T, i.e. i_ν is defined by $u_{i_\nu} = t_\nu$ for $\nu = 1, \dots, k$. Likewise, we define $I_S = \{j_1, ..., j_l\}$ by the relation $u_{j_\nu} = s_\nu$ for $\nu = 1, \dots, l$. Then

$$A^{(m)} := \Big\{(y_1, ..., y_m) \in \mathbb{R}^m \Big| (y_{i_1}, ..., y_{i_k}) \in A,\ y_i \in \mathbb{R}\ \forall i \in M \backslash I_T\Big\},$$

and

$$B^{(m)} := \Big\{(y_1, ..., y_m) \in \mathbb{R}^m \Big| (y_{j_1}, ..., y_{j_l}) \in B,\ y_i \in \mathbb{R}\ \forall i \in M \backslash I_S\Big\}.$$

Take $C := A^{(m)} \cap B^{(m)}$. Clearly, $A^{(m)}$ and $B^{(m)}$ are in $\mathscr{B}(\mathbb{R}^m)$, so is their intersection C, and it holds

$$\pi_U^{-1}(C) = \Big\{x \in D \Big| \big(x(u_{i_1}), ..., x(u_{i_k})\big) \in A, \big(x(u_{j_1}), ..., x(u_{j_l})\big) \in B\Big\}$$
$$= \pi_T^{-1}(A) \cap \pi_S^{-1}(B).$$

■

Proposition 3.7 *If T_0 is dense in $[0,1]$ and contains 0 and 1, then $\mathscr{F}(T_0)$ is a separating class for $\mathscr{D}$.*

Proof. Propositions 2.3, 3.5 and Lemma 3.6 instantly yield the claim. ■

We are now ready to formulate the main result of this section.

Theorem 3.8 *Let $\{P_n\}$ be a sequence of probability measures on $(D,\mathscr{D})$ with the following two properties:*

(1) $\{P_n\}$ is tight.

(2) There exists a measure P on $(D,\mathscr{D})$ such that

$$P_n \circ \pi_T^{-1} \xrightarrow{\mathscr{L}} P \circ \pi_T^{-1}, \qquad \textit{for all finite } T \subset T_P. \tag{9}$$

Then $P_n \xrightarrow{\mathscr{L}} P$ in $(D,\mathscr{D})$.

Proof. By Prokhorov's theorem (2.14) we know that $\{P_n\}$ is relatively compact, i.e. there exist a subsequence $\{P'_n\}$ and a probability measure Q on $(D,\mathscr{D})$, such that

$$P'_n \xrightarrow{\mathscr{L}} Q.$$

For a finite set $S = \{s_1,...,s_l\} \subset [0,1]$ the function $\pi_S = (\pi_{s_1},...,\pi_{s_l})$ is continuous at $x \in D$ if and only if $\pi_{s_1},...,\pi_{s_l}$ are continuous at x. By the σ-additivity it follows that π_S is Q-a.e. continuous, if each component is, i.e. if all the $s_1,...,s_l$ lie in T_Q. Hence, by the CMT (Theorem 2.15):

$$P'_n \circ \pi_S^{-1} \xrightarrow{\mathscr{L}} Q \circ \pi_S^{-1}, \qquad \text{for all finite } S \in T_Q. \tag{10}$$

On the other hand we know that

$$P'_n \circ \pi_T^{-1} \xrightarrow{\mathscr{L}} P \circ \pi_T^{-1}, \qquad \text{for all finite } T \in T_P, \tag{11}$$

since a subsequence of a (weakly) converging sequence necessarily has the same limit. From (10) and (11) follows

$$P \circ \pi_U^{-1} = Q \circ \pi_U^{-1}, \qquad \text{for all finite } U \in T_P \cap T_Q. \tag{12}$$

Let $T_0 = T_P \cap T_Q$ and $\mathscr{F}(T_0)$ be defined as in Proposition 3.5. Then (12) states that P and Q agree on $\mathscr{F}(T_0)$. All we still need to show is that $\mathscr{F}(T_0)$ is a separating class for $\mathscr{D}$, i.e. checking the conditions of Proposition 3.7. By Lemma 3.4 T_P and T_Q have both at most countable complements in $[0,1]$ and both contain 0 and 1. Hence the same is true for their intersection T_0. Since each ball in $[0,1]$ is uncountable, T_0 is dense in $[0,1]$. By Proposition 3.7: $\mathscr{F}(T_0)$ is a separating class for $\mathscr{D}$.

■

In order to really turn Theorem 3.8 into a useful tool for proving weak convergence in $(D, \mathscr{D})$, we still need a handy criterion for tightness. The next theorem is formulated using random variables rather than probability measures, but note that this is a change in notation only. Notably, it allows us to write down expectations.
We write $\{X_n\}$ for a sequence of random elements in $(D, \mathscr{D})$. These may or may not be defined upon common background probability space. We assume here - for reasons of simplicity - that all random elements are defined on the same probability space $(\Omega, \mathscr{A}, \mathbb{P})$. $X_n(t)$, $t \in [0,1]$ then denotes the real random variable $\pi_t \circ X_n$. (Recall that we have proved measurability of π_t.) Finally, $T_X := T_{P_X}$, where P_X is the distribution of X.

Theorem 3.9 *Let $X, X_1, X_2, \dots$ be random variables in $(D, \mathscr{D})$. Suppose that*

(1) $\big(X_n(t_1), ..., X_n(t_k)\big) \xrightarrow{\mathscr{L}} \big(X(t_1), ..., X(t_k)\big)$

as $n \to \infty$ for points $0 \le t_1 < t_2 < ... < t_k \le 1$ in T_X,

(2) there exist a non-decreasing, continuous function F on $[0,1]$ and real numbers $a > 1$ and $b \ge 0$, such that for all $0 \le r < s < t \le 1$ and $n \ge 1$ holds

$$\mathbb{E}\Big(|X_n(s) - X_n(r)|^b |X_n(t) - X_n(s)|^b\Big) \le \big(F(t) - F(r)\big)^a,$$

and

(3) $X(1-) = X(1)$ *a.s.*

Then $X_n \xrightarrow{\mathscr{L}} X$ in $(D, \mathscr{D})$.

Remarks.

(I) Condition (1) is another way of saying $\pi_T \circ X_n \xrightarrow{\mathscr{L}} \pi_T \circ X$, where $T = \{t_1, ..., t_k\}$, and is of course equivalent to (9), with P and P_n being the distributions of X and X_n, respectively.

(II) Because of Theorem 3.8, all that needs to be proved is tightness of $\{X_n\}$.

Proof of Theorem 3.9. This is essentially the same as Theorem 13.5 on page 142 in [Bil99], except for two differences:

(a) Assumption (3) is exchanged for: $X(1) - X(1-\delta) \xrightarrow{p} 0$ as $\delta \to 0$, and

(b) instead of assumption (2) we have: For all $0 \leq r < s < t \leq 1$, $n \geq 1$ and $\lambda > 0$ shall hold

$$\mathbb{P}\Big(|X_n(s) - X_n(r)| \wedge |X_n(t) - X_n(s)| \geq \lambda\Big) \leq \frac{1}{\lambda^{2b}}\big(F(t) - F(r)\big)^a, \quad (13)$$

with α, β and F as above.

Since almost sure convergence implies convergence in probability, (3) follows from (a). So all we do here is show that 3.9 (2) entails (13). This is a simple corollary of Markov's inequality:

$$\mathbb{P}(|Y| \geq \tau) \leq \frac{\mathbb{E}|Y|^k}{\tau^k}, \quad (14)$$

for all $\tau > 0$, $k > 0$ and random variables Y, for which $\mathbb{E}|Y|^k$ exists. Take $\tau = \lambda^2$, $k = b$ and $Y = (X_n(s) - X_n(r))(X_n(t) - X_n(s))$, then (14) turns into

$$\mathbb{P}\big(|X_n(s) - X_n(r)||X_n(t) - X_n(s)| \geq \lambda^2\big) \leq \frac{1}{\lambda^{2b}}\mathbb{E}\big(|X_n(s) - X_n(r)|^b|X_n(t) - X_n(s)|^b\big).$$

Note that

$$\begin{aligned}
&|X_n(s) - X_n(r)||X_n(t) - X_n(s)| \geq \lambda^2 \\
\iff &|X_n(s) - X_n(r)| \geq \lambda,\ |X_n(t) - X_n(s)| \geq \lambda \\
\iff &|X_n(s) - X_n(r)| \wedge |X_n(t) - X_n(s)| \geq \lambda,
\end{aligned}$$

and hence

$$\begin{aligned}
&\mathbb{P}\big(|X_n(s) - X_n(r)| \wedge |X_n(t) - X_n(s)| \geq \lambda\big) \\
&\leq \frac{1}{\lambda^{2b}}\mathbb{E}\big(|X_n(s) - X_n(r)|^b|X_n(t) - X_n(s)|^b\big) \\
&\overset{(2)}{\leq} \frac{1}{\lambda^{2b}}\big(F(t) - F(r)\big)^a.
\end{aligned}$$

■

3.4 The Uniform Empirical Process and the Brownian Bridge

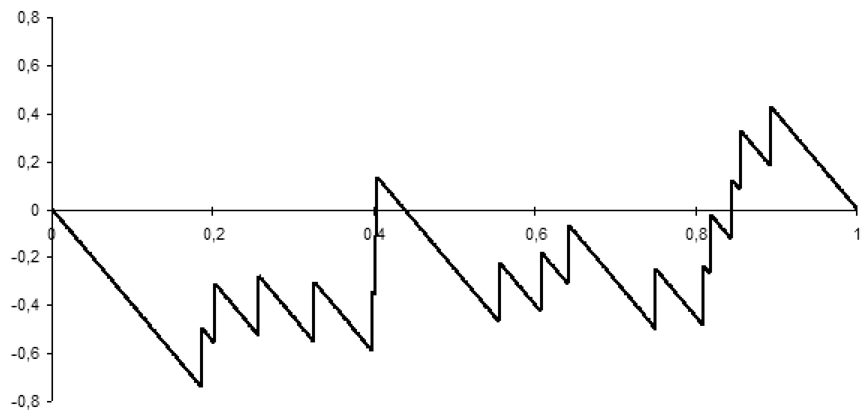
Now we use Theorem 3.9 to show that α_n from the introduction converges in law to the Brownian bridge B_0. As mentioned before, our focus is not on the space $D[0,1]$, and the second section is a preface. So the following serves in the first place as an example of the application of Theorem 3.9. It points in the direction we are heading, though. The sequence $\{\alpha_n\}$ will appear again in the essential forth section, but in a way more generalised guise and embedded in $D(-\infty, \infty)$. We will then refer to Lemma 3.11.
Let $U_1, U_2, ...$ be a sequence of independent, identically uniformly on $[0,1]$ distributed random variables. The empirical distribution function of $U_1, ..., U_n$ is

$$G_n(t) := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{U_k \le t\}}, \qquad t \in [0,1]. \tag{15}$$

Let

$$\begin{aligned} \alpha_n(t) &:= \sqrt{n}(G_n(t) - t) \\ &= \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \left(\mathbb{1}_{\{U_k \le t\}} - t\right), \qquad t \in [0,1]. \end{aligned} \tag{16}$$

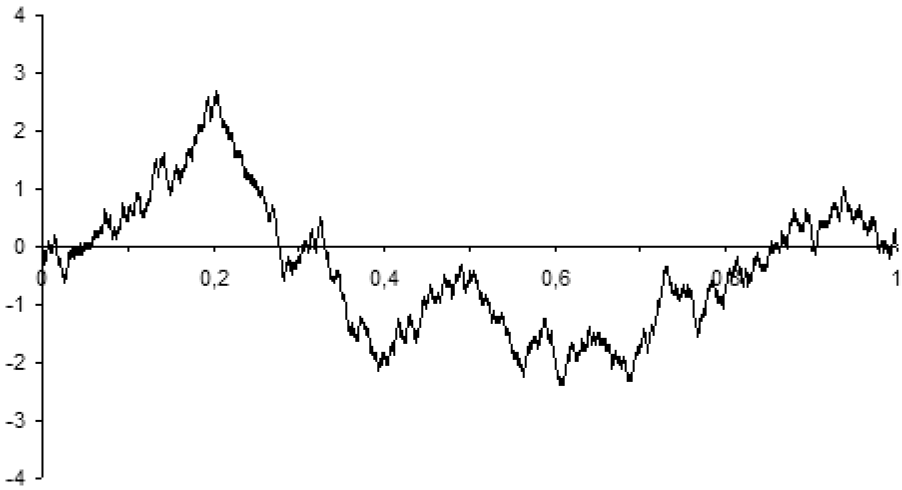
$\alpha_n = \{\alpha_n(t) |\, t \in [0,1]\}$ is a random element in $(D, \mathscr{D})$ and is called the uniform empirical process. Below is a realisiation of α_{16}.



Let furthermore $B_0 = \{B_0(t) |\, t \in [0,1]\}$ a Brownian bridge, i. e.

$$B_0(t) = B(t) - tB(1), \qquad t \in [0,1], \tag{17}$$

where B is a Brownian motion. A typical realisation of a Brownian bridge is shown below.



It holds:

$$\alpha_n \xrightarrow{\mathscr{L}} B_0 \qquad \text{in } (D, \mathscr{D}). \tag{18}$$

This result was first published by Donsker (cf. [Don52]). In the remainder of this section we will prove it by the means provided above, i.e. we have to check the conditions of Theorem 3.9. First, we show that the finite-dimensional distributions of α_n converge to those of B_0 (Proposition 3.10) and then prove in Proposition 3.11 that $\{\alpha_n\}$ satisfies 3.9 (2).

Lemma 3.10 *Let* $0 \le t_1 < ... < t_m \le 1$. *Then*

$$\big(\alpha_n(t_1), ..., \alpha_n(t_m)\big) \xrightarrow{\mathscr{L}} \big(B_0(t_1), ..., B_0(t_m)\big). \tag{19}$$

Proof. Let $\Gamma \in \mathbb{R}^{m\times m}$ be the matrix whose elements in the i-th row and j-th column is $t_i \wedge t_j - t_i t_j$ $(i, j = 1, ..., m)$. It is known that for $0 \le s < t \le 1$ holds

$$\operatorname{cov}(B_0(s), B_0(t)) = s(1-t), \tag{20}$$

which can easily be derived from (17). Hence the vector $(B_0(t_1), ..., B_0(t_m))$ has a centered multivariate normal distribution with covariance matrix Γ.

Now consider random vectors $\xi_i := (\mathbb{1}_{\{U_i \le t_1\}}, ..., \mathbb{1}_{\{U_i \le t_m\}})$, $i = 1, ..., n$. These are independent and identically distributed. Each has mean $\mu := (t_1, ..., t_m)$ and covariance matrix Γ, because for $0 \le s < t \le 1$:

$$\mathrm{cov}(\mathbb{1}_{\{U_i \le s\}}, \mathbb{1}_{\{U_i \le t\}}) = \mathbb{E}(\mathbb{1}_{\{U_i \le s\}}\mathbb{1}_{\{U_i \le t\}}) - \mathbb{E}\mathbb{1}_{\{U_i \le s\}}\mathbb{E}\mathbb{1}_{\{U_i \le t\}} = s - st. \quad (21)$$

Note the following two things:

(a) By the multivariate CLT: $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i - \mu) \xrightarrow{\mathscr{L}} \mathrm{N}(0, \Gamma)$.

(b) $\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(\xi_i - \mu) = (\alpha_n(t_1), ..., \alpha_n(t_m))$

The lemma is proved.

∎

Lemma 3.11 *For $0 \le r < s < t \le 1$ holds*

$$\mathbb{E}|\alpha_n(s) - \alpha_n(r)|^2|\alpha_n(t) - \alpha_n(s)|^2 \le 6(r - s)(t - s). \quad (22)$$

Proof. We introduce some short-hand notation. Let

$$Y_k := \mathbb{1}_{(r,s]}(U_k) - (s - r), \qquad k = 1, ..., m,$$

and

$$Z_k := \mathbb{1}_{(s,t]}(U_k) - (t - s), \qquad k = 1, ..., m.$$

Recall that

$$\alpha_n(t) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\big(\mathbb{1}_{(0,t]}(U_k) - t\big).$$

It follows

$$\alpha_n(t) - \alpha_n(s) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\big(\mathbb{1}_{(s,t]}(U_k) - (t - s)\big) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Z_k,$$

and

$$\alpha_n(s) - \alpha_n(r) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n}\big(\mathbb{1}_{(r,s]}(U_k) - (s - r)\big) = \frac{1}{\sqrt{n}}\sum_{k=1}^{n} Y_k.$$

Let us call M the left-hand side of (22). Then

$$\begin{aligned} M &= \mathbb{E}|\alpha_n(s) - \alpha_n(r)|^2|\alpha_n(t) - \alpha_n(s)|^2 \\ &= \mathbb{E}\Big[\frac{1}{n^2}\big(\sum_{k=1}^{n} Y_k\big)^2\big(\sum_{k=1}^{n} Z_k\big)^2\Big] \\ &= \frac{1}{n^2}\sum_{1\le i,j,k,l\le n} \mathbb{E}Y_iY_jZ_kZ_l \end{aligned}$$

The Y_k and Z_k have each mean zero ($k = 1, ..., n$), and any two of them are independent as long as their indices differ. Hence, $\mathbb{E}Y_iY_jZ_kZ_l$ is zero if any of the indices i, j, k, l differs from all others. So we only need to deal with those summands whose indices appear in pairs.

$$n^2M = \sum_{i=1}^{n} \mathbb{E}Y_i^2Z_i^2 + \sum_{\substack{1\le i,j\le n\\ i\ne j}} \Big(\mathbb{E}Y_i^2Z_j^2 + \mathbb{E}Y_j^2Z_i^2 + 4\mathbb{E}Y_iY_jZ_iZ_j\Big) \tag{23}$$

We are left to compute these four expectations.

$$\begin{aligned} &\mathbb{E}Y_i^2Z_i^2 = \mathbb{E}\big[(\mathbb{1}_{(r,s]}(U_i) - (s-r))^2(\mathbb{1}_{(s,t]}(U_i) - (t-s))^2\big] \\ &= \mathbb{E}\big[\big(\mathbb{1}_{(r,s]}(U_i) - 2(s-r)\mathbb{1}_{(r,s]}(U_i) + (s-r)^2\big) \\ &\qquad \big(\mathbb{1}_{(s,t]}(U_i) - 2(t-s)\mathbb{1}_{(s,t]}(U_i) + (t-s)^2\big)\big] \\ &= \mathbb{E}\big[(t-s)^2\mathbb{1}_{(r,s]}(U_i) - 2(s-r)(t-s)^2\mathbb{1}_{(r,s]}(U_i) + (s-r)^2\mathbb{1}_{(s,t]}(U_i) \\ &\qquad -2(s-r)^2(t-s)\mathbb{1}_{(s,t]}(U_i) + (s-r)^2(t-s)^2\big] \\ &= (s-r)(t-s)^2 + (s-r)^2(t-s) - 3(s-r)^2(t-s)^2 \\ &= (s-r)(t-s)\big[(s-r) + (t-s) - 3(s-r)(t-s)\big] \end{aligned}$$

Note that the last factor is smaller than 1, thus we have $\mathbb{E}Y_i^2Z_i^2 \le (s-r)(t-s)$. By similiar computations as the one above one finds that $\mathbb{E}Y_i^2Z_j^2$, $\mathbb{E}Y_j^2Z_i^2$ and $\mathbb{E}Y_iY_jZ_iZ_j$ are also bound by $(s-r)(t-s)$. Therefore from (23):

$$\begin{aligned} n^2M &\le n\cdot(s-r)(t-s) + n(n-1)\cdot 6(s-r)(t-s) \\ \Longrightarrow M &\le (s-r)(t-s)\Big[\frac{1}{n} + 6\,\frac{n-1}{n}\Big] \le 6(s-r)(t-s). \end{aligned}$$

∎

The next theorem summarizes all our results so far.

Theorem 3.12 $\alpha_n = \{\alpha_n(t)|t \in [0,1]\}$, *defined by (16), converges in distribution to the Brownian bridge* B_0.

Proof. We apply Theorem 3.9 with B_0 and α_n in the roles of X and X_n, respectively. Lemma 3.10 takes care of condition 3.9 (1). Note that $T_{B_0} = [0,1]$, since B_0 has continuous paths. But the set T_{B_0} is of no relevance here. We know that (3.10) holds for all collections $\{t_1, ..., t_m\}$. In particular, the paths of B_0 are continuous at 1, condition 3.9 (3) is satisfied.
Finally, (22) implies 3.9 (2): take $F(t) = \sqrt{6}t$, $a = b = 2$ and note that $(s-r)(t-s) \leq (t-r)^2$. Hence, the sequence $\{\alpha_n\}$ is tight and converges in distribution to B_0.

∎

Final remark. The content of the last theorem is also treated in [Bil99] in Theorem 14.3.

4 Weak Convergence in $D(-\infty,\infty)$

This section presents the content of the previous section rewritten for $D(-\infty,\infty)$. Again, we follow the approach of [Bil99] (section 16), introducing a metric d_∞ on $D(-\infty,\infty)$ by means of the metric d. This approach is originally due to Lindvall ([Lin73]). Billingsley and Lindvall treat the space $D[0,\infty)$, but the main concept is basically the same.

4.1 The Function Space $D(-\infty,\infty)$

Definition 4.1 Let $D_\infty = D(-\infty,\infty)$ be the space of all functions $x : \mathbb{R} \to \mathbb{R}$, that are right-continuous and have left-hand limits, that is, for all $t \in \mathbb{R}$,

(i) $x(t-)$ exists, and

(ii) $x(t+)$ exists and equals $x(t)$.

Elements of D_∞, too, have at most countably many discontinuity points.

The next step is to declare a metric on D_∞. This will require some preparation. First, note that for any real a, b with $a < b$, we can define the space $D[a,b]$ analoguously to $D[0,1]$. All results we have for $D[0,1]$ apply to $D[a,b]$ just as well.

Definition 4.2 For any $a < b$ let

$$D[a,b] := \left\{ x : [a,b] \to \mathbb{R} \,\middle|\, \forall\, a < t \leq b : \exists\, x(t-);\ \forall\, a \leq t < b : \exists\, x(t+), x(t+) = x(t-) \right\}$$

In order to mathematically embody the idea of identifiabilily of $D[a,b]$ with $D[0,1]$, let

$$\varphi : [0,1] \to [a,b] : t \mapsto a + (b-a)t.$$

φ is a bijection, its inverse is

$$\varphi^{-1} : [a,b] \to [0,1] : t \mapsto \frac{t-a}{b-a}.$$

Now,

$$\phi : D[a,b] \to D[0,1] : x \mapsto x \circ \varphi. \tag{24}$$

The mapping ϕ compresses a function defined on $[a,b]$ along the time axis and moves its domain to $[0,1]$. ϕ is an isomorphism. It holds

$$\phi^{-1} : D[0,1] \to D[a,b] : x \mapsto x \circ \varphi^{-1}.$$

We define the Skorokhod metric on $D[a,b]$ by

$$d_{[a,b]}(x,y) := d_{[0,1]}(\phi(x), \phi(y)), \tag{25}$$

where $d_{[0,1]} = d$ as defined in (1). Thus, ϕ is also a isometry.

We use the following short hand notation. For any positive m let

$$D_m := D[-m, m]$$

and

$$d_m := d_{[-m,m]}.$$

The Borel-σ-field on (D_m, d_m) shall be called $\mathscr{D}_m$. Furthermore, define for $m \geq 1$ a function $g_m : \mathbb{R} \to \mathbb{R}$ as follows:

$$g_m(t) = \begin{cases} 1 & \text{if } \ |t| \leq m-1, \\ m - |t| & \text{if } \ m-1 < |t| \leq m, \\ 0 & \text{if } \ m < |t|, \end{cases}$$

and let

$$\psi_m : D_\infty \to D_m : x \mapsto (g_m \cdot x)|_{[-m,m]}.$$

The mapping ψ_m cuts the graph of $x \in D$ off at $-m$ and m, and bends the ends down to zero. This definition is valid for any real $m \geq 1$, but we will need m only to be integer-valued. With these notations we are ready to declare a metric on D_∞:

$$d_\infty(x,y) := \sum_{m=1}^{\infty} \frac{1}{2^m} \Big[1 \wedge d_m\big(\psi_m(x), \psi_m(y)\big) \Big]. \tag{26}$$

Knowing that d_m is a metric, it is easy to see that d_∞ is a metric, too. We want to call $\mathscr{D}_\infty$ the Borel-σ-field on (D_∞, d_∞).

Lemma 4.3 *(D_∞, d_∞) is a separable metric space.*

Proof. cf. [Bil99], page 170, Theorem 16.3.

□

The question remains, why we have brought in ψ_m. Why not take

$$\tilde{d}_\infty(x,y) := \sum_{m=1}^{\infty} \frac{1}{2^m}\Big(1 \wedge d_m\big(x|_{[-m,m]}, y|_{[-m,m]}\big)\Big) \tag{27}$$

as metric on D_∞ instead of d_∞? The answer shall be illustrated by the following example. As mentioned in the previous section, in $D[0,1]$, it holds

$$\mathbb{1}_{[0,\frac{1}{2}-\frac{1}{n})} \longrightarrow \mathbb{1}_{[0,\frac{1}{2})},$$

which remains true if we replace $\frac{1}{2}$ by any $\tau \in (0,1)$. However, it fails for $\tau = 1$. In $(D[0,1], d)$, $\mathbb{1}_{[0,1-\frac{1}{n})}$ does not converge to $\mathbb{1}_{[0,1)}$, neither does $\mathbb{1}_{[\frac{1}{n},1]}$ to $\mathbb{1}_{[0,1]}$. This is due to the fact that every $\lambda \in \Lambda$ is fixed at 0 and 1. For a detailed argumentation see Lemma A.4 in the appendix. It immediately follows that, e.g.

$$\mathbb{1}_{[-1,1-\frac{1}{n})} \nrightarrow \mathbb{1}_{[-1,1)} \qquad \text{in } D[-1,1]$$

From (27) we see that functions $x_n \in D(-\infty,\infty)$ converge x with respect to $\tilde{d}_\infty$ if and only if for all integers $m \geq 1$

$$x_n|_{[-m,m]} \xrightarrow{d_m} x|_{[-m,m]}.$$

Hence we know that

$$\mathbb{1}_{[-1,1-\frac{1}{n})} \nrightarrow \mathbb{1}_{[-1,1)} \qquad \text{in } (D_\infty, \tilde{d}_\infty)$$

now the indicator functions being defined on the whole of $\mathbb{R}$. But of course, in a sensible extension of $D[0,1]$-theory to $D(-\infty,\infty)$, these functions must converge. We see that problems occur, when the limiting function has a discontinuity at a cut-off point. Here the ψ_m help out by making each function continuous before it is cut off. More detailed explanations can be found in [Bil99], pages 167 through 169.

4.2 A Criterion for Weak Convergence in $D(-\infty,\infty)$

We repeat the program of section 3.3. The results will either be proved by making use of the corresponding statement in the $D[0,1]$-case, or the respective proofs can be adopted with little notational change. In the latter case the proofs are omitted.

Recall the projections π_t, π_T. We extend their definition to D_∞, i.e. π_t, π_T are defined as before by (3), (4), respectively, but for all $t \in \mathbb{R}$ and (finite)

$T \subset \mathbb{R}$. Of course, any function on $\mathbb{R}$ or a subset of $\mathbb{R}$ is a valid argument of π_t and π_T. We use the same notation regardless of the particular domain space.
Furthermore, recall $\mathscr{F}(T_0)$, specified by (7). We now assume T_0 to be a subset of $\mathbb{R}$.

Proposition 4.4 *Let T_0 be a dense subset of $\mathbb{R}$. Then $\sigma\{\pi_t|t \in T_0\} = \sigma(\mathscr{F}(T_0)) = \mathscr{D}_\infty$.*

Proof. We only show $\sigma\{\pi_t|t \in T_0\} = \mathscr{D}_\infty$. The equality $\sigma\{\pi_t|t \in T_0\} = \sigma(\mathscr{F}(T_0))$ is easy to get: $\sigma\{\pi_t|t \in T_0\} \subset \sigma(\mathscr{F}(T_0))$ is trivial, and the inverse inclusion follows by the same argument as in the $D[0,1]$-case. In fact, the proof is literally the same as the one of 3.5 (a). Now that we work with $\sigma\{\pi_t|t \in T_0\}$ instead of $\sigma(\mathscr{F}(T_0))$, let us abbreviate:

$$\mathscr{S} := \mathscr{S}(T_0) := \big\{\pi_t^{-1}(A)\big|t \in T_0, A \in \mathscr{B}(\mathbb{R})\big\},$$

thus $\sigma(\mathscr{S}) = \sigma\{\pi_t|t \in T_0\}$.

Proof of $\sigma(\mathscr{S}) \subset \mathscr{D}_\infty$: We show, for $t \in \mathbb{R}$, π_t is $(\mathscr{D}_\infty, \mathscr{B}(\mathbb{R}))$-measurable. It is so as a composition of measurable mappings. For the scope of this proof let $\pi_s^{[0,1]}$, $0 \le s \le 1$, be the projection from (the metric space) $D[0,1]$ into $\mathbb{R}$ (as it is defined in (3)). We know that $\pi_s^{[0,1]}$ is $(\mathscr{D}, \mathscr{B}(\mathbb{R}))$-measurable for every $s \in [0,1]$ (Proposition 3.5). Now let $t \in \mathbb{R}$ be arbitrary and $m \in \mathbb{N}$ such that $m > |t| + 1$. Recall the functions $\psi_m : D_\infty \to D_m$ and $\phi_m := \phi : D_m \to D[0,1]$, the latter defined by (24). From the definitons of the metrics d_∞ and d_m it is immediately clear that these two functions are continuous, hence measurable with respect to the corresponding Borel-σ-fields. Finally, note that

$$\pi_t = \pi^{[0,1]}_{\varphi^{-1}(t)} \circ \phi_m \circ \psi_m.$$

Proof of $\mathscr{D}_\infty \subset \sigma(\mathscr{S})$: The first step is to show that $\psi_m : D_\infty \to D_m$ is $(\sigma(\mathscr{S}), \mathscr{D}_m)$-measurable. Let $T_m := (T_0 \cap [-m,m]) \cup \{-m,m\}$, and note that it is dense in $[-m,m]$. For reasons of clarity we will call $\pi_t^{(m)}$ the projection with the domain D_m. From Proposition 3.5 we know that $\mathscr{D} = \sigma\{\pi_t^{[0,1]}|t \in (T_0 \cap [0,1]) \cup \{0,1\}\}$. By analoguous argument holds $\mathscr{D}_m = \sigma\{\pi_t^{(m)}|t \in T_m\}$. Hence, it suffices to prove that ψ_m is $(\sigma(\mathscr{S}), \sigma\{\pi_t^{(m)}|t \in T_m\})$-measurable. We apply Theorem 2.6, and show that

$$\psi_m^{-1}\big((\pi_t^{(m)})^{-1}(A)\big) \in \sigma(\mathscr{S})$$

for all $A \in \mathscr{B}(\mathbb{R})$ and $t \in T_m$. The key is to notice that

$$\pi_t^{(m)}(\psi_m(x)) = \psi_m(x)(t) = g_m(t)\pi_t(x) = g_m(t)x(t).$$

We distinguish two cases:

(I) $t \in \{-m, m\}$:
It holds $x(t)g_m(t) = 0$ for all $x \in D_\infty$, hence

$$\psi_m^{-1}\big((\pi_t^{(m)})^{-1}(A)\big) = \begin{cases} \emptyset & \text{if } 0 \notin A \\ D_\infty & \text{if } 0 \in A. \end{cases}$$

These two sets are elements of $\sigma(\mathscr{S})$.

(II) $t \in (-m, m) \cap T_0$:
It is $g_m(t) \neq 0$, and we can write

$$\begin{aligned} \psi_m^{-1}\big((\pi_t^{(m)})^{-1}(A)\big) &= \big\{x \in D_\infty \big| \pi_t^{(m)}(\psi_m(x)) \in A\big\} \\ &= \big\{x \in D_\infty \big| g_m(t)\pi_t(x) \in A\big\} \\ &= \big\{x \in D_\infty \big| \pi_t(x) \in \frac{1}{g_m(t)} \cdot A\big\} \end{aligned}$$

This set evidently lies in $\sigma(\mathscr{S})$, too.

Now we know that ψ_m is $(\sigma(\mathscr{S}), \mathscr{D}_m)$-measurable. Using this result we show next: $d_\infty(\cdot, y) : D_\infty \to \mathbb{R}$ is $(\sigma(\mathscr{S}), \mathscr{B}(\mathbb{R}))$-measurable for every $y \in D_\infty$. Note that for $y \in D_m$, $d_m(\cdot, y)$ is $(\mathscr{D}_m, \mathscr{B}(\mathbb{R}))$-measurable, since it is continuous. Then $d_m(\psi_m(\cdot), y)$ is $(\sigma(\mathscr{S}), \mathscr{B}(\mathbb{R}))$-measurable (composition), and so is

$$d_\infty(\cdot, y) := \sum_{m=1}^{\infty} \frac{1}{2^m}\Big(1 \wedge d_m\big(\psi_m(\cdot), \psi_m(y)\big)\Big)$$

(countable sum of measurable functions).
This means, $\sigma(\mathscr{S})$ contains all (open) balls. These are just the sets $d_\infty(\cdot, y)^{-1}([0, \varepsilon))$, $\varepsilon > 0$, $y \in D_\infty$. The space (D_∞, d_∞) is separable, hence $\sigma(\mathscr{S})$ contains all open sets and consequently the whole σ-field $\mathscr{D}_\infty$. (For a detailed argumentation see appendix, Lemma A.5.) The proof is completed.

■

Lemma 4.5 *For any $T_0 \subset \mathbb{R}$, $\mathscr{F}(T_0)$ is a π-system.*

Proof. See proof of Lemma 3.6. □

Proposition 4.6 *If T_0 is dense in $\mathbb{R}$, then $\mathscr{F}(T_0)$ is separating class for $\mathscr{D}_\infty$.*

Proof. This follows by Theorem 2.3 from Proposition 4.4 and Lemma 4.5. ■

Now consider a probability measure P on $(D_\infty, \mathscr{D}_\infty)$. As before, let

$$T_P := \Big\{t \in \mathbb{R} \Big| P(\{x \in D_\infty | \pi_t \text{ is continuous at } x\}) = 1\Big\}. \tag{28}$$

Despite the same symbol this definition can not be mixed up with (6), due to the different domain of the measure P.

Proposition 4.7 *The complement of T_P in $\mathbb{R}$ is at most countable.*

Proof. See [Bil99], page 174. ■

Now we can formulate and prove the generalisation of Theorem 3.8.

Theorem 4.8 *Let $\{P_n\}$ be a sequence of probability measures on $(D_\infty, \mathscr{D}_\infty)$ with the following two properties:*

(1) $\{P_n\}$ is tight.

(2) There exists a measure P on $(D_\infty, \mathscr{D}_\infty)$ such that

$$P_n \circ \pi_T^{-1} \xrightarrow{\mathscr{L}} P \circ \pi_T^{-1}, \qquad \textit{for all finite } T \subset T_P. \tag{29}$$

Then $P_n \xrightarrow{\mathscr{L}} P$ in $(D_\infty, \mathscr{D}_\infty)$.

Proof. The proof is the same as the one of Theorem 3.8, just take Proposition 4.7 instead of 3.4 and Proposition 4.6 instead of 3.7. ■

4.3 A Criterion for Tightness in $D(-\infty, \infty)$

As before in $D[0,1]$, we still need a nice tightness criterion in order to make use of Theorem 4.8. In this section we state and prove Theorem 4.9, a D_∞-analogon of Theorem 3.9, which will be used in subsection 4.4 to show tightness of $\{\beta_n\}$. There are actually several possibilities of tightness criteria that work for this particular sequence. For a short discussion see the remarks at the end of this subsection.

Theorem 3.9 is more or less a citation from [Bil99]. This book also treats the space $D[0,\infty)$, but does not contain a similiar result for this space. A proposition of this type can be found, for example, in [JS02](Theorem 4.1, page 355), but for the proof the authors just refer to [Bil99]. I am going to give a detailed proof of Theorem 4.9 in this paper, which, of course, is inspired by Billingsley's proof of Theorem 3.9. As before, when we derived the metric d_∞, a few details must be taken care of when moving from $D[0,1]$ to $D(-\infty,\infty)$. Therefore, theorem and proof are organised in this separate subsection.

Just like 3.9, Theorem 4.9 is formulated in terms of random variables. That way, notation gets more convenient. All random variables we consider are defined on the same probability space $(\Omega, \mathscr{A}, \mathbb{P})$. We denote the distribution of the random element X by P_X, and write T_X for T_{P_X}.

Theorem 4.9 *Let $X, X_1, X_2, ...$ be random variables in $(D_\infty, \mathscr{D}_\infty)$. Suppose that*

(1) $\big(X_n(t_1), ..., X_n(t_k)\big) \xrightarrow{\mathscr{L}} \big(X(t_1), ..., X(t_k)\big)$

as $n \to \infty$ for points $t_1 < t_2 < ... < t_k$ in T_X, and

(2) there exist a non-decreasing, continuous function $F : \mathbb{R} \to \mathbb{R}$ and real numbers $a > 1$ and $b \geq 0$, such that

$$\mathbb{E}\Big(|X_n(s) - X_n(r)|^b |X_n(t) - X_n(s)|^b\Big) \leq \big(F(t) - F(r)\big)^a.$$

holds for all $r < s < t$, and $n \geq 1$

Then $X_n \xrightarrow{\mathscr{L}} X$ in $(D_\infty, \mathscr{D}_\infty)$.

Remarks.

(I) Theorem 4.9 states convergence in law. I refer to it as a tightness criterion, because assumptions (1) and (2) imply tightness, from which by Theorem 4.8 convergence follows.

(II) The main and almost only difference to Theorem 3.9 is the absence of a third condition similiar to 3.9 (3). The point 1 plays a special role in $D[0,1]$-theory, since the value of a $D[0,1]$-function at 1 is not connected by any continuity-type relation to values at neighboring points. It does not play any special role in $D(-\infty,\infty)$.

But in fact, 3.9 (3) is a pretty strong condition, it is far from being necessary. Take, e.g., a constant sequence $\{P_n\}$, where P_n has all the mass concentrated in one point x, and that x has a jump at 1. Condition 3.9 (3) is just one way to ensure (in combination with the convergence of the finite-dimensional distributions) that forall $\varepsilon, \eta > 0$ there exist a $0 < \delta < 1$ and an $n_0 \in \mathbb{N}$ such that

$$\mathbb{P}\big(|X_n(1-) - X_n(1-\delta)| \geq \varepsilon\big) \leq \eta, \qquad \forall\, n \geq n_0,$$

i.e. $X_n(1-\delta)$ converges to $X_n(1-)$ for all n at the same rate. This condition is indeed necessary (cf. [Bil99], Theorem 13.2 and the subsequent corollary, in particular (13.8), see also bottom of page 141). Note that this does not impose any condition on $X(1)$ or $X_n(1)$.

The rest of this subsection is devoted to the proof of Theorem 4.9. Some preparation is necessary before we get to the actual proof. Let us temporarily return to the space $D = D[0,1]$.
For an arbitrary function $x : [0,1] \to \mathbb{R}$ we can define

$$w(x,\delta) := \sup_{|s-t|\leq\delta} |x(s) - s(t)|, \qquad 0 < \delta \leq 1. \tag{30}$$

The functional w is called the *modulus of continuity* and characterizes continuous functions the following way.

Lemma 4.10 *The function $x : [0,1] \to \mathbb{R}$ is continuous if and only if $w(x,\delta) \to 0$ as $\delta \to 0$.*

Proof. cf. [Bil99], page 80.

□

For an arbitrary subset T of $[0,1]$ let

$$w_x(T) := \sup_{s,t\in T} |x(s) - x(t)|. \tag{31}$$

Another way of expressing $w(x,\delta)$, using w_x, is

$$w(x,\delta) = \sup_{t\in[0,1-\delta]} w_x([t,t+\delta]), \qquad 0 < \delta \leq 1. \tag{32}$$

Now let $S = \{s_0, ..., s_\nu\}$ be a grid on $[0, 1]$ (cf. subsection 3.2), and $0 < \delta < 1$. A grid is called *δ-sparse*, or short a *δ-grid*, if

$$s_i - s_{i-1} > \delta \qquad \forall\, i = 1, ..., \nu,$$

i.e. the distance between two neighboring grid points is greater than δ. Let $\mathscr{S}_0(\delta)$ be the set of all δ-sparse grids on $[0, 1]$, i.e.

$$\mathscr{S}_0(\delta) := \Big\{\{s_0, ..., s_\nu\}\Big|\nu \in \mathbb{N}, s_0 = 0, s_\nu = 1, s_i - s_{i-1} > \delta\;\forall\; i = 1, ..., \nu\Big\} \quad (33)$$

We want to call the grid S a *δ-grid with soft boundaries*, if

$$s_i - s_{i-1} > \delta \qquad \forall\, i = 2, ..., \nu - 1,$$

i.e. all intervalls must be wider than δ except those at the left and the right end. Let $\mathscr{S}_1(\delta)$ be the set of all δ-grids with soft boundaries on $[0, 1]$, i.e.

$$\mathscr{S}_1(\delta) := \Big\{\{s_0, ..., s_\nu\}\Big|\nu \in \mathbb{N}, s_0 = 0, s_\nu = 1, s_i - s_{i-1} > \delta\;\forall\, i = 2, ..., \nu - 1\Big\}.$$

Now define the following two moduli:

$$w'(x, \delta) := \inf_{\mathscr{S}_0(\delta)} \max_{1\le i\le\nu} w_x[s_{i-1}, s_i),$$

and

$$\hat{w}(x, \delta) := \inf_{\mathscr{S}_1(\delta)} \max_{1\le i\le\nu} w_x[s_{i-1}, s_i). \quad (34)$$

The modulus w' plays the same role for $D[0, 1]$ as w does for $C[0, 1]$.

Lemma 4.11 *The function $x : [0, 1] \to \mathbb{R}$ is an element of $D[0, 1]$ if and only if $w'(x, \delta)$ tends to zero as $\delta \to 0$.*

Proof. cf. [Bil99], page 123.

□

The purpose of $\hat{w}$ will become clear later. First, we introduce another modulus. Let $\mathscr{T}_0(\delta)$ be the set of all triples $\{t_1, t, t_2\}$, where $0 \le t_1 \le t \le t_2 \le 1$ and $t_2 - t_1 \le \delta$, and

$$w''(x, \delta) := \sup_{\mathscr{T}_0(\delta)} \big\{|x(t) - x(t_1)| \wedge |x(t_2) - x(t)|\big\}. \quad (35)$$

We now turn to the question how these moduli are related to each other. Obviously,

$$\hat{w}(x,\delta) \leq w'(x,\delta) \tag{36}$$

for all x and δ, since the infimum on the left side extends over a larger set. Moreover, it holds

$$w''(x,\delta) \leq w'(x,\delta), \tag{37}$$

(cf. [Bil99], page 131, (12.28)). There can be now general inequality like (37) in the opposite direction (allowing appropriate multiples), as the following example illustrates (taken from [Bil99]). Let

$$y_n := \mathbb{1}_{[0,\frac{1}{n})}, \qquad n \geq 1. \tag{38}$$

It holds

$$w''(y_n,\delta) = 0 \qquad \text{for all } \delta > 0$$

(y_n has two constant pieces),

$$\hat{w}(y_n,\delta) = 0 \qquad \text{for all } \delta > 0$$

(take the grid $\{0, \frac{1}{n}, 1\}$), but

$$w'(y_n,\delta) = \begin{cases} 0 & \text{if } \delta < \frac{1}{n} \quad (n \geq 1), \\ 1 & \text{if } \delta \geq \frac{1}{n} \quad (n \geq 2). \end{cases}$$

There *is* an inequality in the opposite direction if we take into account additional information about the behaviour at the end points.

$$w'(x,\frac{\delta}{2}) \leq 24\Big\{w''(x,\delta) \vee |x(\delta) - x(0)| \vee |x(1-) - x(1-\delta)|\Big\}, \tag{39}$$

(cf. [Bil99], page 132, (12.32)).
For our purpose here we primarily need a relationship between w'' and $\hat{w}$. There we have

$$\hat{w}(x,\frac{\delta}{2}) \leq 6w''(x,\delta) \tag{40}$$

This can be shown by the same means as (39) and is treated in Lemma A.7 in the appendix.

Finally, we define w_m, w'_m, $\hat{w}_m$ and w''_m to be the corresponding functionals for functions living on $[-m, m]$ or on a superset of $[-m, m]$. The definitions look exactly the same except that 0 and 1 are replaced by $-m$ and m, respectively. Then it is easy to see that, for example,

$$w''_m(x, \delta) = w''(\varphi_m(x), \frac{\delta}{2m}), \tag{41}$$

which may as well serve as a definition. The same is true for w_m, $\hat{w}_m$ and w'_m. This makes it also immediately clear that all results (36),(37),(39) and (40) hold as well, in particular

$$\hat{w}_m(x, \frac{\delta}{2}) \leq 6w''_m(x, \delta). \tag{42}$$

We explicitely want to allow this notation for functions x defined on the whole of $\mathbb{R}$, i.e. $w'_m(x, \delta) = w'_m(x|_{[-m,m]}, \delta)$, likewise for w_m, $\hat{w}_m$ and w''_m.

The next proposition establishes a link between these moduli and the problem at hand, to prove Theorem 4.9.

Proposition 4.12 *Let $m > 0$ and Y be a random element in $D[-m, m]$, furthermore $a > 1$, $b \geq 0$ and μ a finite measure on $[-m, m]$. If, for positive numbers λ and δ, holds*

$$\mathbb{P}\big(|Y(s) - Y(r)| \wedge |Y(t) - Y(s)| \geq \lambda\big) \leq \frac{1}{\lambda^{2b}}\mu^a(r, t] \tag{43}$$

for all $-m \leq r \leq s \leq t \leq m$ with $t - r < 2\delta$, then a constant c (depending only on a and b) exists such that

$$\mathbb{P}\big(w''_m(Y, \delta) \geq \lambda\big) \leq \frac{c}{\lambda^{2b}}\mu[-m, m] \sup_{t\in[-m,m-2\delta]} \mu^{a-1}[t, t + 2\delta].$$

Proof. If we write down the proposition for $[0, 1]$ instead of $[-m, m]$, then it is an immediate corollary of Theorem 10.4 on page 112 in [Bil99]. The case $[-m, m]$ can be treated in an absolute analoguous manner, or, even easier, can be derived from the $[0, 1]$-case by considering (instead of Y and μ) the random variable $\varphi_m(Y)$ and the finite measure μ_0 on $[0, 1]$, defined by

$$\mu_0(s, t] := \mu(\varphi_m(s), \varphi_m(t)], \qquad 0 \leq s \leq t \leq 1.$$

If we know that the theorem holds for these, then by the relation (41), it also holds for Y and μ.

∎

Corollary 4.13 *Let Y be a random element in $D[-m,m]$, $a > 1$, $b \geq 0$ and F a non-decreasing, continuous function on $[-m,m]$. If, for positive numbers δ and λ, holds*

$$\mathbb{P}\big(|Y(s) - Y(r)| \wedge |Y(t) - Y(s)| \geq \lambda\big) \leq \frac{1}{\lambda^{2b}}\big(F(t) - F(s)\big)^a \tag{44}$$

for all $-m \leq r \leq s \leq t \leq m$ with $t - r < 2\delta$, then a constant c exists such that

$$\mathbb{P}\big(\hat{w}_m(Y,\delta) \geq \lambda\big) \leq c\Big(\frac{6}{b}\Big)^{2b}\Big[F(m) - F(-m)\Big]\Big[w_m(F, 4\delta)\Big]^{2a-1}.$$

Proof. $\mu(s,t] := F(t) - F(s)$ for all $-m \leq s \leq t \leq m$ defines a finite measure on $[-m,m]$. Due to the continuity and monotony of F we have

$$\mu[t, t+\delta) = F(t+\delta) - F(t) = \sup_{s,r\in[t,t+\delta]} |F(s) - F(r)| = w_F([t, t+\delta]),$$

and the modulus of continuity (cf. (32)) of F reads as

$$w_m(F,\delta) = \sup_{t\in[-m,m-\delta]} \mu[t, t+\delta].$$

Since the mapping $t \to t^{a-1}$, $(t \in \mathbb{R}, a > 1)$ is monotonuous and continuous as well,

$$w_m^{a-1}(F,\delta) = \sup_{t\in[-m,m-\delta]} \mu^{a-1}[t, t+\delta].$$

Then, (44) is the same as (43), thus by Proposition 4.12,

$$\mathbb{P}\big(w_m''(Y,\delta) \geq \lambda\big) \leq \frac{c}{\lambda^{2b}}\big[F(m) - F(-m)\big] w_m^{2a-1}(F, 2\delta).$$

Because of (42) we have the following implication

$$\hat{w}_m(y,\delta) \geq \lambda \Longrightarrow 6w_m''(y, 2\delta) \geq \lambda,$$

hence

$$\mathbb{P}\big(\hat{w}_m(Y,\delta) \geq \lambda\big) \leq \mathbb{P}\big(w_m''(Y,2\delta) \geq \frac{\lambda}{6}\big) \leq c\Big(\frac{6}{b}\Big)^{2b}\Big[F(m) - F(-m)\Big]\Big[w_m(F,4\delta)\Big]^{2a-1}.$$

■

This finishes the preparations. We are now ready to prove Theorem 4.9. We will make use of the following proposition, which is the D_∞-equivalent of Theorem 16.8 in combination with the subsequent corollary, to be found in [Bil99]. Theorem 16.8 is formulated for $D[0,\infty)$, the proof works just the same.

Theorem 4.14 *A sequence of random variables $\{X_n\}$ in $(D_\infty, \mathscr{D}_\infty)$ is tight if and only if the following two conditions hold.*

(1) For all t in a dense subset T_0 of $\mathbb{R}$ holds

$$\lim_{a\to\infty} \limsup_n \mathbb{P}(|X_n(t)| \geq a) = 0. \tag{45}$$

(2) For every $m \in \mathbb{N}$ and $\varepsilon > 0$ holds

$$\lim_{\delta\to 0} \limsup_n \mathbb{P}(\hat{w}_m(X_n, \delta) \geq \varepsilon) = 0. \tag{46}$$

Proof of Theorem 4.9. This proof is based upon the demonstration of the analoguous result for the $D[0,1]$-case, as to be found in [Bil99] (see remarks at the end of this subsection).
We have to show that the assumptions of Theorem 4.9 imply those of Theorem 4.14. In fact, 4.14 (1) follows already from 4.9 (1), while 4.14 (2) is a conclusion from 4.9 (2).

We start by showing 4.14 (1), which is the easier part and literally the same as in $D[0,1]$. Billingsley treats it in three lines on page 141. Here, it is written down in a more detailed way. Suppose we have 4.9 (1). T_X is dense in $\mathbb{R}$ (see Proposition 4.7), and takes on the role of T_0 in Theorem 4.14. Let μ and μ_n be the distributions of $X(t)$ and $X_n(t)$, in other words $\mu := \mathbb{P} \circ X^{-1} \circ \pi_t^{-1}$ and $\mu_n := \mathbb{P} \circ X_n^{-1} \circ \pi_t^{-1}$. By the assumption we have $X_n(t) \xrightarrow{\mathscr{L}} X(t)$, which is the same as $\mu_n \xrightarrow{\mathscr{L}} \mu$. This means, $\{\mu_n\}$ is a weakly convergent sequence of probability measures on the complete and separable metric space $\mathbb{R}$, hence it is tight (cf. Corollary 2.13 and Theorem 2.14 (2)). Recall the definition of tightness (2.11): for every $\varepsilon > 0$ exists a compact set $K \in \mathbb{R}$ such that

$$\mu_n(K) > 1 - \varepsilon, \qquad n \in \mathbb{N}.$$

K is bounded and hence contained in a set of the kind $(-a, a)$ for a suited $a > 0$. Therefore: for every $\varepsilon > 0$ exists an $a > 0$ such that

$$\mu_n(-a, a) > 1 - \varepsilon,$$

and consequently

$$\mathbb{P}(|X_n(t)| \geq a) = \mu_n\big[(-\infty, -a] \cup [a, \infty)\big] < \varepsilon.$$

This holds for all $n \in N$, hence

$$\limsup_n \mathbb{P}(|X_n(t)| \geq a) \leq \varepsilon. \tag{47}$$

If (47) holds for an a_0 it holds also for all $a > a_0$ (monotony of measures), so this is just the definition of the convergence

$$\lim_{a\to\infty} \limsup_n \mathbb{P}(|X_n(t)| \geq a) = 0.$$

This completes the first part of the proof and we turn to the second part, deriving 4.14 (2) from 4.9 (2). With the preparation we have put in so far, this is not longer than the first part. First of all, 4.9 (2) implies

$$\mathbb{P}\Big(|X_n(s) - X_n(r)| \wedge |X_n(t) - X_n(s)| \geq \lambda\Big) \leq \frac{1}{\lambda^{2b}}\big(F(t) - F(r)\big)^a \tag{48}$$

for all $r \leq s \leq t$, $\lambda > 0$ and $n \in \mathbb{N}$. This follows by exactly the same argument as in the $D[0,1]$-case (Markov inequality, cf. proof of 3.9). The next step is to note that (48) holds as well if we restrict r, s and t to be from the intervall $[-m, m]$, and $t - r < 2\delta$ for any positive δ. This means that the assumptions of Corollary 4.13 are met. In order to argue precisely at this point we have to note the following two things. If X_n is a random variable in $(D_\infty, \mathscr{D}_\infty)$, then $X_n^{(m)}$, defined by

$$X_n^{(m)}(\omega) := X_n(\omega)\Big|_{[-m,m]}, \qquad \omega \in \Omega,$$

is a random variable in $(D_m, \mathscr{D}_m)$, and for this random variable holds

$$\Big(X_n^{(m)}(t)\Big)(\omega) = \Big(X_n(t)\Big)(\omega), \qquad \omega \in \Omega, t \in [-m, m].$$

This means that the distributions of X_n and $X_n^{(m)}$ (which are measures on $\mathscr{D}_\infty$ and $\mathscr{D}_m$, respectively) are the same on sets that are only defined by values at times t within the intervall $[-m, m]$, especially

$$\mathbb{P}\Big(|X_n(s) - X_n(r)| \wedge |X_n(t) - X_n(s)| \geq \lambda\Big) = \mathbb{P}\Big(|X_n^{(m)}(s) - X_n^{(m)}(r)| \wedge |X_n^{(m)}(t) - X_n^{(m)}(s)| \geq \lambda\Big).$$

for all $-m \leq r \leq s \leq t \leq m$. To this $X_n^{(m)}$ (in the role Y) we can apply Corollary 4.13. It yields

$$\mathbb{P}\big(\hat{w}_m(X_n, \delta) \geq \lambda\big) \leq c\Big(\frac{6}{b}\Big)^{2b}\big[F(m) - F(-m)\big]\, w_m^{2a-1}(F, 4\delta).$$

This holds for all $n \in \mathbb{N}$, $m \in \mathbb{N}$ and $\delta > 0$, hence

$$\limsup_{n} \mathbb{P}\big(\hat{w}_m(X_n, \delta) \geq \lambda\big) \leq c\Big(\frac{6}{b}\Big)^{2b} \big[F(m) - F(-m)\big] \big[w_m(F, 4\delta)\big]^{2a-1}. \quad (49)$$

Since F is continuous, due to 4.10 the right-hand side goes to zero as $\delta \to 0$,

$$\lim_{\delta \to 0} \limsup_{n} \mathbb{P}\big(\hat{w}_m(X_n, \delta) \geq \lambda\big) = 0, \qquad m \in \mathbb{N}.$$

This completes the proof of Theorem 4.9.

∎

Remarks.

(I) A few words about how this proof differs from the $D[0,1]$-case: The central piece of the proof is Theorem 4.14, its counterpart in $D[0,1]$ is the following.

Theorem 4.15 *A sequence $\{P_n\}$ of probability measures on $(D, \mathscr{D})$ is tight if and only if the following two conditions hold.*

(1) T_0 is a dense set in $[0,1]$ containing 1, and for each $t \in T_0$ holds

$$\lim_{a \to \infty} \limsup_{n} \mathbb{P}(|X_n(t)| \geq a) = 0.$$

(2) For every $\varepsilon > 0$ holds

$$\lim_{\delta \to 0} \limsup_{n} \mathbb{P}(w'(X_n, \delta) \geq \varepsilon) = 0. \quad (50)$$

This can be found in [Bil99] as Theorem 13.2 and the subsequent corollary. The main difference between 4.9 and 4.15 is that in the second conditon of 4.9 we have $\hat{w}$ instead of w'. This accounts for the fact, that the cut-off points at $-m$ and m are chosen arbitrarily.

It means that in $D[0,1]$ one uses relation (39) in the place of (40) to derive from Proposition 4.12 a corollary, similiar to 4.13 but involving w' instead of $\hat{w}$.

Since (39) has a more complicated structure than (40) the proof is in fact more complicated in the $D[0,1]$-case. Also keep in mind, in $D[0,1]$ we need an additional assumption (cf. 3.9 (3))

(II) The discussion promised at the beginning of this subsection consists of citing the following, alternative tightness criterion from [JS02]. Briefly, it states, if the converging sequence as well as the limit are point processes, then convergence the finite-dimensionals automatically implies convergence in law. In detail, let $\mathscr{Z}$ be the following set of sequences:

$$\mathscr{Z} := \big\{\{t_n\}\big|\, t_n \in (0,\infty] \forall\, n \in \mathbb{N}; t_n < t_{n+1} \text{ if } t_n < \infty; t_n \to \infty\big\},$$

i.e. the elements of $\mathscr{Z}$ are unbounded, increasing sequences of positive numbers. The value $+\infty$ is allowed, too, and the sequence must be strictly increasing as long as it is not ∞. We want to call a stochastic process X on $[0,\infty)$ a *point process* or a *counting process* if its trajectories lie in the following set $\mathscr{V}$, the set of all *counting functions*:

$$\mathscr{V} := \big\{x : [0,\infty) \to \mathbb{R}\big| x = \sum_{n=1}^{\infty} \mathbb{1}_{[t_n,\infty)}, \{t_n\} \in \mathscr{Z}\big\},$$

i.e. a counting functions is an element of $D[0,\infty)$, which is increasing and piecewise constant, it may be bounded, but has only jumps of height one.

Proposition 4.16 *If $X_n, n \in \mathbb{N}$ and X are point processes, and if the finite-dimensional distributions of X_n converge to those X, i.e.*

$$\big(X_n(t_1), ..., X_n(t_k)\big) \xrightarrow{\mathscr{L}} \big(X(t_1), ..., X(t_k)\big)$$

as $n \to \infty$ for points $0 \le t_1 < t_2 < ... < t_k$ in T_X, then $X_n \xrightarrow{\mathscr{L}} X$ in $D[0,\infty)$.

Proof. see [JS02], page 354, Theorem 3.37. □

I just mention this here in order to give an idea of what else is possible. Therefore we do not worry at this point about details such as what exactly the metric on $D[0,\infty)$ may look like. $D[0,\infty)$-theory is analogue to $D(-\infty,\infty)$, and extensively treated in standard volumes such as [JS02], [Bil99], [Pol84] or [EK86].

The criterion is obviously very handy for the example $\{\beta_n\}$ in the next subsection. It applies directly to β_n if restricted to $[0,\infty)$, or would need to be adjusted in a suitable way for the whole real line.

However, we will make use of Theorem 4.9, which we have proved thoroughly and which works also for the sequences we deal with later on in section 5.2.

4.4 The Rescaled Empirical Distribution Function

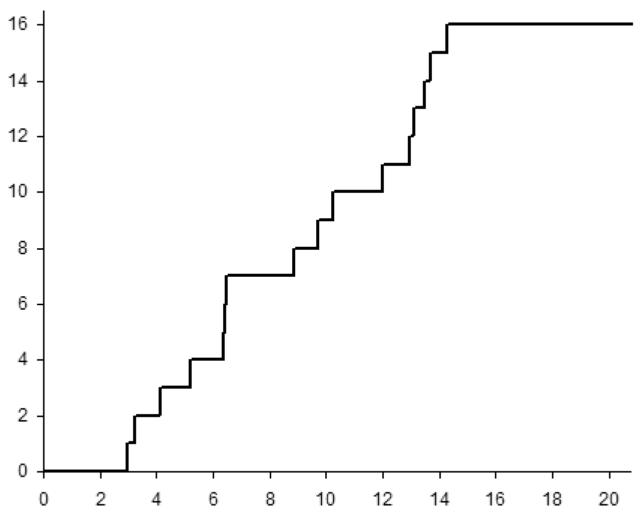
We now present an example of weak convergence in $(D_\infty, \mathscr{D}_\infty)$. This will be $\beta_n \xrightarrow{\mathscr{L}} N$, mentioned in the introduction. The paths of these processes are naturally defined on $D[0,\infty)$. We view them as random elements in $(D_\infty, \mathscr{D}_\infty)$ and treat them with the theory developped so far.
Let $U_1, U_2, ...$ be a sequence of independent, identically uniformly on $[0,1]$ distributed random variables. Recall the empirical distribution function of $U_1, ..., U_n$:

$$G_n(t) := \frac{1}{n} \sum_{k=1}^{n} \mathbb{1}_{\{U_k \le t\}}, \qquad t \in \mathbb{R}. \tag{51}$$

Let

$$\beta_n(t) := n\, G_n\left(\tfrac{t}{n}\right) = \sum_{k=1}^{n} \mathbb{1}_{\{U_k \le \frac{t}{n}\}}, \qquad t \in \mathbb{R}. \tag{52}$$

We want to call $\beta_n = \{\beta_n(t) | t \in \mathbb{R}\}$ the *rescaled empirical distribution function* of $U_1, ..., U_n$. It is a random element in $(D_\infty, \mathscr{D}_\infty)$. The figure below depicts a path of β_{16}, which is formed by the same realisation of $U_1, ..., U_{16}$ as the path of α_{16} shown on page 21.



Let $N = \{N(t) | t \in [0,\infty)\}$ be a Poisson process with rate 1. We define $\tilde{N}$ to be the following random element in $(D_\infty, \mathscr{D}_\infty)$:

$$\tilde{N}(t) := \begin{cases} 0, & \text{if } t < 0, \\ N(t) & \text{if } t \ge 0 \end{cases} \tag{53}$$

Theorem 4.17 $\beta_n \xrightarrow{\mathscr{L}} \tilde{N}$ *in* $(D_\infty, \mathscr{D}_\infty)$.

For the proof we need the following two lemmas.

Lemma 4.18 *Let* $-\infty < t_1 < ... < t_m < \infty$. *Then*

$$\big(\beta_n(t_1), ..., \beta_n(t_m)\big) \xrightarrow{\mathscr{L}} \big(\tilde{N}(t_1), ..., \tilde{N}(t_m)\big).$$

Proof. In the analoguous Lemma 3.10 we used the Central Limit Theorem. The limiting vector here does not have a normal distribution, a different approach is necessary. We are in the lucky situation though, that the converging vectors and the limit vector all are discrete. Proving the convergence of the elementary probabilities will lead to the aim.
Since $\beta_n(t) = \tilde{N}(t) = 0$ for all $t < 0$ and $n \in \mathbb{N}$, we can assume without loss of generality that $0 \le t_1$. The key is to see that

$$Z_n^{(m+1)} := \big(\beta_n(t_1), \beta_n(t_2) - \beta_n(t_1), \ldots, \beta_n(t_m) - \beta_n(t_{m-1}), n - \beta_n(t_m)\big)$$

has a $(m+1)$-dimensional multinomial distribution (see appendix, Definition A.8). The claim then follows by a variant of the Poisson limit theorem (see appendix, Lemma A.9). Since we are considering convergence for $n \to \infty$ we may as well assume that $t_m < n$. In this case the multinomial distribution of $Z_n^{(m+1)}$ has parameters n, $\frac{t_1}{n}$, $\frac{t_2 - t_1}{n}$, $\ldots$, $\frac{t_m - t_{m-1}}{n}$, $1 - \frac{t_m}{n}$. (If $t_j \le n < t_{j+1}$, then the parameters are n, $\frac{t_1}{n}$, $\frac{t_2-t_1}{n}$, $\ldots$, $\frac{t_j - t_{j-1}}{n}$, $1 - \frac{t_j}{n}, 0, \ldots, 0$.)
Let $0 \le k_1 \le \ldots \le k_m \le n$, $k_i \in \mathbb{N}$ for $i = 1, ..., m$. The random vectors $(\beta_n(t_1), ..., \beta_n(t_m))$ and $(\tilde{N}(t_1), ..., \tilde{N}(t_m))$ both admit only such vectors $(k_1, ..., k_m)$ with positive probability.

$$\begin{aligned}
& \mathbb{P}\Big((\beta_n(t_1), ..., \beta_n(t_m)) = (k_1, ..., k_m)\Big) \\
= \quad & \mathbb{P}\Big(Z_n^{(m+1)} = (k_1, k_2 - k_1, ..., k_m - k_{m-1}, n - k_m)\Big) \\
= \quad & \frac{n!}{k_1!(k_2-k_1)! \ldots (n-k_m)!} \left[\frac{t_1}{n}\right]^{k_1} \left[\frac{t_2 - t_1}{n}\right]^{k_2 - k_1} \cdots \left[\frac{t_m - t_{m-1}}{n}\right]^{k_m - k_{m-1}} \left[1 - \frac{t_m}{n}\right]^{n - k_m} \\
= \quad & \underbrace{\frac{n!}{(n-k_m)! n^{k_m}}}_{\longrightarrow 1} \cdot \underbrace{\frac{t_1^{k_1}}{k_1!} \frac{(t_2 - t_1)^{k_2 - k_1}}{(k_2 - k_1)!} \cdots \frac{(t_m - t_{m-1})^{k_m - k_{m-1}}}{(k_m - k_{m-1})!}}_{\text{constant with respect to } n} \cdot \underbrace{\left(1 - \frac{t_m}{n}\right)^{n - k_m}}_{\longrightarrow e^{-t_m}} \\
\longrightarrow \quad & \frac{t_1^{k_1}}{k_1!} \frac{(t_2 - t_1)^{k_2 - k_1}}{(k_2 - k_1)!} \cdots \frac{(t_m - t_{m-1})^{k_m - k_{m-1}}}{(k_m - k_{m-1})!} \cdot e^{-t_m} \\
= \quad & \prod_{i=1}^{m} \frac{(t_i - t_{i-1})^{k_i - k_{i-1}}}{(k_i - k_{i-1})!} e^{-t_i - t_{i-1}} \qquad (\text{with } t_0 := k_0 := 0) \\
= \quad & \mathbb{P}\Big((\tilde{N}(t_1), ..., \tilde{N}(t_m)) = (k_1, ..., k_m)\Big)
\end{aligned}$$

■

Lemma 4.19 *For $r \leq s \leq t$, $s, r, t \in \mathbb{R}$ holds*

$$\mathbb{E}|\beta_n(s) - \beta_n(r)||\beta_n(t) - \beta_n(s)| \leq (s-r)(t-s).$$

Proof. For the time being, assume r, s and t to lie in $[0, n]$. Recall

$$\beta_n(t) = \sum_{k=1}^{n} \mathbb{1}_{\{U_k \leq \frac{t}{n}\}},$$

thus

$$\beta_n(t) - \beta_n(s) = \sum_{k=1}^{n} \mathbb{1}_{(\frac{s}{n}, \frac{t}{n}]}(U_k).$$

Note that this random variables is a.s. non-negative, so we can drop the absolute values

$$\mathbb{E}\Big[\beta_n(s) - \beta_n(r)\Big]\Big[\beta_n(t) - \beta_n(s)\Big] = \mathbb{E}\Big[\sum_{k=1}^{n} \mathbb{1}_{(\frac{r}{n}, \frac{s}{n}]}(U_k)\Big]\Big[\sum_{l=1}^{n} \mathbb{1}_{(\frac{s}{n}, \frac{t}{n}]}(U_l)\Big]$$
$$= \sum_{k=1}^{n} \mathbb{E}\Big[\mathbb{1}_{(\frac{r}{n}, \frac{s}{n}]}(U_k)\mathbb{1}_{(\frac{s}{n}, \frac{t}{n}]}(U_k)\Big] + \sum_{k \neq l} \mathbb{E}\Big[\mathbb{1}_{(\frac{r}{n}, \frac{s}{n}]}(U_k)\mathbb{1}_{(\frac{s}{n}, \frac{t}{n}]}(U_l)\Big].$$

Each summand in the left sum is zero since U_k cannot lie in two disjoint intervalls at once. Due to the independence of U_k and U_l ($k \neq l$), each summand in the right sum equals $n^{-2}(s-r)(t-s)$, hence

$$\mathbb{E}\Big[\beta_n(s) - \beta_n(r)\Big]\Big[\beta_n(t) - \beta_n(s)\Big] = n(n-1)\Big(\frac{s-r}{n}\Big)\Big(\frac{t-s}{n}\Big) \leq (s-r)(t-s). \tag{54}$$

Now we drop the restriction from the beginning. For arbitrary real numbers $s \leq t$ holds

$$\mathbb{E}\mathbb{1}_{(\frac{s}{n}, \frac{t}{n}]}(U_k) = \begin{cases} 0 & \text{if } n \leq s, \\ 1 - \frac{s}{n} & \text{if } s < n \leq t, \\ \frac{t-s}{n} & \text{if } 0 \leq s \leq t < n, \\ 1 & \text{if } s < 0, n \leq t, \\ \frac{t}{n} & \text{if } s < 0 \leq t, \\ 0 & \text{if } t < 0, \end{cases}$$

so in any case $\mathbb{E}\mathbb{1}_{(\frac{s}{n}, \frac{t}{n}]}(U_k) \leq \frac{t-s}{n}$, and the inequality (54) holds as well.

■

Proof of Theorem 4.17. Apply Theorem 4.9. Lemma 4.18 implies 4.9 (1), and lemma 4.19 yields 4.9 (2).

■

With view of the generalisation of β_n that we are going to deal with in section 5.2, one might ask, why we do not treat the general case right away (which would plainly motivate the use of $D(-\infty, \infty)$ instead of $D[0, \infty)$). I would like to give two reasons for my choice: First, the special case is easier to overlook. The main idea is better to get at, since it is less blurred by technical details. Second, the content of Lemma 4.18 will neither be used again later nor refered to, and speaking of technical details, the proof would be unnecessarily lengthy in the general case.

5 Weak Convergence in the Skorokhod Product Space

We give a criterion for weak convergence in $D(-\infty,\infty) \times D(-\infty,\infty)$ of the same type as Theorems 3.8 and 4.8, that is convergence of the finite-dimensional distributions and tightness. We will see that tightness of both "component sequences" already suffices (cf. Lemma 5.2).
Take the product space

$$E := D_\infty \times D_\infty,$$

naturally equipped with the product topology, which is induced, for example, by the metric

$$e\big((x_1,x_2),(y_1,y_2)\big) := d_\infty(x_1,y_1) \vee d_\infty(x_2,y_2).$$

(E,e) is a separable metric space.
The elements of E are pairs of functions $\mathbb{R} \to \mathbb{R}$. Often one wishes to interpret the elements of E as mappings themselves. For example, they can be viewed as functions $\mathbb{R}^2 \to \mathbb{R}^2$, whose first component depends only on the first argument, and the second component only on the second argument, i.e. the pair (x,y), $x,y \in D_\infty$, is assigned to the mapping

$$(s,t) \mapsto (x(s),y(t)), \qquad s,t \in \mathbb{R}. \tag{55}$$

However, the standard - since more convenient - interpretation of elements of E is as functions from $\mathbb{R}$ to $\mathbb{R}^2$, i.e. (x,y) gets assigned to the mapping

$$t \mapsto (x(t),y(t)), \qquad t \in \mathbb{R}. \tag{56}$$

One usually identifies the Cartesian product of function spaces as a space of functions mapping into the Cartesian product of the range spaces. This, of course, is only feasible if the functions have the same domain.
Note that (55) and (56) are fully compatible, knowing one of them means knowing the other, neither does "contain more information" than the other. It is best to stick to regarding the elements of E primarily as pairs of functions, and not exclusively think of interpretation (56). Confusion may arise from the following. Let $D_\infty(\mathbb{R}^2)$ be the space of all cadlag functions from $\mathbb{R}$ to $\mathbb{R}^2$. Now, if x and y are elements of D_∞, it is easy to verify that (x,y), interpreted as $t \mapsto (x(t),y(t))$, is an element of $D_\infty(\mathbb{R}^2)$. Vice versa, if $z : t \mapsto (z_1(t),z_2(t))$ is an element of $D_\infty(\mathbb{R}^2)$, then the components z_1 and z_2 are one-dimensional cadlag functions. This is simply due to the fact that

a vector converges (in $\mathbb{R}^k$) if and only if all its components do. Thus, we can identify $D_\infty \times D_\infty$ with $D_\infty(\mathbb{R}^2)$. On the latter one may, of course, define a Skorokhod topology as well (see e.g. [JS02], page 327), and the question arises, how does this Skorokhod topology relate to the product topology on $D_\infty \times D_\infty$. The answer is, it is strictly finer (cf. [JS02], page 329).

I am not going to describe the Skorokhod topology on $D_\infty(\mathbb{R}^2)$, but in order to exemplify the difference, I would like to resort to the spaces $D[0,1]\times D[0,1]$ (equipped with the product topology) and $D_{[0,1]}(\mathbb{R}^2)$ (equipped with the Skorokhod topology), respectively. This keeps things simpler but illustrates the point well enough. The Skorokhod topology on $D_{[0,1]}(\mathbb{R}^2)$ is, for example, given by the metric

$$d^{(2)}(z_1, z_2) := \inf_{\lambda\in\Lambda} \big\{ \sup_{t\in[0,1]} |\lambda(t)-t| \vee \sup_{t\in[0,1]} |z_1(\lambda(t))-z_2(t)|\big\}, \quad z_1, z_2 \in D_{[0,1]}(\mathbb{R}^2). \tag{57}$$

The only difference to (1) is that the second (and only the second) absolute value denotes the vector norm in $\mathbb{R}^2$. Suppose (x, y) and (x_n, y_n), $n \in \mathbb{N}$, are elements of $D[0,1] \times D[0,1]$. We know, (x_n, y_n) converges to (x, y) in $D[0,1] \times D[0,1]$ if and only if $x_n \to x$ and $y_n \to y$ in $D[0,1]$. According to Lemma A.1 this means, there exist sequences $\{\lambda_n^{(1)}\}$, $\{\lambda_n^{(2)}\} \subset \Lambda$ such that

$$\begin{cases} |\, x_n(t) - x(\lambda_n^{(1)}(t))| \to 0 \\ \lambda_n^{(1)}(t) \to t \end{cases} \qquad \text{uniformly in } t \in [0,1], \tag{58}$$

and

$$\begin{cases} |\, y_n(t) - y(\lambda_n^{(2)}(t))| \to 0 \\ \lambda_n^{(2)}(t) \to t \end{cases} \qquad \text{uniformly in } t \in [0,1]. \tag{59}$$

According to (57), (x_n, y_n) converges to (x, y) with respect to $d^{(2)}$ if and only if (cf. Lemma A.10) there exists a sequence $\{\lambda_n\}$ such that

$$\begin{cases} |\, x_n(t) - x(\lambda_n(t))| \to 0 \\ |\, y_n(t) - y(\lambda_n(t))| \to 0 \\ \lambda_n(t) \to t \end{cases} \qquad \text{uniformly in } t \in [0,1]. \tag{60}$$

This is evidently a stronger condition than (58) and (59) together. For (x_n, y_n) to converge in $D_{[0,1]}(\mathbb{R}^2)$ the sequence $\{\lambda_n\}$ of time deformations must be the same in every component. In order to show that condition (60) is indeed strictly stronger, take the following example:

$$x_n := \mathbb{1}_{[0,\frac{1}{2}-\frac{1}{n})}$$

$$y_n := \mathbb{1}_{[0,\frac{1}{2}+\frac{1}{n})}$$
$$x := y := \mathbb{1}_{[0,\frac{1}{2})}$$

From Lemma A.2 we know that (x_n, y_n) converges to (x, y) in $D[0,1] \times D[0,1]$, but it does not converge with respect to $d^{(2)}$ (Lemma A.11).
It becomes clear how to construct an analoguous example on the whole real line. Thus the topological space $D_\infty(\mathbb{R}^2)$ has less convergent sequences and more open sets than $D_\infty \times D_\infty$. Here, "less" and "more" are meant in the sense of a strict set inclusion. Moreover one can conclude, $D_\infty(\mathbb{R}^2)$ has less compact sets, but this does not have to be a strict inclusion.
However, we are not concerned with the Skorokhod space $D_\infty(\mathbb{R}^2)$. The space $D_{[0,\infty)}(\mathbb{R}^k)$ is extensively treated in [JS02]. The book [EK86] examines the even more general space $D_{[0,\infty)}(M)$, where M is an arbitrary metric space. [Whi02], section 12, sheds further light on the differences of the k-dimensional Skorokhod topology and the weaker product topology.
We deal with elements of D_∞, i.e. one-dimensional cadlag functions, and pairs of those. One goal is to prove stochastic independence, so the natural choice of the space we model things in must be product space. We use results from D_∞-theory and extend them by product space argumentations to $D_\infty \times D_\infty$.
Let

$$\mathscr{E} := \mathscr{D}_\infty \otimes \mathscr{D}_\infty = \sigma(\mathscr{D}_\infty \times \mathscr{D}_\infty).$$

Lemma 5.1 *$\mathscr{E}$ is the Borel-σ-field on E.*

Proof. cf. Proposition 2.8. Separability is needed.

□

5.1 A Criterion for Weak Convergence in the Space $D_\infty \times D_\infty$

Before we state and prove theorem 5.5 we introduce some more notation and present three small lemmas that will abbreviate the proof.
Recall the projection π_T from the previous section. Now we deal with projections of the following kind: Let $S = \{s_1, ..., s_k\}$, $T := \{t_1, ..., t_l\}$. Then

$$\pi_{S,T} : E \to \mathbb{R}^{|S|+|T|} : (x, y) \mapsto (\pi_S(x), \pi_T(y)) = \big((x(s_1), ..., x(s_k), y(t_1), ..., y(t_l)\big),$$

(61)

where x and y are elements of D_∞. In particular,

$$\pi_{S,\emptyset} : E \to \mathbb{R}^{|S|} : (x,y) \mapsto \pi_S(x) = \big(x(s_1), ..., x(s_k)\big),$$

and

$$\pi_{\emptyset,T} : E \to \mathbb{R}^{|T|} : (x,y) \mapsto \pi_T(y) = \big(y(t_1), ..., y(t_l)\big).$$

Let T_1, T_2 be arbitrary subsets of $\mathbb{R}$ and

$$\mathscr{F}(T_1,T_2) := \big\{\pi_{S,T}^{-1}(A)\big|S,T \text{ finite}, S \subset T_1, T \subset T_2, A \in \mathscr{B}(\mathbb{R}^{|S|+|T|})\big\}.$$

Now consider a probability measure P on $(E,\mathscr{E})$. We will often write $P = (P^{(1)}, P^{(2)})$. This shall mean, $P^{(1)}$ and $P^{(2)}$ are the marginal probabilities of P, each on $\mathscr{D}_\infty$, i.e.

$$P^{(1)}(A) = P(A \times D_\infty) \qquad \forall\, A \in \mathscr{D}_\infty,$$

$$P^{(2)}(B) = P(D_\infty \times B) \qquad \forall\, B \in \mathscr{D}_\infty.$$

Keep in mind that this defines $P^{(1)}$ and $P^{(2)}$ from P, but not vice versa.

Lemma 5.2 *Let $\{P_n = (P_n^{(1)}, P_n^{(2)})\}$ be a sequence of probability measures on $(E,\mathscr{E})$. If $\{P_n^{(1)}\}$ and $\{P_n^{(2)}\}$ both are tight in $(D_\infty, \mathscr{D}_\infty)$, then so is $\{P_n\}$.*

Proof. This is a corollary of Tikhonov's theorem. Let $\varepsilon > 0$. According to the assumptions there exists a compact set $K_1 \subset D_\infty$, such that

$$P_n^{(1)}(K_1) > 1 - \frac{\varepsilon}{2} \qquad \forall\, n \in \mathbb{N},$$

hence

$$P_n^{(1)}(D_\infty \setminus K_1) = P_n((D_\infty \setminus K_1) \times D_\infty) < \frac{\varepsilon}{2}.$$

Likewise, there exists a compact set $K_2 \in D_\infty$ such that

$$P_n^{(2)}(D_\infty \setminus K_2) = P_n(D_\infty \times (D_\infty \setminus K_2)) < \frac{\varepsilon}{2}.$$

By De Morgan's law:

$$\begin{aligned}
& \big[(D_\infty \setminus K_1) \times D_\infty\big] \cup \big[D_\infty \times (D_\infty \setminus K_2)\big] \\
= \;& (D_\infty \times D_\infty) \setminus (K_1 \times D_\infty) \cup (D_\infty \times D_\infty) \setminus (D_\infty \times K_2) \\
= \;& E \setminus \big[(K_1 \times D_\infty) \cap (D_\infty \times K_2)\big] \\
= \;& E \setminus (K_1 \times K_2)
\end{aligned}$$

Therefore:

$$\begin{aligned} P_n\big(E \setminus (K_1 \times K_2)\big) &= P_n\Big(\big[(D_\infty \setminus K_1) \times D_\infty\big] \cup \big[D_\infty \times (D_\infty \setminus K_2)\big]\Big) \\ &\leq P_n\Big((D_\infty \setminus K_1) \times D_\infty\Big) + P_n\Big(D_\infty \times (D_\infty \setminus K_2)\Big) < \varepsilon \end{aligned}$$

and

$$P_n(K_1 \times K_2) > 1 - \varepsilon.$$

By Tikhonov's theorem, $K := K_1 \times K_2$ is compact in the product space (E, e). Hence $\{P_n\}$ is tight.

■

Lemma 5.3 *Let $P = (P^{(1)}, P^{(2)})$ be a probability measure on $(E, \mathscr{E})$, and S,T finite subsets of $T_{P^{(1)}}$, $T_{P^{(2)}}$, respectively. Then $\pi_{S,T}$ is P-a. e. continuous.*

Proof. Let $A_1, A_2 \subset D_\infty$ be the discontinuity sets of π_S, π_T, respectively. Then $A_1 \times D_\infty$ and $D_\infty \times A_2$ are the discontinuity sets of $\pi_{S,\emptyset}$ and $\pi_{\emptyset,T}$. This can for example easily be seen by the convergence characterization of continuity.
From previous considerations we know that π_S is $P^{(1)}$-a.e. continuous, and that π_T is $P^{(2)}$-a.e. continuous (cf. proof of Theorem 3.8), i.e.

$$\begin{aligned} &P^{(1)}(A_1) = P(A_1 \times D_\infty) = 0, \\ &P^{(2)}(A_2) = P(D_\infty \times A_2) = 0. \end{aligned}$$

Note that

$$\pi_{S,T} = (\pi_{S,\emptyset}, \pi_{\emptyset,T}),$$

and keep in mind that a vector-valued function is continuous at a point, if every component is. Hence, the discontinuity set of $\pi_{S,T}$ is $(A_1 \times D_\infty) \cup (D_\infty \times A_2)$.

$$P\big([A_1 \times D_\infty] \cup [D_\infty \times A_2]\big) \leq P(A_1 \times D_\infty) + P(D_\infty \times A_2) = 0.$$

$\pi_{S,T}$ is Q-a.e. continuous.

■

Lemma 5.4 *Let $\mathscr{S}_1$ and $\mathscr{S}_2$ be arbitrary system of sets. If $\mathscr{S}_1$ and $\mathscr{S}_2$ both are π-systems, their Cartesian product $\mathscr{T} := \mathscr{S}_1 \times \mathscr{S}_2$ is a π-system, too.*

Proof. Let $A, B \in \mathscr{T}$. Then there exist sets $A_1, B_1 \in \mathscr{S}_1$ and $A_2, B_2 \in \mathscr{S}_2$ such that $A = A_1 \times A_2$ and $B = B_1 \times B_2$. Because of the assumption:

$$C := (A_1 \cap B_1) \times (A_2 \cap B_2) \in \mathscr{T}$$

The claim reads:

$$D := (A_1 \times A_2) \cap (B_1 \times B_2) \in \mathscr{T},$$

which becomes evident when noting that the sets C anc D both equal

$$\big\{(x, y)\big| x \in A_1, x \in B_1, y \in B_1, y \in B_2\big\}.$$

∎

Theorem 5.5 *Let $\{P_n = (P_n^{(1)}, P_n^{(2)})\}$ be a sequence of probability measures on $(E, \mathscr{E})$ for which holds:*

(1) $\{P_n^{(1)}\}$ and $\{P_n^{(2)}\}$ are both tight.

(2) There exists a measure $P = (P^{(1)}, P^{(2)})$ on $(E, \mathscr{E})$ such that

$$P_n \circ \pi_{S,T}^{-1} \xrightarrow{\mathscr{L}} P \circ \pi_{S,T}^{-1}, \tag{62}$$

for all finite S and T, $S \subset T_{P^{(1)}}$ and $T \subset T_{P^{(2)}}$.

Then $P_n \xrightarrow{\mathscr{L}} P$ in $(E, \mathscr{E})$.

Before we prove Theorem 5.5, we rephrase it in terms of random variables. Mainly we do so in order to have a more descriptive formulation of condition (62).

Corollary 5.6 *Let $\{Z_n = (X_n, Y_n)\}$ be a sequence of random variables in $(E, \mathscr{E})$. If*

(1) the sequences $\{X_n\}$ and $\{Y_n\}$ are tight, and

(2) there is a random variable $Z = (X, Y)$ in $(E, \mathscr{E})$ such that

$$\big((X_n(s_1), ..., X_n(s_k), Y_n(t_1), ..., Y_n(t_l)\big) \xrightarrow{\mathscr{L}} \big((X(s_1), ..., X(s_k), Y(t_1), ..., Y(t_l)\big) \tag{63}$$

for all $k, l \in \mathbb{N}$, $s_1, ..., s_k \in T_X$, and $t_1, ..., t_l \in T_Y$,

then $Z_n \xrightarrow{\mathscr{L}} Z$.

Proof of Theorem 5.5. The proof follows the same pattern as the one of Theorem 3.8. Quite a few details require special attention, most of which we have already accounted for by Lemmas 5.1 5.2, 5.3 to 5.4.
From Lemma 5.2 we know that $\{P_n\}$ is tight, and due to Prokhorov's theorem it is relatively compact. Here it is important that $\mathscr{E}$ is the Borel-σ-field on E (Lemma 5.1). Thus we have a subsequence $\{P_n'\}$ and measure $Q = (Q^{(1)}, Q^{(2)})$ on $(E, \mathscr{E})$ such that P_n' weakly converges to Q. If S and T are finite subsets of $T_{Q^{(1)}}$ and $T_{Q^{(2)}}$, respectively, then $\pi_{S,T}$ is Q-a.e. continuous (Lemma 5.3), hence

$$P_n' \circ \pi_{S,T}^{-1} \xrightarrow{\mathscr{L}} Q \circ \pi_{S,T}^{-1}, \qquad\qquad S,\, T \text{ finite, } S \subset T_{Q^{(1)}}, T \subset T_{Q^{(2)}}.$$

On the other hand, (62) implies

$$P_n' \circ \pi_{S,T}^{-1} \xrightarrow{\mathscr{L}} P \circ \pi_{S,T}^{-1}, \qquad\qquad S,\, T \text{ finite, } S \subset T_{P^{(1)}}, T \subset T_{P^{(2)}}.$$

Therefore, if $T_1 := T_{P^{(1)}} \cap T_{Q^{(1)}}$, and $T_2 := T_{P^{(2)}} \cap T_{Q^{(2)}}$,

$$Q \circ \pi_{S,T}^{-1} = P \circ \pi_{S,T}^{-1} \qquad\qquad \forall\, S,\, T \text{ finite, } S \subset T_1, T \subset T_2.$$

This means, P and Q agree upon the class $\mathscr{F}(T_1, T_2)$. We are left to show that $\mathscr{F}(T_1, T_2)$ is a separating class for $\mathscr{E}$. First of all, since $T_{P^{(1)}}$, $T_{Q^{(1)}}$, $T_{P^{(2)}}$ and $T_{Q^{(2)}}$ have at most countable complements in $\mathbb{R}$ (Proposition 4.7), the intersections T_1 and T_2 are dense in $\mathbb{R}$. Hence $\mathscr{F}(T_1)$ and $\mathscr{F}(T_2)$ are both π-systems (Lemma 4.5) and generate both D_∞ (Proposition 4.4). By Lemma 5.4 the product

$$\mathscr{F}(T_1) \times \mathscr{F}(T_2) =$$

$$\left\{\pi_S^{-1}(A) \times \pi_T^{-1}(B) \middle| S, T \text{ finite } S \subset T_1, T \subset T_2, A \subset \mathscr{B}(\mathbb{R}^{|S|}), B \subset \mathscr{B}(\mathbb{R}^{|T|})\right\}$$

is a π-system, and it also generates $\mathscr{E}$ (Theorem 2.7). Therefore it is a seperating class for $\mathscr{E}$. Finally,

$$\mathscr{F}(T_1, T_2) \supset \mathscr{F}(T_1) \times \mathscr{F}(T_2),$$

which implies, that $\mathscr{F}(T_1, T_2)$ is a separating class, too.

■

As the last proposition in this subsection we state and prove a slightly simpler version of Corollary 5.6. In order to do so, define two further classes of sets. For $T_0 \subset \mathbb{R}$, let

$$\mathscr{G}(T_0) := \left\{\pi_T^{-1}(A) \times \pi_T^{-1}(B) \middle| T \subset T_0, |T| < \infty,\ A, B \in \mathscr{B}(\mathbb{R}^{|T|})\right\} \quad (64)$$

and

$$\mathscr{H}(T_0) := \left\{\pi_{T,T}^{-1}(C) \,\middle|\, T \subset T_0, |T| < \infty,\ C \in \mathscr{B}(\mathbb{R}^{2|T|})\right\}. \tag{65}$$

$\pi_T : D_\infty \to \mathbb{R}^{|T|}$ and $\pi_{S,T} : E \to \mathbb{R}^{2|T|}$ are the projections defined in subsection 4.2 and by (61), respectively. Apparently

$$\mathscr{G}(T_0) \subset \mathscr{H}(T_0). \tag{66}$$

Moreover, we have the following result.

Lemma 5.7 *$\mathscr{F}(T_0) \times \mathscr{F}(T_0) = \mathscr{G}(T_0)$.*

Proof. The inclusion "$\supset$" is evident, we are left to show $\mathscr{F}(T_0) \times \mathscr{F}(T_0) \subset \mathscr{G}(T_0)$. Let T and S be finite subsets of T_0, and $A \in \mathscr{B}(R^{|T|})$, $B \in \mathscr{B}(R^{|S|})$. Reproduce the construction of the sets U, $A^{(m)}$ and $B^{(m)}$ from the proof of Lemma 3.6. $U = \{u_1, ..., u_m\}$ is the union of T and S and thus contained in T_0. Then

$$\pi_T^{-1}(A) \times \pi_S^{-1}(B) = \pi_U^{-1}(A^{(m)}) \times \pi_U^{-1}(B^{(m)}).$$

Any set in $\mathscr{F}(T_0) \times \mathscr{F}(T_0)$ can be written as the left-hand side. The right-hand side is an element of $\mathscr{G}(T_0)$.

■

One can also show $\mathscr{F}(T_0, T_0) = \mathscr{H}(T_0)$, but Lemma 5.7 suffices to prove the following proposition.

Theorem 5.8 *Let $\{Z_n = (X_n, Y_n)\}$ be a sequence of random variables in $(E, \mathscr{E})$. If*

(1) the sequences $\{X_n\}$ and $\{Y_n\}$ are tight, and

(2) there is a random variable $Z = (X, Y)$ in $(E, \mathscr{E})$ such that

$$\big((X_n(t_1), ..., X_n(t_k), Y_n(t_1), ..., Y_n(t_k)\big) \xrightarrow{\mathscr{L}} \big((X(t_1), ..., X(t_k), Y(t_1), ..., Y(t_k)\big) \tag{67}$$

for all $k \in \mathbb{N}$, $t_1, ..., t_k \in T_X \cap T_Y$,

then $Z_n \xrightarrow{\mathscr{L}} Z$.

Proof. We use the notation from Theorem 5.5, i.e. $P_n = (P_n^{(1)}, P_n^{(2)})$ and $P = (P^{(1)}, P^{(2)})$ are the distributions of $Z_n = (X_n, Y_n)$ and $Z = (X, Y)$, respectively. Condition 5.8 (2) is the same as the weak convergence of the corresponding distributions, i.e.

$$P_n \circ \pi_{T,T}^{-1} \xrightarrow{\mathscr{L}} P \circ \pi_{T,T}^{-1}, \qquad \forall \text{ finite } T \subset T_{P^{(1)}} \cap T_{P^{(2)}}.$$

Just like in Theorem 5.5 we can conclude that there exist a measure Q and a subsequence $\{P_n'\}$ such that

$$P_n' \circ \pi_{T,T}^{-1} \xrightarrow{\mathscr{L}} Q \circ \pi_{T,T}^{-1}, \qquad \forall \text{ finite } T \subset T_{Q^{(1)}} \cap T_{Q^{(2)}},$$

hence

$$P \circ \pi_{T,T}^{-1} = Q \circ \pi_{T,T}^{-1} \qquad \forall \text{ finite } T \subset T_{Q^{(1)}} \cap T_{Q^{(2)}} \cap T_{P^{(1)}} \cap T_{P^{(2)}}. \tag{68}$$

Call $T_0 := T_{Q^{(1)}} \cap T_{Q^{(2)}} \cap T_{P^{(1)}} \cap T_{P^{(2)}}$. Then (68) implies that P and Q agree on the class

$$\mathscr{H}(T_0) = \left\{\pi_{T,T}^{-1}(C) \,\middle|\, T \text{ finite}, T \subset T_0, C \in \mathscr{B}(\mathbb{R}^{2|T|})\right\}.$$

Keep in mind that $\pi_{T,T} : (x, y) \mapsto (\pi_T(x), \pi_T(y))$ is a function from $D_\infty \times D_\infty$ into $\mathbb{R}^{2|T|}$, and not into $\mathbb{R}^{|T|}$. In particular, (68) does *not* (only) imply that P and Q agree on the (much smaller) set

$$\left\{\pi_T^{-1}(A) \times \pi_T^{-1}(A) \,\middle|\, T \text{ finite}, T \subset T_0, A \in \mathscr{B}(\mathbb{R}^{|T|})\right\}.$$

From Lemma 5.7 and (66) we know

$$\mathscr{H}(T_0) \supset \mathscr{F}(T_0) \times \mathscr{F}(T_0).$$

We are left to show that $\mathscr{F}(T_0) \times \mathscr{F}(T_0)$ is a separating class for $\mathscr{E}$, which works just the same as to show the corresponding claim for $\mathscr{F}(T_1) \times \mathscr{F}(T_2)$, as done in the proof of 5.5. The key is: T_0 is still a dense set in $\mathbb{R}$. Apply (in this order) Proposition 4.4, Proposition 2.7, Lemma 4.5, Lemma 5.4 and Theorem 2.3.

∎

The question arises why we did not put Lemma 5.7 first, proved Corollary 5.8 and skipped 5.5 and 5.6 altogether. The answer is, to emphasize that condition 5.8 (2) is very little weaker than, and in the context of the proof equally strong as 5.6 (2) - a fact that may seem counterintuitive at first, and which took me some time to realize.

Moreover, 5.8 (2) is usually not harder to show than 5.6 (2). For example, look at the proof of Proposition 5.10: Replacing $s_1, ..., s_p$ by $t_1, ..., t_m$ would mean a change in notation only.

Another way of deducting Theorem 5.8, instead of introducing the sets $\mathscr{G}(T_0)$ and $\mathscr{H}(T_0)$, is as follows. One can show, that, if a random vector in $\mathbb{R}^k$ converges in law, then every sub-vector (i.e. any vector consisting of a subset of the components) does as well. The fastest way is probably by the CMT. Then,

$$\pi_{S,T}(X_n, Y_n) \xrightarrow{\mathscr{L}} \pi_{S,T}(X, Y) \qquad \forall\ S, T \text{ finite, } S, T \subset T_X \cap T_Y$$

follows right away from

$$\pi_{T\cup S,T\cup S}(X_n, Y_n) \xrightarrow{\mathscr{L}} \pi_{T\cup S,T\cup S}(X, Y) \qquad \forall\ S, T \text{ finite, } S, T \subset T_X \cap T_Y.$$

From there one can conclude that (with the notation of the proof above) P and Q agree on the class $\mathscr{F}(T_0, T_0)$, which is a superset of $\mathscr{F}(T_0) \times \mathscr{F}(T_0)$, and then continue as in the proof above.
This argumentation makes it also immediately clear that 5.8 (2) and 5.6 (2) are equivalent if $T_X = T_Y$. But note, that, if $T_X \neq T_Y$, 5.8 (2) is technically a stronger condition than 5.6 (2), and thus 5.8 is not a direct corollary of 5.6.

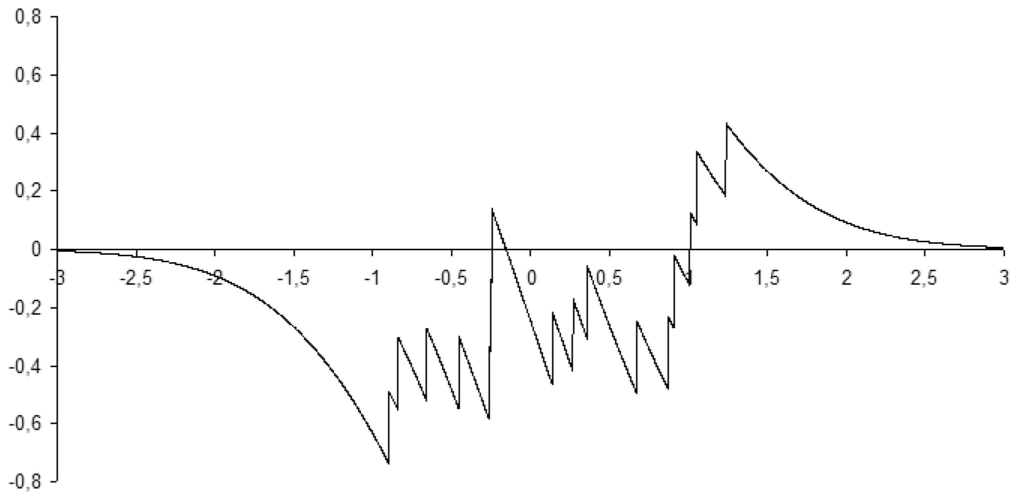
5.2 The Limit of the Common Distribution of the Empirical Process and the Rescaled Empirical Distribution Function

We now come back to the initial problem. We apply Corollary 5.6 to (α_n, β_n) and by doing so prove the asymptotical independence of α_n and β_n. So far we have just considered the *uniform* empirical process and the rescaled *uniform* empiricial distribution function. Now we drop this restriction and re-define α_n and β_n, allowing the random variables they were derived from to have an arbitrary distribution function F. Moreover, the time point τ around which β_n is stretched may as well be arbitrary. So far, τ has been zero. We will write α_n^F and $\beta_n^{F,\tau}$ in order to indicate the difference to the uniform processes we dealt with in sections 3.4 and 4.4, respectively. We are going to prove, that, for continuous F, $(\alpha_n^F, \beta_n^{F,\tau})$ converges to a random variable (B_1, N_0), where B_1 and N_0 are two independent random elements in D_∞ yet to be specified. This will be the content of Theorem 5.9.

Let F be an arbitrary distribution function, and $\{X_n\}$ a sequence of i.i.d. random variables, being distributed according to F. Define $\alpha_n^F = \{\alpha_n^F(t) | \, t \in \mathbb{R}\}$, $n \in \mathbb{N}$, to be the following random element in D_∞:

$$\alpha_n^F(t) := \sqrt{n}\big(F_n(t) - F(t)\big) = \frac{1}{\sqrt{n}} \sum_{k=1}^{n} \big(\mathbb{1}_{\{X_k \le t\}} - F(t)\big), \qquad t \in \mathbb{R}. \tag{69}$$

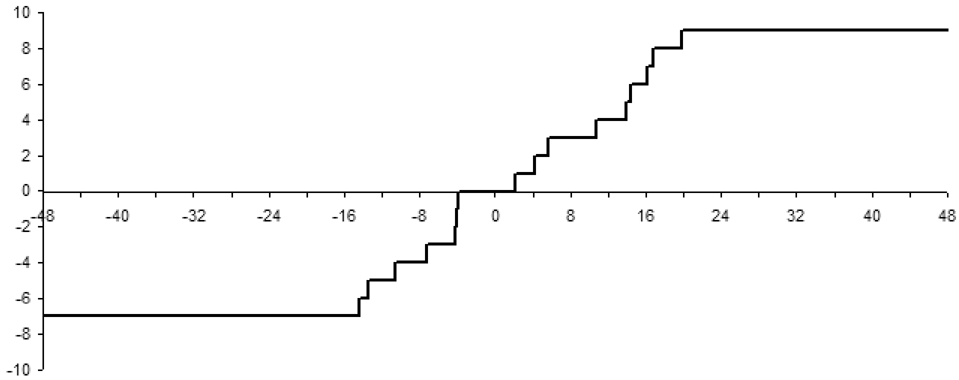
Below is picture of a realisation of α_{16}^F, where F is the standard normal distribution function.



Let furthermore $\tau \in \mathbb{R}$ be arbitrary, and $\beta_n^{F,\tau} = \{\beta_n^{F,\tau}(t) |\, t \in \mathbb{R}\}$, $n \in \mathbb{N}$, defined as

$$\begin{aligned}\beta_n^{F,\tau}(t) \;&:=\; n\big[F_n(\tau + \frac{t}{n}) - F_n(\tau)\big], \qquad t \in \mathbb{R}, \\ &= \begin{cases} \sum_{k=1}^{n} \mathbb{1}_{(\tau,\tau+\frac{t}{n}]}(X_k), & \text{if } t \geq 0, \\ \sum_{k=1}^{n} -\mathbb{1}_{(\tau+\frac{t}{n},\tau]}(X_k), & \text{if } t < 0. \end{cases}\end{aligned} \tag{70}$$

Here we have a picture of $\beta_{16}^{F,\tau}$, with $\tau = 0$ and the same F as above, also formed by the same sample as the graph above.



Then, let $\gamma_n := \gamma_n^{F,\tau} := (\alpha_n^F, \beta_n^{F,\tau})$. This is a function of $X_1, ..., X_n$ and a random element in $(E, \mathscr{E})$. Its distribution is determined by F and τ.

We assume from now on that the following two conditions hold.

Condition C.1 F is continuous.

Condition C.2 F has both, left- and right-hand side derivatives in τ. Call the former ϱ_1 and the latter ϱ_2, i.e.

$$\varrho_1 := \lim_{h \nearrow 0} \frac{F(\tau + h) - F(\tau)}{h} \tag{71}$$

and

$$\varrho_2 := \lim_{h \searrow 0} \frac{F(\tau + h) - F(\tau)}{h}. \tag{72}$$

Here are two examples where Condition C.2 is not satisfied. In both cases take $\tau = 0$.

(I) F is the distribution function of a χ^2_1 distribution, i.e. a Chi-sqared distribution with one degree of freedom. The right-hand side derivative at 0 is $+\infty$, F is continuous.

(II) Take F to be the following:

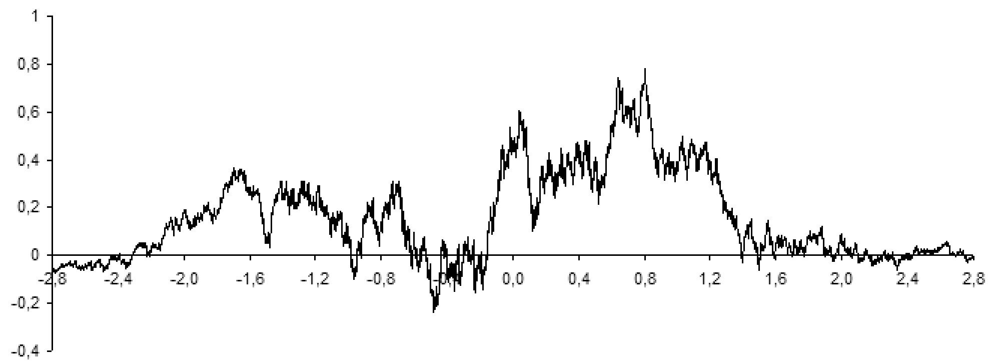
$$F(t) := \begin{cases} 1, & t \geq 1, \\ (\frac{1}{2})^n, & (\frac{1}{2})^n \leq t < (\frac{1}{2})^{n-1}, \qquad n \geq 1 \\ 0, & t \leq 0, \end{cases}$$

F is continuous at 0, but the right-hand side derivative does not exist.

We still need to specify the potential limit process of γ_n. Let $B_0 = \{B_0(t) | \, t \in [0,1]\}$ be a Brownian bridge. For any, not necessarily continuous distribution function F, the process $B_1 = B_1^F = \{B_1(t) | \, t \in \mathbb{R}\}$, definded by

$$B_1(t) := B_0(F(t)), \qquad t \in \mathbb{R}, \tag{73}$$

has also trajectories in D_∞. B_0 has continuous paths, and F is an element of D_∞, it is then easy to verify that B_1 has cadlag paths, too. We also want to allow the notation $B_1^F = B_0 \circ F$, meaning an ω-wise definiton. Below is a picture of a typical trajectory of B_1^F, where F is the standard normal distribution function.



Furthermore, let N_1, N_2 be two Poisson processes with the following properties:

- N_1 and N_2 are independent of B_1, and of each other.
- N_i has rate ϱ_i, $i = 1, 2$.

- N_2 has, as usual, right-continuous trajectories while those of N_1 are left-continuous, i.e. the value at a jump point is always set to the left-hand limit. Note that this leaves the finite-dimensional distributions unchanged.

Then, $\tilde{N}_0 = \tilde{N}_0(\varrho_1, \varrho_2) = \{\tilde{N}_0(t) | \, t \in \mathbb{R}\}$, defined by

$$\tilde{N}_0(t) := \begin{cases} -N_1(-t), & t < 0, \\ N_2(t), & t \geq 0, \end{cases}$$

is a random variable in D_∞. In analogy to $\beta_n \to \tilde{N}$ (cf. Theorem 4.17), we have: If Conditions C.1 and C.2 hold, then $\beta_n^{F,\tau} \to \tilde{N}_0$ in D_∞ (cf. [Fer03b]). Finally, let $M := M^{F,\tau} := (B_1, \tilde{N}_0)$. Now here is the paper's main result:

Theorem 5.9 *If Conditions C.1 and C.2 are fulfilled, then $\gamma_n^{F,\tau}$ converges in distribution to M in $(E, \mathscr{E})$.*

Proof. According to Corollary 5.6, three things need to be verified:

(a) $\{\alpha_n^F\}$ is tight.

(b) $\{\beta_n^{F,\tau}\}$ is tight.

(c) For all finite sets $\{t_1, ..., t_m\} \in T_{B_1}$ and $\{s_1, ..., s_p\} \in T_{\tilde{N}_0}$, it holds

$$\begin{aligned} &\Big(\alpha_n^F(t_1), ..., \alpha_n^F(t_m), \beta_n^{F,\tau}(s_1), ..., \beta_n^{F,\tau}(s_p)\Big) \\ &\xrightarrow{\mathscr{L}} \Big(B_1(t_1), ..., B_1(t_m), \tilde{N}_0(s_1), ..., \tilde{N}_0(s_p)\Big). \end{aligned} \tag{74}$$

Part (c) is the content of Proposition 5.10. We do not need to bother what T_{B_1} and $T_{\tilde{N}_0}$ look like, since (74) holds for *all* collections $\{t_1, ..., t_m\}$ and $\{s_1, ..., s_p\}$.
By re-enacting the proofs of Lemmas 3.11 and 4.19 we find that for all $r, s, t \in \mathbb{R}$, $r < s < t$, and $n \in \mathbb{N}$,

$$\begin{aligned} &\mathbb{E}|\alpha_n^F(s) - \alpha_n^F(r)|^2 |\alpha_n^F(t) - \alpha_n^F(s)|^2 \\ &\leq 6(F(r) - F(s))(F(t) - F(s)) \leq 6(F(t) - F(r))^2 \end{aligned}$$

and

$$\mathbb{E}|\beta_n^{F,\tau}(s) - \beta_n^{F,\tau}(r)||\beta_n^{F,\tau}(t) - \beta_n^{F,\tau}(s)|$$

$$\leq (F(r) - F(s))(F(t) - F(s)) \leq (F(t) - F(r))^2.$$

Hence, $\{\alpha_n^F\}$ and $\{\beta_n^{F,\tau}\}$ both satisfy 4.9 (2). Here is where we need Condition C.1, the continuity of F.
Furthermore, the finite-dimensional distributions of both sequences converge. This follows as a special case from Proposition 5.10, or it can be deducted analoguously to Lemmas 3.10 and 4.18, respectively. Thus $\{\alpha_n^F\}$ and $\{\beta_n^{F,\tau}\}$ satisfy 4.9 (1) as well, and so by Theorem 4.9 they both are tight and converge in distribution.

■

Proposition 5.10 *Let $\tau \in \mathbb{R}$ and $F : \mathbb{R} \to [0, 1]$ be an arbitrary distribution function. If Condition C.2 is satisfied, then (74) holds for all $t_1 \leq ... \leq t_m$ and $s_1 \leq ... \leq s_p$.*

It is important to notice that it does not suffice to show the convergence in law of the finite-dimensional distributions of α_n^F and $\beta_n^{F,\tau}$ separately. Here is the point where the independence of the limiting processes - and the dependence of the two converging sequences - come into play.
When we proved the convergence of the finite-dimensionals for α_n (Lemma 3.10), we applied the Central limit theorem, while for β_n (Lemma 4.18) we showed the convergence of the probablilities of the elementary events using a Poisson-limit type result. Obviously, neither approach can be succesfull here.The vector $(B_1(t_1), ..., B_1(t_m), \tilde{N}_0(s_1), ..., \tilde{N}_0(s_p))$ has a "mixed distribution", i.e. some components have discrete, others have continuous (marginal) distributions.
We prove Proposition 5.10 by showing the convergence of corresponding characteristic functions (cf. Theorems 2.16 and 2.17). There are two important principles when it comes to computing characteristic functions (see e.g. [Bil95] or [Mül91]).

(I) The characteristic function of a random vector with independent components is the product of the characteristic functions of each component.

(II) The characteristic function of the sum of independent random variables is also the product of the individual characteristic functions, but all being the function of one single variable.

$\tilde{N}_0$ is a process with independent increments, so its finite-dimensional distributions can be written as the linear transform of independent random variables. B_1 is linked to the Brownian motion, a process with independent

increments, so its finite-dimensionals can as well be expressed as a linear function of independent random variables. Hence, (I) allows us to compute the characteristic function of $(B_1(t_1), ..., B_1(t_m), \tilde{N}_0(s_1), ..., \tilde{N}_0(s_p))$.
As for $(\alpha_n^F(t_1), ..., \alpha_n^F(t_m), \beta_n^{F,\tau}(s_1), ..., \beta_n^{F,\tau}(s_p))$, we basically have to deal with sums of mutually independent indicator functions. Here we can make use of (II).
Now let us turn to the proof itself. But before we do so we have to declare some notation. In order to make the whole calculation a little bit easier to follow, we re-label the time points. Let

$$Y_n := \big(\alpha_n(t_1), ..., \alpha_n(t_m), \beta_n(r_1), ..., \beta_n(r_q), \beta_n(s_1), ..., \beta_n(s_p)\big), \qquad n \in \mathbb{N},$$

and

$$Y := \big(B_1(t_1), ..., B_1(t_m), \tilde{N}_0(r_1), ..., \tilde{N}_0(r_q), \tilde{N}_0(s_1), ..., \tilde{N}_0(s_p)\big),$$

where $t_1 < ... < t_m$ and $r_1 < ... < r_q < 0 \leq s_1 < ... < s_p$. Furthermore we introduce the following short-hand notation:

$$\begin{aligned}
x &:= (x_1, ..., x_m)^T \in \mathbb{R}^m, \\
y &:= (y_1, ..., y_q)^T \in \mathbb{R}^q, \\
z &:= (z_1, ..., z_p)^T \in \mathbb{R}^p, \\
r_{q+1} &:= s_0 := 0, \\
f_k &:= F(t_k) \qquad (k = 1, ..., m), \qquad\qquad f_{m+1} := 1, \qquad f_0 := 0, \\
\lambda &:= \sum_{j=1}^{m} -f_j\, x_j, \\
\mu_k &:= \sum_{j=k}^{m} x_j \qquad (k = 1, ..., m), \qquad\qquad \mu_{m+1} := 0, \\
\xi_k &:= \sum_{j=1}^{k} -y_j \qquad (k = 1, ..., q), \qquad\qquad \xi_0 := 0, \\
\nu_k &:= \sum_{j=k}^{p} z_j \qquad (k = 1, ..., p), \qquad\qquad \nu_{p+1} := 0.
\end{aligned}$$

Finally, set t_0 to $-\infty$ and t_{m+1} to $+\infty$. The characteristic function of any random variable X shall be denoted by φ_X.
The main work is to compute the characteristic functions of Y and Y_n, which is the content of the next two lemmas.

Lemma 5.11 *The characteristic function* $\psi : \mathbb{R}^{m+q+p} \to \mathbb{C}$ *of* Y *is*

$$\begin{aligned} \psi(x,z,y) &= \exp\Bigg\{ -\frac{1}{2}\sum_{k=1}^{m+1}(\mu_k+\lambda)^2(f_k-f_{k-1}) \\ &\quad +\sum_{k=1}^{q}\varrho_1(e^{i\xi_k}-1)(r_{k+1}-r_k)+\sum_{k=1}^{p}\varrho_2(e^{i\nu_k}-1)(s_k-s_{k-1})\Bigg\}. \end{aligned}$$

Proof. In the first step we compute the characteristic function of $(B_1(t_1), ..., B_1(t_m))$ According to (17) and (73), there exists a Brownian motion B on $[0,1]$ such that

$$
\begin{aligned}
\begin{pmatrix} B_1(t_1) \\ \vdots \\ B_1(t_m) \end{pmatrix} &= \begin{pmatrix} B_0(F(t_1)) \\ \vdots \\ B_0(F(t_m)) \end{pmatrix} = \begin{pmatrix} B_0(f_1) \\ \vdots \\ B_0(f_m) \end{pmatrix} \\
&= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & -f_1 \\ 0 & 1 & & 0 & 0 & -f_2 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & 0 & & 1 & 0 & -f_{m-1} \\ 0 & 0 & \dots & 0 & 1 & -f_m \end{pmatrix} \begin{pmatrix} B(f_1) \\ \vdots \\ B(f_m) \\ B(1) \end{pmatrix}
\end{aligned}
$$

$$
= \begin{pmatrix} 1 & 0 & \dots & 0 & 0 & -f_1 \\ 0 & 1 & & 0 & 0 & -f_2 \\ \vdots & & \ddots & & \vdots & \vdots \\ 0 & 0 & & 1 & 0 & -f_{m-1} \\ 0 & 0 & \dots & 0 & 1 & -f_m \end{pmatrix} \begin{pmatrix} 1 & 0 & \dots & 0 & 0 \\ 1 & 1 & & 0 & 0 \\ \vdots & \vdots & \ddots & & \vdots \\ 1 & 1 & \dots & 1 & 0 \\ 1 & 1 & \dots & 1 & 1 \end{pmatrix} \begin{pmatrix} B(f_1) \\ B(f_2) - B(f_1) \\ \vdots \\ B(f_m) - B(f_{m-1}) \\ B(1) - B(f_m) \end{pmatrix}.
$$

The first matrix is $m \times (m+1)$, the second $(m+1) \times (m+1)$. Their product is the $m \times (m+1)$-matrix

$$
H := \begin{pmatrix} 1-f_1 & -f_1 & -f_1 & \dots & -f_1 & -f_1 \\ 1-f_2 & 1-f_2 & -f_2 & \dots & -f_2 & -f_2 \\ 1-f_3 & 1-f_3 & 1-f_3 & & -f_3 & -f_3 \\ \vdots & \vdots & \vdots & \ddots & & \vdots \\ 1-f_m & 1-f_m & 1-f_m & \dots & 1-f_m & -f_m \end{pmatrix}.
$$

Call the vector on the right $B^{(m+1)}$. It has independent components, its distribution is known.

$$
B^{(m+1)} \sim \mathsf{N}_{m+1}(0, \Gamma)
$$

with

$$\Gamma = \begin{pmatrix} f_1 & 0 & \dots & 0 & 0 \\ 0 & f_2 - f_1 & & 0 & 0 \\ \vdots & & \ddots & & \vdots \\ 0 & 0 & & f_m - f_{m-1} & 0 \\ 0 & 0 & \dots & 0 & 1 - f_m \end{pmatrix} = \mathrm{diag}(f_k - f_{k-1})_{k=1,\dots,m+1}.$$

Knowing the distribution of $B^{(m+1)}$, we also know its characteristic function. Let $a := (a_1, ..., a_{m+1})^T \in \mathbb{R}^{m+1}$. Then (cf. Lemma A.12)

$$\varphi_{B^{(m+1)}}(a) = \exp\Big(-\frac{1}{2}a^T\Gamma a\Big)$$

and (cf. Lemma A.13)

$$\varphi_{(B_1(t_1),\dots,B_1(t_m))}(x) = \varphi_{HB^{(m+1)}}(x) = \varphi_{B^{(m+1)}}(H^T x) = \exp\Big\{\frac{1}{2}(H^T x)^T \Gamma H^T x\Big\}.$$

We give this term a more descriptive shape.

$$\begin{aligned} H^T x &= \begin{pmatrix} 1-f_1 & 1-f_2 & \dots & 1-f_m \\ -f_1 & 1-f_2 & \dots & 1-f_m \\ \vdots & & \ddots & \vdots \\ -f_1 & -f_2 & & 1-f_m \\ -f_1 & -f_2 & \dots & -f_m \end{pmatrix} \begin{pmatrix} x_1 \\ x_2 \\ \vdots \\ x_m \end{pmatrix} \\ &= \begin{pmatrix} \sum_1^m -f_j x_j + \sum_1^m x_j \\ \sum_1^m -f_j x_j + \sum_2^m x_j \\ \vdots \\ \sum_1^m -f_j x_j + x_m \\ \sum_1^m -f_j x_j \end{pmatrix} = \begin{pmatrix} \lambda + \mu_1 \\ \lambda + \mu_2 \\ \vdots \\ \lambda + \mu_m \\ \lambda \end{pmatrix} \end{aligned}$$

Γ is a diagonal matrix, thus

$$(H^T x)^T \Gamma H^T x = \sum_{k=1}^{m+1} (f_k - f_{k-1})(\lambda + \mu_k)^2$$

and

$$\varphi_{(B_1(t_1),...,B_1(t_m))}(x) = \exp\Big\{ -\frac{1}{2}\sum_{k=1}^{m+1}(f_k - f_{k-1})(\lambda + \mu_k)^2\Big\}.$$

Now we determine the characteristic function of $(\tilde{N}_0(r_1), ..., \tilde{N}_0(r_q))$.

$$\begin{pmatrix} \tilde{N}_0(r_1) \\ \tilde{N}_0(r_2) \\ \vdots \\ \tilde{N}_0(r_q) \end{pmatrix} = \begin{pmatrix} -N_1(-r_1) \\ -N_1(-r_2) \\ \vdots \\ -N_1(-r_q) \end{pmatrix} = \begin{pmatrix} -1 & -1 & \dots & -1 \\ 0 & -1 & \dots & -1 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & & -1 \end{pmatrix} \begin{pmatrix} N_1(-r_1) - N_1(-r_2) \\ N_1(-r_2) - N_1(-r_3) \\ \vdots \\ N_1(-r_q) \end{pmatrix}$$

Call the matrix R and the vector on the right $N_1^{(q)}$. The distribution of $N_1^{(q)}$ again is known. It has independent components, and the k-th component has a Poisson distribution with parameter $\varrho_1(r_{k+1} - r_k)$, $k = 1, ..., q$. Hence, with $a := (a_1, ..., a_q) \in \mathbb{R}^q$,

$$\varphi_{N_1^{(q)}}(a) = \exp\Big\{ \sum_{k=1}^{q} \varrho_1(r_{k+1} - r_k)(e^{ia_k} - 1)\Big\}$$

(cf. Lemma A.14). Note that

$$R^T y = \begin{pmatrix} -1 & 0 & \dots & 0 \\ -1 & -1 & & 0 \\ \vdots & \vdots & \ddots & \\ -1 & -1 & \dots & -1 \end{pmatrix} \begin{pmatrix} y_1 \\ y_2 \\ \vdots \\ y_q \end{pmatrix} = \begin{pmatrix} -y_1 \\ -y_1 - y_2 \\ \vdots \\ \sum_{j=1}^q -y_j \end{pmatrix} = \begin{pmatrix} \xi_1 \\ \xi_2 \\ \vdots \\ \xi_q \end{pmatrix},$$

therefore (cf. Lemma A.13)

$$\begin{aligned} \varphi_{(\tilde{N}_0(r_1),...,\tilde{N}_0(r_q))}(y) &= \varphi_{RN_1^{(p)}}(y) = \varphi_{N_1^{(p)}}(R^T y) \\ &= \exp\Big\{ \sum_{k=1}^{q} \varrho_1(r_{k+1} - r_k)(e^{i\xi_k} - 1)\Big\}. \end{aligned}$$

As for $(\tilde{N}_0(s_1), ..., \tilde{N}_0(s_p))$ we have

$$\begin{pmatrix} \tilde{N}_0(s_1) \\ \tilde{N}_0(s_2) \\ \vdots \\ \tilde{N}_0(s_p) \end{pmatrix} = \begin{pmatrix} 1 & 0 & \dots & 0 \\ 1 & 1 & & 0 \\ \vdots & \vdots & \ddots & \\ 1 & 1 & \dots & 1 \end{pmatrix} \begin{pmatrix} N_2(s_1) \\ N_2(s_2) - N_2(s_1) \\ \vdots \\ N_2(s_p) - N_2(s_{p-1}) \end{pmatrix}$$

The vector on the right has independent components. The k-th component has a Poisson distribution with parameter $\varrho_2(s_k - s_{k-1})$, $k = 1, ..., p$. Furthermore

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ 0 & 1 & \dots & 1 \\ \vdots & & \ddots & \vdots \\ 0 & 0 & & 1 \end{pmatrix} \begin{pmatrix} z_1 \\ z_2 \\ \vdots \\ z_p \end{pmatrix} = \begin{pmatrix} \sum_{j=1}^{p} z_j \\ \sum_{j=2}^{p} z_j \\ \vdots \\ z_p \end{pmatrix} = \begin{pmatrix} \nu_1 \\ \nu_2 \\ \vdots \\ \nu_p \end{pmatrix},$$

and hence

$$\varphi_{(\tilde{N}_0(s_1),...,\tilde{N}_0(s_p))}(y) = \exp\Big\{ \sum_{k=1}^{p} \varrho_2(s_k - s_{k-1})(e^{i\nu_k} - 1) \Big\}.$$

Since B_1, N_1 and N_2 are mutually independent, the vectors $(B_1(t_1), ..., B_1(t_m))$, $(\tilde{N}_0(r_1), ..., \tilde{N}_0(r_q))$ and $(\tilde{N}_0(s_1), ..., \tilde{N}_0(s_p))$ are as well. This yields

$$\begin{aligned} \psi(x, y, z) &= \varphi_{(B_1(t_1),...,B_1(t_m))}(x)\ \varphi_{(\tilde{N}_0(r_1),...,\tilde{N}_0(r_q))}(y)\ \varphi_{(\tilde{N}_0(s_1),...,\tilde{N}_0(s_p))}(z) \\ &= \exp\Big\{ -\frac{1}{2} \sum_{k=1}^{m+1} (\mu_k + \lambda)^2 (f_k - f_{k-1}) \\ &\qquad + \sum_{k=1}^{q} \varrho_1 (e^{i\xi_k} - 1)(r_{k+1} - r_k) + \sum_{k=1}^{p} \varrho_2 (e^{i\nu_k} - 1)(s_k - s_{k-1}) \Big\}, \end{aligned}$$

and the lemma is proved. ∎

The next lemma deals with the characteristic function of Y_n. Here we have to distinguish two cases depending on the location of the time points. The distribution of Y_n looks different if one of the t_i coincides with τ.

Lemma 5.12

(1) If $\tau \neq t_i$ for all $1 \leq i \leq n$, then there is a $w \in \{0, ..., m\}$ such that $t_w < \tau < t_{w+1}$. If n is large enough that $t_w < \tau + \frac{r_1}{n}$ and $\tau + \frac{s_p}{n} < t_{w+1}$, the the characteristic function $\psi_n : \mathbb{R}^{m+q+p} \to \mathbb{C}$ of Y_n is

$$\begin{aligned} \psi_n(x, y, z) = e^{i\lambda\sqrt{n}} \Big\{ & \sum_{k=1}^{m+1} (f_k - f_{k-1}) e^{\frac{i\mu_k}{\sqrt{n}}} \\ & + e^{\frac{i\mu_{w+1}}{\sqrt{n}}} \Big[\sum_{k=1}^{q} \big[F(\tau + \tfrac{r_{k+1}}{n}) - F(\tau + \tfrac{r_k}{n})\big](e^{i\xi_k} - 1) \\ & + \sum_{k=1}^{p} \big[F(\tau + \tfrac{s_k}{n}) - F(\tau + \tfrac{s_{k-1}}{n})\big](e^{i\nu_k} - 1) \Big] \Big\}^n. \end{aligned}$$

$$(75)$$

(2) If $\tau = t_w$ for a $w \in \{1, ..., m\}$ and n is large enough that $t_{w-1} < \tau + \frac{r_1}{n}$ and $\tau + \frac{s_p}{n} < t_{w+1}$, then the characteristic function $\psi_n : \mathbb{R}^{m+q+p} \to \mathbb{C}$ of Y_n is

$$\begin{aligned}\psi_n(x,y,z) = e^{i\lambda\sqrt{n}}\Big[&\sum_{k=1}^{m+1}(f_k - f_{k-1})e^{\frac{i\mu_k}{\sqrt{n}}}\\ &+e^{\frac{i\,\mu_w}{\sqrt{n}}}\sum_{k=1}^{q}(e^{i\xi_k}-1)\big[F(\tau+\tfrac{r_{k+1}}{n}) - F(\tau+\tfrac{r_k}{n})\big]\\ &+e^{\frac{i\,\mu_{w+1}}{\sqrt{n}}}\sum_{k=1}^{p}(e^{i\nu_k}-1)\big[F(\tau+\tfrac{s_k}{n}) - F(\tau+\tfrac{s_{k-1}}{n})\big]\Big]^n.\end{aligned} \tag{76}$$

Proof. Part (1) will be treated in a detailed way. Afterwards we point out how part (2) differs and present the important intermediate results. Suppose that the assumptions of (1) hold.

$$\psi_n(x,y,z) = \mathbb{E}\exp\big\{\ i\ \langle Y_n, (x,y,z)\rangle\ \big\},$$

where $\langle\cdot,\cdot\rangle$ denotes the usual dot product.

$$\begin{aligned}\langle Y_n,(x,y,z)\rangle &= \sum_{j=1}^{m}\alpha_n(t_j)x_j + \sum_{j=1}^{q}\beta_n(r_j)y_j + \sum_{j=1}^{p}\beta_n(s_j)z_j\\ &= \sum_{j=1}^{m}\Big[\frac{1}{\sqrt{n}}\sum_{k=1}^{n}\big(\mathbb{1}_{(-\infty,t_j]}(X_k) - F(t_j)\big)\Big]x_j\\ &\quad+\sum_{j=1}^{q}\Big[\sum_{k=1}^{n}-\mathbb{1}_{(\tau+\frac{r_j}{n},\tau]}(X_k)\Big]y_j\ +\ \sum_{j=1}^{p}\Big[\sum_{k=1}^{n}\mathbb{1}_{(\tau,\tau+\frac{s_j}{n}]}(X_k)\Big]z_j\\ &= \sum_{k=1}^{n}\Big[\frac{1}{\sqrt{n}}\sum_{j=1}^{m}\big(\mathbb{1}_{(-\infty,t_j]}(X_k) - f_j\big)x_j\\ &\quad+\sum_{j=1}^{q}-y_j\mathbb{1}_{(\tau+\frac{r_j}{n},\tau]}(X_k)\ +\ \sum_{j=1}^{p}z_j\mathbb{1}_{(\tau,\tau+\frac{s_j}{n}]}(X_k)\Big]\end{aligned}$$

$$= \sum_{k=1}^{n}\Big[\frac{1}{\sqrt{n}}\sum_{j=1}^{m}x_j\mathbb{1}_{(-\infty,t_j]}(X_k) + \frac{\lambda}{\sqrt{n}} + \sum_{j=1}^{q}-y_j\mathbb{1}_{(\tau+\frac{r_j}{n},\tau]}(X_k) + \sum_{j=1}^{p}z_j\mathbb{1}_{(\tau,\tau+\frac{s_j}{n}]}(X_k)\Big]$$

We want to call the expression in the squared brackets $\zeta_{n,k}$, that is

$$\zeta_{n,k} := \frac{\lambda}{\sqrt{n}} + \frac{1}{\sqrt{n}} \sum_{j=1}^{m} x_j \mathbb{1}_{(-\infty,t_j]}(X_k) + \sum_{j=1}^{q} -y_j \mathbb{1}_{(\tau+\frac{r_j}{n},\tau]}(X_k) + \sum_{j=1}^{p} z_j \mathbb{1}_{(\tau,\tau+\frac{s_j}{n}]}(X_k)$$

Since the X_k, $k \in \mathbb{N}$, are i.i.d., we have

$$\psi_n(x,y,z) = \mathbb{E}\exp\Big\{i\sum_{k=1}^{n}\zeta_{n,k}\Big\} = \prod_{k=1}^{n}\mathbb{E}\, e^{i\,\zeta_{n,k}} = \Big(\mathbb{E}\, e^{i\,\zeta_{n,1}}\Big)^n.$$

Denote X_1 by X and let

$$Z := Z(X) := (Z_1, ..., Z_{m+q+p})$$

$$:= \Big(\mathbb{1}_{(-\infty,t_1]}(X), ..., \mathbb{1}_{(-\infty,t_m]}(X), \mathbb{1}_{(\tau+\frac{r_1}{n},\tau]}(X), ..., \mathbb{1}_{(\tau+\frac{r_q}{n},\tau]}(X), \mathbb{1}_{(\tau,\tau+\frac{s_1}{n}]}(X), ..., \mathbb{1}_{(\tau,\tau+\frac{s_p}{n}]}(X)\Big).$$

Then

$$\zeta_{n,1} = \frac{\lambda}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{j=1}^{m} x_j Z_j + \sum_{j=1}^{q} -y_j Z_{m+j} + \sum_{j=1}^{p} z_j Z_{m+q+j}. \tag{77}$$

Z is a random element in $\{0,1\}^{m+q+p}$. Under the assumptions of 5.12 (1), the order of the time points is fixed (with respect to n), and we distinguish $m+q+p+2$ disjoint events depending on between which two of adjacent time points X falls. This leads to (at most) $m+q+p+1$ different values that Z takes on with positive probability. Table 1 on page 69 lists these events, their respective probabilities and the corresponding value of Z. Note that there are two lines containing the same Z-value.
Let θ_n be the n-th root of $\psi_n(x,y,z)$, i.e.

$$\theta_n := \mathbb{E}\, e^{i\,\zeta_{n,1}}.$$

Because of (77) we can write down the expectation θ_n as a sum of $m+q+p+1$ summands.

$$\begin{aligned}
\theta_n \;=\; & \sum_{\substack{k=1\\k\neq w+1}}^{m+1} \exp\Big\{i\Big[\frac{\lambda}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{j=k}^{m} x_j\Big]\Big\}\big(F(t_k)-F(t_{k-1})\big)\\
& + \sum_{k=1}^{q} \exp\Big\{i\Big[\frac{\lambda}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{j=w+1}^{m} x_j + \sum_{j=1}^{k} -y_j\Big]\Big\}\big(F(\tau+\frac{r_{k+1}}{n}) - F(\tau+\frac{r_k}{n})\big)\\
& + \sum_{k=1}^{p} \exp\Big\{i\Big[\frac{\lambda}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{j=w+1}^{m} x_j + \sum_{j=k}^{p} z_j\Big]\Big\}\big(F(\tau+\frac{s_k}{n}) - F(\tau+\frac{s_{k-1}}{n})\big)\\
& + \exp\Big\{i\Big[\frac{\lambda}{\sqrt{n}} + \frac{1}{\sqrt{n}}\sum_{j=w+1}^{m} x_j\Big]\Big\}\Big(F(t_{w+1}) - F(\tau+\frac{s_p}{n}) + F(\tau+\frac{r_1}{n}) - F(t_w)\Big)
\end{aligned}$$

I	$\mathbb{P}(X \in I)$	Value of $Z(X)$ if $X \in I$											
		$\mathbb{1}^{(X)}_{(-\infty,t_1]}$	$\dots$	$\mathbb{1}^{(X)}_{(-\infty,t_w]}$	$\mathbb{1}^{(X)}_{(-\infty,t_{w+1}]}$	$\dots$	$\mathbb{1}^{(X)}_{(-\infty,t_m]}$	$\mathbb{1}^{(X)}_{(\tau+\frac{r_1}{n},\tau]}$	$\dots$	$\mathbb{1}^{(X)}_{(\tau+\frac{r_q}{n},\tau]}$	$\mathbb{1}^{(X)}_{(\tau,\tau+\frac{s_1}{n}]}$	$\dots$	$\mathbb{1}^{(X)}_{(\tau,\tau+\frac{s_p}{n}]}$
$(t_m, +\infty)$	$1 - F(t_m)$	0	$\dots$	0	0	$\dots$	0	0	$\dots$	0	0	$\dots$	0
$(t_{m-1}, t_m]$	$F(t_m) - F(t_{m-1})$	0	$\dots$	0	0		1	0	$\dots$	0	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$		$\iddots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$
$(\tau + \frac{s_p}{n}, t_{w+1}]$	$F(t_{w+1}) - F(\tau + \frac{s_p}{n})$	0	$\dots$	0	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0
$(\tau + \frac{s_{p-1}}{n}, \tau + \frac{s_p}{n}]$	$F(\tau + \frac{s_p}{n}) - F(\tau + \frac{s_{p-1}}{n})$	0	$\dots$	0	1	$\dots$	1	0	$\dots$	0	0		1
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\iddots$	$\vdots$
$(\tau, \tau + \frac{s_1}{n}]$	$F(\tau + \frac{s_1}{n}) - F(\tau)$	0	$\dots$	0	1	$\dots$	1	0	$\dots$	0	1	$\dots$	1
$(\tau + \frac{r_q}{n}, \tau]$	$F(\tau) - F(\tau + \frac{r_q}{n})$	0	$\dots$	0	1	$\dots$	1	1	$\dots$	1	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\iddots$		$\vdots$		$\vdots$
$(\tau + \frac{r_1}{n}, \tau + \frac{r_2}{n}]$	$F(\tau + \frac{r_2}{n}) - F(\tau + \frac{r_1}{n})$	0	$\dots$	0	1	$\dots$	1	1		0	0	$\dots$	0
$(t_w, \tau + \frac{r_1}{n}]$	$F(\tau + \frac{r_1}{n}) - F(t_w)$	0	$\dots$	0	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0
$(t_{w-1}, t_w]$	$F(t_w) - F(t_{w-1})$	0		1	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0
$\vdots$	$\vdots$		$\iddots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$
$(-\infty, t-1]$	$F(t_1)$	1	$\dots$	1	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0

Table 1: Values of Z if $t_w < \tau < t_{w+1}$

$$\begin{aligned}\theta_n = e^{\frac{i\lambda}{\sqrt{n}}}\Big[&\sum_{\substack{k=1\\k\neq w+1}}^{m+1} e^{\frac{i\,\mu_k}{\sqrt{n}}}(f_k - f_{k-1}) \\ &+\sum_{k=1}^{q}\exp\Big\{i\big[\frac{\mu_{w+1}}{\sqrt{n}}+\xi_k\big]\Big\}\big(F(\tau+\frac{r_{k+1}}{n}) - F(\tau+\frac{r_k}{n})\big) \\ &+\sum_{k=1}^{p}\exp\Big\{i\big[\frac{\mu_{w+1}}{\sqrt{n}}+\nu_k\big]\Big\}\big(F(\tau+\frac{s_k}{n}) - F(\tau+\frac{s_{k-1}}{n})\big) \\ &+e^{\frac{i\,\mu_{w+1}}{\sqrt{n}}}\big(f_{w+1} - f_w + F(\tau+\frac{r_1}{n}) - F(\tau+\frac{s_p}{n})\big)\Big]\end{aligned}$$

We split the last summand and add it to the others in a suitable way. Note that (telescoping sum)

$$F(\tau+\frac{r_1}{n}) = F(\tau) - \sum_{k=1}^{q}\big(F(\tau+\frac{r_{k+1}}{n}) - F(\tau+\frac{r_k}{n})\big) \tag{78}$$

and

$$F(\tau+\frac{s_p}{n}) = -F(\tau) - \sum_{k=1}^{p}\big(F(\tau+\frac{s_k}{n}) - F(\tau+\frac{s_{k-1}}{n})\big). \tag{79}$$

Hence

$$\begin{aligned}\theta_n = e^{\frac{i\lambda}{\sqrt{n}}}\Big[&\textstyle\sum_{k=1}^{m+1}(f_k - f_{k-1})e^{\frac{i\mu_k}{\sqrt{n}}} \\ &+e^{\frac{i\,\mu_{w+1}}{\sqrt{n}}}\Big\{ \textstyle\sum_{k=1}^{q}(e^{i\xi_k}-1)\big[F(\tau+\frac{r_{k+1}}{n}) - F(\tau+\frac{r_k}{n})\big] \\ &\qquad +\textstyle\sum_{k=1}^{p}(e^{i\nu_k}-1)\big[F(\tau+\frac{s_k}{n}) - F(\tau+\frac{s_{k-1}}{n})\big]\Big\}\Big]\end{aligned}$$

and $\psi_n(x,y,z) = \theta_n^n$ looks like (75).
As for 5.12 (2), the possible values of Z and the respective probabilities are shown in Table 2 on page 71. This leads to the following expression for θ_n.

I	$\mathbb{P}(X \in I)$	Value of $Z(X)$ if $X \in I$												
		$\mathbb{1}^{(X)}_{(-\infty,t_1]}$	$\dots$	$\mathbb{1}^{(X)}_{(-\infty,t_{w-1}]}$	$\mathbb{1}^{(X)}_{(-\infty,t_w]}$	$\mathbb{1}^{(X)}_{(-\infty,t_{w+1}]}$	$\dots$	$\mathbb{1}^{(X)}_{(-\infty,t_m]}$	$\mathbb{1}^{(X)}_{(\tau+\frac{r_1}{n},\tau]}$	$\dots$	$\mathbb{1}^{(X)}_{(\tau+\frac{r_q}{n},\tau]}$	$\mathbb{1}^{(X)}_{(\tau,\tau+\frac{s_1}{n}]}$	$\dots$	$\mathbb{1}^{(X)}_{(\tau,\tau+\frac{s_p}{n}]}$
$(t_m, +\infty)$	$1 - F(t_m)$	0	$\dots$	0	0	0	$\dots$	0	0	$\dots$	0	0	$\dots$	0
$(t_{m-1}, t_m]$	$F(t_m) - F(t_{m-1})$	0	$\dots$	0	0	0		1	0	$\dots$	0	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\iddots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$
$(\tau + \frac{s_p}{n}, t_{w+1}]$	$F(t_{w+1}) - F(\tau + \frac{s_p}{n})$	0	$\dots$	0	0	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0
$(\tau + \frac{s_{p-1}}{n}, \tau + \frac{s_p}{n}]$	$F(\tau + \frac{s_p}{n}) - F(\tau + \frac{s_{p-1}}{n})$	0	$\dots$	0	0	1	$\dots$	1	0	$\dots$	0	0		1
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$		$\iddots$	$\vdots$
$(\tau, \tau + \frac{s_1}{n}]$	$F(\tau + \frac{s_1}{n}) - F(\tau)$	0	$\dots$	0	0	1	$\dots$	1	0	$\dots$	0	1	$\dots$	1
$(\tau + \frac{r_q}{n}, \tau]$	$F(\tau) - F(\tau + \frac{r_q}{n})$	0	$\dots$	0	1	1	$\dots$	1	1	$\dots$	1	0	$\dots$	0
$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$	$\iddots$		$\vdots$		$\vdots$
$(\tau + \frac{r_1}{n}, \tau + \frac{r_2}{n}]$	$F(\tau + \frac{r_2}{n}) - F(\tau + \frac{r_1}{n})$	0	$\dots$	0	1	1	$\dots$	1	1		0	0	$\dots$	0
$(t_{w-1}, \tau + \frac{r_1}{n}]$	$F(\tau + \frac{r_1}{n}) - F(t_{w-1})$	0	$\dots$	0	1	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0
$(t_{w-2}, t_{w-1}]$	$F(t_{w-1}) - F(t_{w-2})$	0		1	1	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0
$\vdots$	$\vdots$		$\iddots$	$\vdots$	$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$	$\vdots$		$\vdots$
$(-\infty, t-1]$	$F(t_1)$	1	$\dots$	1	1	1	$\dots$	1	0	$\dots$	0	0	$\dots$	0

Table 2: Values of Z if $\tau = t_w$

$$
\begin{aligned}
\theta_n = \; e^{\frac{i\lambda}{\sqrt{n}}} \Big[\; & \sum_{\substack{k=1 \\ k \neq w, w+1}}^{m+1} e^{\frac{i\,\mu_k}{\sqrt{n}}} (f_k - f_{k-1}) \\
& + \sum_{k=1}^{q} \exp\Big\{ i\big[\frac{\mu_w}{\sqrt{n}} + \xi_k\big]\Big\} \big(F(\tau + \frac{r_{k+1}}{n}) - F(\tau + \frac{r_k}{n})\big) \\
& + \sum_{k=1}^{p} \exp\Big\{ i\big[\frac{\mu_{w+1}}{\sqrt{n}} + \nu_k\big]\Big\} \big(F(\tau + \frac{s_k}{n}) - F(\tau + \frac{s_{k-1}}{n})\big) \\
& + e^{\frac{i\,\mu_{w+1}}{\sqrt{n}}} \big(f_{w+1} - F(\tau + \frac{s_p}{n})\big) + e^{\frac{i\,\mu_w}{\sqrt{n}}} \big(F(\tau + \frac{r_1}{n}) - f_{w-1}\big) \Big]
\end{aligned}
$$

The last two summands equal

$$
\begin{aligned}
& e^{\frac{i\,\mu_{w+1}}{\sqrt{n}}} (f_{w+1} - f_w) + e^{\frac{i\,\mu_w}{\sqrt{n}}} (f_w - f_{w-1}) \\
& - e^{\frac{i\,\mu_{w+1}}{\sqrt{n}}} \sum_{k=1}^{q} \big(F(\tau + \frac{r_{k+1}}{n}) - F(\tau + \frac{r_k}{n})\big) - e^{\frac{i\,\mu_w}{\sqrt{n}}} \sum_{k=1}^{p} \big(F(\tau + \frac{s_k}{n}) - F(\tau + \frac{s_{k-1}}{n})\big).
\end{aligned}
$$

Use again (78) and (79). Hence

$$
\begin{aligned}
\theta_n = e^{\frac{i\lambda}{\sqrt{n}}} \Big[& \sum_{k=1}^{m+1} (f_k - f_{k-1}) e^{\frac{i\mu_k}{\sqrt{n}}} \\
& + e^{\frac{i\,\mu_w}{\sqrt{n}}} \sum_{k=1}^{q} (e^{i\xi_k} - 1)\big[F(\tau + \tfrac{r_{k+1}}{n}) - F(\tau + \tfrac{r_k}{n})\big] \\
& + e^{\frac{i\,\mu_{w+1}}{\sqrt{n}}} \sum_{k=1}^{p} (e^{i\nu_k} - 1)\big[F(\tau + \tfrac{s_k}{n}) - F(\tau + \tfrac{s_{k-1}}{n})\big] \Big],
\end{aligned}
$$

and $\psi_n(x, y, z) = \theta_n^n$ is given by (76).

■

Proof of Proposition 5.10. We show $\psi_n(x, y, z) \to \psi(x, y, z)$ as $n \to \infty$ for all $x \in \mathbb{R}^m$, $y \in \mathbb{R}^q$ and $z \in \mathbb{R}^p$ (cf. 2.16, 2.17). We restrict ourselves to the case 5.12 (1), it will be obvious that the second case works absolutely analoguous and leads to the same result.
We use the following limit theorem. If, for complex numbers x, x_n, $n \in \mathbb{N}$, $x_n \to x$, then

$$
\Big(1 + \frac{x_n}{n}\Big)^n \longrightarrow e^x. \tag{80}
$$

Hence, in order to prove

$$\psi_n(x,y) \longrightarrow \psi(x,y),$$

it suffices to show that

$$n\big(\psi_n^{\frac{1}{n}}(x,y) - 1\big) \longrightarrow \ln \psi(x,y).$$

Call the left-hand side h_n and the right-hand side h. By Lemmas 5.12 and 5.11 we have

$$\begin{aligned} h_n = \; & n \cdot e^{\frac{i\lambda}{\sqrt{n}}} \Bigg\{ \sum_{k=1}^{m+1} (f_k - f_{k-1}) e^{\frac{i\mu_k}{\sqrt{n}}} \\ & \quad + e^{\frac{i\mu_{w+1}}{\sqrt{n}}} \Bigg[\sum_{k=1}^{q} \big[F(\tau + \tfrac{r_{k+1}}{n}) - F(\tau + \tfrac{r_k}{n})\big](e^{i\xi_k} - 1) \\ & \qquad + \sum_{k=1}^{p} \big[F(\tau + \tfrac{s_k}{n}) - F(\tau + \tfrac{s_{k-1}}{n})\big](e^{i\nu_k} - 1)\Bigg]\Bigg\} - n. \end{aligned}$$

$$\begin{aligned} = \; & n \sum_{k=1}^{m+1} (f_k - f_{k-1})(e^{\frac{i}{\sqrt{n}}(\lambda + \mu_k)} - 1) \\ & + \exp\big\{\frac{i}{\sqrt{n}}(\lambda + \mu_{w+1})\big\} \sum_{k=1}^{q} n\big[F(\tau + \tfrac{r_{k+1}}{n}) - F(\tau + \tfrac{r_k}{n})\big](e^{i\xi_k} - 1) \\ & + \exp\big\{\frac{i}{\sqrt{n}}(\lambda + \mu_{w+1})\big\} \sum_{k=1}^{p} n\big[F(\tau + \tfrac{s_k}{n}) - F(\tau + \tfrac{s_{k-1}}{n})\big](e^{i\nu_k} - 1) \end{aligned}$$

and

$$h = -\frac{1}{2}\sum_{k=1}^{m+1}(\mu_k+\lambda)^2(f_k-f_{k-1}) + \sum_{k=1}^{q} \varrho_1(e^{i\xi_k}-1)(r_{k+1}-r_k) + \sum_{k=1}^{p} \varrho_2(e^{i\nu_k}-1)(s_k-s_{k-1}).$$

We are done when we have proved the following three convergence statements (i.e. we show the convergence of each of the three summands separately).

(a) $n \sum_{k=1}^{m+1} (f_k - f_{k-1})(e^{\frac{i}{\sqrt{n}}(\lambda + \mu_k)} - 1) \longrightarrow -\frac{1}{2} \sum_{k=1}^{m+1} (\mu_k + \lambda)^2 (f_k - f_{k-1})$

(b) $\exp\big\{\frac{i}{\sqrt{n}}(\lambda + \mu_{w+1})\big\} \sum_{k=1}^{q} n\big[F(\tau + \tfrac{r_{k+1}}{n}) - F(\tau + \tfrac{r_k}{n})\big](e^{i\xi_k} - 1)$

$\longrightarrow \sum_{k=1}^{q} \varrho_1 (e^{i\xi_k} - 1)(r_{k+1} - r_k)$

(c) $\exp\Big\{\dfrac{i}{\sqrt{n}}(\lambda+\mu_{w+1})\Big\}\displaystyle\sum_{k=1}^{p} n\big[F(\tau+\tfrac{s_k}{n})-F(\tau+\tfrac{s_{k-1}}{n})\big](e^{i\nu_k}-1)$

$$\longrightarrow \sum_{k=1}^{p} \varrho_2(e^{i\nu_k}-1)(s_k-s_{k-1})$$

Proof of (a). Use the Taylor expansion of exponential function,

$$e^c = \sum_{j=0}^{\infty}\frac{c^j}{j!}, \qquad c\in\mathbb{C}.$$

The fact that it converges uniformly on compact sets in $\mathbb{C}$, allows us to write

$$e^{\frac{i}{\sqrt{n}}(\lambda+\mu_k)} = 1+\frac{i}{\sqrt{n}}(\lambda+\mu_k)-\frac{(\lambda+\mu_k)^2}{2n}+o\big(\frac{1}{n}\big).$$

This yields

$$\begin{aligned}
& n\sum_{k=1}^{m+1}\Big(e^{\frac{i}{\sqrt{n}}(\lambda+\mu_k)}-1\Big)(f_k-f_{k-1}) \\
= \;& n\sum_{k=1}^{m+1}\Big[\frac{i}{\sqrt{n}}(\lambda+\mu_k)-\frac{(\lambda+\mu_k)^2}{2n}+o\big(\frac{1}{n}\big)\Big](f_k-f_{k-1}) \\
= \;& \sum_{k=1}^{m+1} i\sqrt{n}(\lambda+\mu_k)(f_k-f_{k-1})-\frac{1}{2}\sum_{k=1}^{m+1}(\mu_k+\lambda)^2(f_k-f_{k-1})+o(1)
\end{aligned}$$

The last step is to notice that $\sum_{k=1}^{m+1} i\sqrt{n}(\lambda+\mu_k)(f_k-f_{k-1})$ equals zero. (Note that this is true for the sum but not each summand.)

$$\begin{aligned}
\sum_{k=1}^{m+1}(\lambda+\mu_k)(f_k-f_{k-1}) &= \sum_{k=1}^{m+1}(\lambda+\mu_k)f_k-\sum_{k=0}^{m}(\lambda+\mu_{k+1})t_k \\
&= \lambda+\sum_{k=1}^{m}(\mu_k-\mu_{k+1})f_k=\lambda+\sum_{k=1}^{m}x_kf_k=\lambda-\lambda
\end{aligned}$$

Proof of (b). Since $\exp\{\frac{i}{\sqrt{n}}(\lambda+\mu)\}\to 1$ as $n\to\infty$, it suffices to show

$$\lim_{n\to\infty} n\big[F(\tau+\tfrac{r_{k+1}}{n})-F(\tau+\tfrac{r_k}{n})\big]=\varrho_1(r_{k+1}-r_k).$$

This can be seen as follows. Recall the definition (71) of ϱ_1 and that $r_j < 0$ $(j = 1, ..., q)$.

$$\begin{aligned} n\big[F(\tau+\tfrac{r_{k+1}}{n}) - F(\tau+\tfrac{r_k}{n})\big] &= n\big[F(\tau+\tfrac{r_{k+1}}{n}) - F(\tau)\big] - n\big[F(\tau+\tfrac{r_k}{n}) - F(\tau)\big] \\ &= r_{k+1}\frac{F(\tau+\frac{r_{k+1}}{n}) - F(\tau)}{\frac{r_{k+1}}{n}} - r_k\frac{F(\tau+\frac{r_k}{n}) - F(\tau)}{\frac{r_k}{n}} \\ &\overset{n\to\infty}{\longrightarrow} r_{k+1}\varrho_1 - r_k\varrho_1 = \varrho_1(r_{k+1} - r_k) \end{aligned}$$

Proof of (c). By the same means as above show

$$\lim_{n\to\infty} n\big[F(\tau+\tfrac{s_k}{n}) - F(\tau+\tfrac{s_{k-1}}{n})\big] = \varrho_2(s_k - s_{k-1}).$$

Thus, having proved (a), (b) and (c), we know that $h_n \to h$, which implies by (80) $\psi_n(x, y) \to \psi(x, y)$ and by Theorems 2.17 and 2.16, $Y_n \xrightarrow{\mathscr{L}} Y$. Proposition 5.10 is proved. ∎

The following corollary (and last proposition in this paper) takes us back to the special case we started from. Recall the processes α_n, B_0, β_n and $\tilde{N}$, defined by (16), (17), (52) and (53), respectively. Let $\tilde{B}_0 = \{\tilde{B}_0(t) | \, t \in \mathbb{R}\}$ and $\tilde{\alpha}_n = \{\tilde{\alpha}_n(t) | \, t \in \mathbb{R}\}$ be B_0 and α_n embedded in D_∞, i.e.

$$\tilde{B}_0(t) := \begin{cases} B_0(t), & t \in [0,1], \\ 0, & t \notin [0,1], \end{cases}$$

and

$$\begin{aligned} \tilde{\alpha}_n(t) &:= \sqrt{n}(G_n(t) - G(t)), \qquad t \in \mathbb{R}, \\ &= \begin{cases} \alpha_n(t), & t \in [0,1], \\ 0 \text{ a.s.}, & t \notin [0,1], \end{cases} \end{aligned}$$

where G denotes the distribution function of a Uniform$(0, 1)$ distribution, and G_n the corresponding empirical distribution function.

Corollary 5.13 $(\tilde{\alpha}_n, \beta_n) \xrightarrow{\mathscr{L}} (\tilde{B}_0, \tilde{N})$ *in* $(E, \mathscr{E})$.

Proof. We have $\tilde{\alpha}_n \overset{\mathscr{L}}{=} \alpha_n^G$, $\beta_n \overset{\mathscr{L}}{=} \beta_n^{G,0}$ and $\tilde{B}_0 \overset{\mathscr{L}}{=} B_1^G$. Moreover, the left-hand side and right-hand side derivatives of G at $\tau = 0$ are 0 and 1, respectively. Note that a Poisson process with rate 0 is a.s. identical zero, hence $\tilde{N} \overset{\mathscr{L}}{=} \tilde{N}_0(0, 1)$. The function G is continuous, so by Theorem 5.9,

$$(\tilde{\alpha}_n, \beta_n) \overset{\mathscr{L}}{=} \gamma_n^{G,0} \xrightarrow{\mathscr{L}} M^{G,0} = (B_1^G, \tilde{N}_0(0,1)) \overset{\mathscr{L}}{=} (\tilde{B}_0, \tilde{N}).$$

∎

Concluding remark. The question remains what happens if we drop the restriction of F being continuous. I conjecture that the convergence statement of Theorem 5.9 holds just as well. Recall that Proposition 5.10 does not require continuity of F. It does require Condition C.2, though, i.e left- and right-hand side derivative in τ must exist. From the proof of 5.10 it is obvious that we cannot expect any sensible limit without this restriction.
The Continuity is needed when we apply Theorem 4.9 to show tightness of $\{\alpha_n^F\}$ and $\{\beta_n^{F,\tau}\}$. Condition (2) of Theorem 4.9 is very well suited for these sequences if F is continuous, and in particular for the "standard special case" $\{\tilde{\alpha}_n\}$ and $\{\beta_n\}$. We can identify the distribution function F with the F in 4.9 (2), the latter being required to be continuous. Since the continuity is essential in the derivation of Theorem 4.9 (cf. (49)), there seems to be no "easy fix", in the sense that Theorem 4.9 could be easily adjusted to accomodate discontinuous F as well. But (the good news) 4.9 is just one possibility of proving tightness, it sufficient but not necessary. To sum it up, with the means provided in this paper we get the convergence statement 5.9 only for continuous distribution functions F.
Here is my idea how to get around. I suggest two different approaches for $\{\alpha_n^F\}$ and $\{\beta_n^{F,\tau}\}$. Concerning $\{\beta_n^{F,\tau}\}$ I would like to point to the alternative tightness criterion 4.16. It cannot be applied directly. It is formulated for processes on $D[0,\infty)$, and only for point processes in the classical sense, meaning the process may only jump by one at a time. As long as F is continuous, $\beta_n^{F,\tau}$ has only jumps of height one, but if F has a discontinuity at, say, t_0 (i.e. the distribution has mass concentrated in t_0), then $\beta_n^{F,\tau}$ has a jump of more than one at $n\,t_0$ with positive probability. I expect that the proposition can be extended in an appropriate way to such "labelled point processes".
As for $\{\alpha_n^F\}$, we know that $\{\alpha_n\}$ is convergent in $D[0,1]$ (cf. Theorem 3.12). α_n^F is the quantile transformation of α_n, i.e. $\alpha_n^F = \alpha_n \circ F$. It is easy to verify that the quantile transformation $D[0,1] \to D(-\infty,\infty) : x \mapsto x \circ F$ is continuous on $C[0,1]$, i.e. $\mathscr{L}(B_0)$-a.e. continuous. I do not know whether it is continuous (with respect to $\mathscr{D}$ and $\mathscr{D}_\infty$). By the CMT (cf. 2.15) α_n^F converges to B_1^F, hence $\{\alpha_n^F\}$ is relatively compact (cf. Corollary 2.13). We cannot conclude, though, that $\{\alpha_n^F\}$ is tight, since (D_∞, d_∞), as we have declared it, is not complete (cf. Theorem 2.14 (2)). Here is another obstacle. We need tightness of both sequences $\{\alpha_n^F\}$ and $\{\beta_n^{F,\tau}\}$ to show tightness of the product sequence $\gamma_n^{F,\tau}$ (Lemma 5.2), which implies its relative compactness (Prokhorov's theorem, cf. 2.14). What to do? Either one tries to show that relative compactness also transfers from the components to the Cartesian product (which won't work as nicely as it does for tightness), or one makes use of the following: Although (D_∞, d_∞) is not complete, the induced Skorokhod

topology is metrizable with a complete metric. Just like in (D, d) (see remarks on page 13), there exists a complete metric, say $\tilde{d_\infty}$, which is equivalent to d_∞ (see [Bil99], pages 168 and 170). Completeness is a property of the metric, while separability is a topological property. Tightness is a topological property, too (its definition involves compact sets), which means, if $\{\alpha_n^F\}$ is tight in $(D_\infty, \tilde{d_\infty})$, then it is just as well tight in (D_∞, d_∞), because the topologies of both metric spaces agree.

A Lemmas and References

In the subsequent lemmas, dealing with Skorokhod convergence, we will employ the following characterization.

Lemma A.1 *The following are equivalent:*

(I) $x_n \to x$ *in* $(D[0,1], d)$.

(II) There exists a sequence $\{\lambda_n\} \subset \Lambda$ *such that*

$$\begin{cases} x_n(\lambda_n(t)) \to x(t) & \text{uniformly in } t \in [0,1], \\ \lambda_n(t) \to t & \text{uniformly in } t \in [0,1]. \end{cases} \tag{81}$$

(III) There exists a sequence $\{\lambda_n\} \subset \Lambda$ *such that*

$$\begin{cases} |\, x_n(t) - x(\lambda_n(t))| \to 0 & \text{uniformly in } t \in [0,1], \\ \lambda_n(t) \to t & \text{uniformly in } t \in [0,1]. \end{cases} \tag{82}$$

Proof. This is an immediate corollary from the definiton of the metric d, keeping in mind the definition of infimum and supremum. (II) is a citation from [Bil99], page 124. The equivalence (II)-(III) follows right away from (2): If $\{\lambda_n\}$ satisfies (81), then $\{\lambda_n^- 1\}$ does satisfy (82).

■

Lemma A.2 $\mathbb{1}_{[0,\frac{1}{2}-\frac{1}{n})} \to \mathbb{1}_{[0,\frac{1}{2})}$ *in* D.

Proof. Apply A.1 (II). Take λ_n to be the linear interpolation of the three points $(0,0)$, $(\frac{1}{2}-\frac{1}{n},\frac{1}{2})$ and $(1,1)$, i.e. λ_n lifts $\frac{1}{2}-\frac{1}{n}$ up to $\frac{1}{2}$ and is linear otherwise, or explicitely:

$$\lambda_n(t) := \begin{cases} \frac{t}{1-\frac{2}{n}}, & t < \frac{1}{2} - \frac{1}{n}, \\ \frac{t}{1+\frac{2}{n}} + \frac{1}{1+\frac{n}{2}}, & t \geq \frac{1}{2} - \frac{1}{n}. \end{cases}$$

Then

$$\lambda_n(t) := \begin{cases} < \frac{1}{2} & \text{for } t < \frac{1}{2} - \frac{1}{n}, \\ \geq \frac{1}{2} & \text{for } t \geq \frac{1}{2} - \frac{1}{n}, \end{cases}$$

hence $\mathbb{1}_{[0,\frac{1}{2}-\frac{1}{n})}(\lambda_n(t)) = \mathbb{1}_{[0,\frac{1}{2})}(t)$.

■

Lemma A.3 $\mathbb{1}_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2})} \nrightarrow \mathbb{0}$ *in* D.

Proof. Apply A.1 (III). Suppose there is a sequence $\{\lambda_n\} \subset \Lambda$ satisfying (82), then, with $x_n = \mathbb{1}_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2})}$ and $x = \mathbb{0}$,

$$|\, x_n(t) - x(\lambda_n(t))| = \mathbb{1}_{[\frac{1}{2}-\frac{1}{n},\frac{1}{2})}(t),$$

which does not converge to zero uniformly in $t \in [0,1]$.

■

Lemma A.4 $\mathbb{1}_{[0,1-\frac{1}{n})} \nrightarrow \mathbb{1}_{[0,1)}$ *in* D.

Proof. Apply A.1 (III). With $x_n = \mathbb{1}_{[0,1-\frac{1}{n})}$ and $x = \mathbb{1}_{[0,1)}$ we have

$$x([0,1]) = x_n([0,1]) = \{0,1\}$$

for all $n \in \mathbb{N}$. Therefore (82) implies, that there exists an n_0 such that

$$x(\lambda_n(t)) = x_n(t) \qquad \forall\, t \in [0,1], n \geq n_0,$$

because otherwise

$$\sup_{t\in[0,1]} |\, x(\lambda_n(t)) = x_n(t)| \ = 1$$

for infinitely many $n \in \mathbb{N}$. That means

$$\mathbb{1}_{[0,1)}(\lambda_n(t)) = x(\lambda_n(t)) = x_n(t) = \mathbb{1}_{[0,1-\frac{1}{n})}(t) = 0 \qquad \forall\, t \geq 1 - \frac{1}{n},$$

hence

$$\lambda_n(t) = 1 \qquad \forall\, t \geq 1 - \frac{1}{n},$$

which is a contradiction to the strict monotony of λ_n.

■

Lemma A.5 *Let S be a separable metric space and $\mathscr{T}$ be a σ-field on S. If $\mathscr{T}$ contains all open ball, then $\mathscr{B}(S) \subset \mathscr{T}$.*

Proof. This is a standard topology argumentation. Let A be countable, dense set in S, which exists according to the assumptions. Then, let B the set of all open balls with rational radii around the elements of A. The system B is a countable base of the topology S, hence every open set is a countable union of elements of B. But B is a subset of $\mathscr{T}$, hence, with $\mathscr{T}$ being a σ-field, every open set is contained $\mathscr{T}$, so is the σ-field generated by all open sets.

∎

Recall the moduli $\hat{w}$ and w'', defined by (34) and (35):

$$\begin{aligned} \hat{w}(x,\delta) &= \inf_{\mathscr{S}_1(\delta)} \max_{1\le i\le \nu} w_x[s_{i-1}, s_i), \\ w''(x,\delta) &= \sup_{\mathscr{T}_0(\delta)} \left\{ |x(t)-x(t_1)| \wedge |x(t_2)-x(t)| \right\}, \end{aligned}$$

where

$$\mathscr{S}_1(\delta) = \left\{ \{s_0, ..., s_\nu\} \,\middle|\, \nu \in \mathbb{N}, s_0 = 0, s_\nu = 1, s_i - s_{i-1} > \delta \;\forall\, i = 2, ..., \nu - 1 \right\}.$$

and

$$\mathscr{T}_0(\delta) = \left\{ \{t_1, t, t_2\} \,\middle|\, 0 \le t_1 \le t \le t_2 \le 1,\; t_2 - t_1 \le \delta \right\}.$$

The following inequality is repeatedly used in the proof of Lemma A.7.

Lemma A.6 *It holds*

$$|\, x(s) - x(t_1)| \wedge |\, x(t_2) - x(t)| \le 2w''(x,\delta) \tag{83}$$

for all functions $x \in D$ and all $0 \le t_1 \le s \le t \le t_2$ with $t_2 - t_1 \le \delta$.

Proof. This is (12.33) in [Bil99], but the formula there contains a typo. The inequality is not hard to get, though. We show: if $|\, x(s) - x(t_1)| > w''(x,\delta)$, then $|\, x(t_2) - x(t)| \le 2w''(x,\delta)$. The assumption implies by the definition of w'',

$$|\, x(t_2) - x(s)| \le w''(x,\delta)$$

and

$$|\, x(t) - x(s)| \le w''(x,\delta),$$

hence

$$|\, x(t_2) - x(t)| = |\, x(t_2) - x(s) + x(s) - x(t_1)| \le |\, x(t_2) - x(s)| + |\, x(s) - x(t_1)| \le 2w''(x,\delta).$$

∎

Lemma A.7 *For each $x \in D$ and $\delta > 0$,*

$$\hat{w}(x, \frac{\delta}{2}) \leq 6w''(x, \delta).$$

Proof. Let $a > w''(x, \delta)$. It suffices to prove $\hat{w}(x, \frac{\delta}{2}) \leq 6a$. The first step towards this end is to show that x cannot have two jumps exceeding $2a$ within the distance smaller than δ of one another. Suppose there are two points $u_1 < u_2$ at which x has jumps of height larger than $2a$. If $u_2 - u_1 < \delta$, then there exist points s_1 and s_2 such that $s_1 < u_1 < s_2 < u_2$ and $u_2 - s_1 < \delta$. Since the left-hand side limit of x at each point exists, we have

$$\lim_{s_1 \nearrow u_1} |\, x(u_1) - x(s_1)| \; = |\, x(u_1) - x(u_1-)| \; > 2a.$$

Thus by moving s_1 sufficiently close to u_1 one can always achieve that

$$|\, x(u_1) - x(s_1)| \; > 2a,$$

and by the same means

$$|\, x(u_2) - x(s_2)| \; > 2a,$$

hence

$$|\, x(u_1) - x(s_1)| \wedge |\, x(u_2) - x(s_2)| \; > 2a.$$

But this is a contradiction to

$$|\, x(u_1) - x(s_1)| \wedge |\, x(u_2) - x(s_2)| \; \leq 2w''(x, \delta) < 2a,$$

which must be true due to Lemma A.6. Therefore u_1 and u_2 have a distance of at least δ from one another.
Next, we construct a grid $S = \{s_0, ..., s_k\}$ on $[0, 1]$ (cf. subsection 3.2) having the following properties:

(I) $s_i - s_{i-1} \leq \delta \quad \forall\, i = 1, ..., k,$
i.e. the mesh of S is less than or equal to δ.

(II) $s_i - s_{i-1} > \frac{\delta}{2} \quad \forall\, i = 2, ..., k-1,$
i.e. S is a $\frac{\delta}{2}$-grid with soft boundares.

(III) If there is a point $u \in [0, 1]$ for which holds $|\, x(u) - x(u-)| \; > 2a$, then $u \in S$.

Such a grid can be obtained as follows. Take all points $u \in [0,1]$ with $|\,x(u) - x(u-)| > 2a$, plus the endpoints 0 and 1. This leads to a grid $\tilde{S} = \{\tilde{s}_0, ..., \tilde{s}_{\tilde{k}}\}$, which has, as shown above, the property $\tilde{s}_i - \tilde{s}_{i-1} \geq \delta$ for $i = 1, ..., \tilde{k}$. New points are added the following way:

- In the interval $(\tilde{s}_0, \tilde{s_1})$: Successively add the points $\tilde{s}_1 - \delta, \tilde{s}_1 - 2\delta, ...$ as long as they still greater than 0.
- In the interval $(\tilde{s}_0, \tilde{s_1})$, $i = 2, ..., \tilde{r}-1$: If $\tilde{s}_i - \tilde{s}_{i-1} > \delta$, add the midpoint $\frac{\tilde{s}_i - \tilde{s}_{i-1}}{2}$. Iterate.
- In the interval $(\tilde{s}_{\tilde{k}-1}, \tilde{s}_{\tilde{k}})$: Successively add the points $\tilde{s}_{\tilde{k}-1} + \delta, \tilde{s}_{\tilde{k}-1} + 2\delta, ...$ as long as they are smaller than 1.

It is evident that the thus obtained grid $S = \{s_1, ..., s_k\}$ satisfies conditions (I) to (III).
From (I) and (III) we are able to deduct the following property:

$$w_x[s_{i-1} - s_i) \leq 6a \qquad \forall\, i = 1, ..., r.$$

Keep in mind, that x, δ and a are connected via the relation $a \geq w''(x, \delta)$. Again, Lemma A.6 will be the key argument. Let $s_{i-1} \leq t_1 < t_2 < s_i$. What we need to show is $|\,x(t_1) - x(t_2)| \leq 6a$. Let

$$\sigma_1 := \sup\big\{s \in [t_1, t_2]\big|\ \sup_{t_1 \leq u \leq s} |\,x(u) - x(t_1)| \leq 2a\big\}. \tag{84}$$

In words, σ_1 is the point where the graph of x first exceeds $2a$-distance from $x(t_1)$ to the right of t_1. This is the same as

$$\inf\big\{s \in [t_1, t_2]\big|\,|\,x(s) - x(t_1)| > 2a\big\}, \tag{85}$$

but only if this infimum exists, that is, if the set is not empty. In this case we want σ_1 to equal t_2, and thus use definition (84) instead of (85). Likewise

$$\sigma_2 := \inf\big\{s \in [t_1, t_2]\big|\ \sup_{s \leq u \leq t_2} |\,x(t_2) - x(u)| \leq 2a\big\}. \tag{86}$$

Claim:

$$\sigma_2 \leq \sigma_1. \tag{87}$$

Proof: Suppose $\sigma_1 < \sigma_2$. Then there exist points s and t such that $\sigma_1 < s, t < \sigma_2$,

$$|\,x(t_1) - x(s)| > 2a \quad \text{and} \quad |\,x(t_2) - x(t)| > 2a. \tag{88}$$

Due to the properties of the supremum and infimum we can move s and t arbitrarily close to σ_1 and σ_2, respectively, and by doing so, always arrange $s < t$. Since $t_2 - t_1 < \delta$ (property (I)), we have by (83),

$$|\,x(t_1) - x(s)| \wedge |\,x(t_2) - x(t)| \,\leq 2w''(x, \delta) < 2a,$$

which contradicts (88), therefore (87) must hold.
Thus, we summarize,

- $|\,x(t_1) - x(\sigma_1-)| \,\leq 2a$ (due to (84)),
- $|\,x(\sigma_1-) - x(\sigma_1)| \,\leq 2a$ (due to property (III)),
- $|\,x(\sigma_1) - x(t_2)| \,\leq 2a$ (due to (86)and (87)).

Hence, by the triangle inequality,

$$|\,x(t_2) - x(t_1)| \,\leq 6a.$$

This proves the lemma, because

$$w_x[s_{i-1} - s_i) = \sup_{t_1, t_2 \in [s_{i-1} - s_i)} |\,x(t_2) - x(t_1)| \,\leq 6a,$$

and since $S = \{s_1, ..., s_2\} \in \mathscr{S}_1(\frac{\delta}{2})$,

$$\hat{w}(x, \frac{\delta}{2}) = \inf_{\mathscr{S}_1(\frac{\delta}{2})} \max_{1 \leq i \leq \nu} w_x[t_{i-1}, t_i) \leq \max_{1 \leq i \leq k} w_x[s_{i-1}, s_i) \leq 6a.$$

This holds for all $a > w''(x, \delta)$, hence $\hat{w}(x, \frac{\delta}{2}) \leq 6w''(x, \delta)$.

■

Definition A.8 (Multinomial distribution) Let $X = (X_1, ..., X_k)$, $k \geq 2$ be a real random vector. X is said to have a *multinomial* or *polynomial distribution with parameters* $n, p_1, ..., p_k$ ($n \in \mathbb{N}, \sum_1^k p_i = 1$), if

$$\mathbb{P}(X = r) = \frac{n!}{r_1!...r_k!} p_1^{r_1} \cdot ... \cdot p_k^{r_k}$$

for all $r = (r_1, ..., r_k) \in \{0, ..., n\}^k$ with $\sum_1^k r_i = n$.
Interpretation: Suppose the possible outcomes of random experiment are the k different events $A_1, ..., A_k$, occuring with probabilities $p_1, ..., p_k$, respectively. Suppose further that the experiment is carried independently n times. Then $\mathbb{P}(X = r)$ is the probability that the event A_i occurs exactly r_i times, $i = 1, ...k$. (See also [Mül91], page 303.)

Proposition A.9 (Poisson limit theorem) *Let $\{X_n\}$ be a sequence of real random variables, with X_n having a Binomial distribution with parameters n and p_n such that $\lim_{n\to\infty} np_n =: \lambda > 0$. Then*

$$\mathbb{P}(X_n = k) = \binom{n}{k} p_n^k (1-p_n)^{n-k} \longrightarrow e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, 2, ...,$$

as $n \to \infty$, which implies that

$$X_n \xrightarrow{\mathscr{L}} \text{Poisson}(\lambda),$$

where Poisson(λ) *denotes a random variable having a Poisson distribution with parameter λ.*

Proof. See, e.g. [Mül91], page 155.

□

Lemma A.10 *The following are equivalent:*

(I) $(x_n, y_n) \to (x, y)$ *in* $(D_{[0,1]}(\mathbb{R}^2), d^{(2)})$.

(II) There exists a sequence $\{\lambda_n\} \subset \Lambda$ such that

$$\begin{cases} |\, x_n(t) - x(\lambda_n(t))| \to 0 & \textit{uniformly in } t \in [0,1], \\ |\, y_n(t) - y(\lambda_n(t))| \to 0 & \textit{uniformly in } t \in [0,1], \\ \lambda_n(t) \to t & \textit{uniformly in } t \in [0,1]. \end{cases} \tag{89}$$

Proof. Just as in Lemma A.1, $(x_n, y_n) \to (x, y)$ is equivalent to: There is $\{\lambda_n\} \subset \Lambda$ with

$$\begin{cases} |\,(x_n(t), y_n(t)) - (x(\lambda_n(t)), y(\lambda_n(t)))| \to 0 & \text{uniformly in } t \in [0,1], \\ \lambda_n(t) \to t & \text{uniformly in } t \in [0,1]. \end{cases}$$

The former of the conditions is equivalent to

$$\left.\begin{cases} |\, x_n(t) - x(\lambda_n(t))| \to 0 \\ |\, y_n(t) - y(\lambda_n(t))| \to 0 \end{cases}\right. \qquad \text{uniformly in } t \in [0,1].$$

■

Lemma A.11 $(\mathbb{1}_{[0,\frac{1}{2}-\frac{1}{n})}, \mathbb{1}_{[0,\frac{1}{2}+\frac{1}{n})}) \nrightarrow (\mathbb{1}_{[0,\frac{1}{2})}, \mathbb{1}_{[0,\frac{1}{2})})$ *in* $D_{[0,1]}(\mathbb{R}^2)$.

Proof. Apply A.10. With $x_n = \mathbb{1}_{[0,\frac{1}{2}-\frac{1}{n})}$, $y_n = \mathbb{1}_{[0,\frac{1}{2}+\frac{1}{n})}$ and $x = y = \mathbb{1}_{[0,\frac{1}{2})}$, suppose there is a sequence $\{\lambda_n\} \subset \Lambda$ such that (89) is fulfilled. Then

$$\begin{aligned} & |\, x_n(t) - x(\lambda_n(t))| \; + |\, y(\lambda_n(t)) - y_n(t)| \\ \geq \;& |\, x_n(t) - x(\lambda_n(t)) + y(\lambda_n(t)) - y_n(t)| \\ = \;& |\, x_n(t) - y_n(t)| \\ = \;& 1 \quad \text{if } \tfrac{1}{2} - \tfrac{1}{n} \leq t < \tfrac{1}{2} + \tfrac{1}{n}. \end{aligned}$$

This means, $|\, x_n(t) - x(\lambda_n(t))| \; + |\, y(\lambda_n(t)) - y_n(t)|$ cannot converge to zero uniformly in t, which implies that none of the two summands does either. Hence (89) cannot be true.

■

Lemma A.12 *Let X be normal random vector in $\mathbb{R}^k$, having expectation $\mu \in \mathbb{R}^k$ and covariance matrix $\Sigma \in \mathbb{R}^{k\times k}$. The characteristic function $\varphi_X : \mathbb{R}^k \to \mathbb{C}$ of X is given by*

$$\varphi_X(z) = \mathbb{E}e^{i\langle z,X\rangle} = \exp\{i\langle z, \mu\rangle \; - \; \frac{1}{2}\, z^T \, \Sigma \, z\}, \qquad z \in \mathbb{R}^k.$$

Proof. cf. [Fer02a], Satz 5.4.

□

Lemma A.13 *Let X be a random variable in $\mathbb{R}^k$, $A \in \mathbb{R}^{m\times k}$, $b \in \mathbb{R}^m$ and $Y = AX + b$ a.s., moreover let $\varphi_X : \mathbb{R}^k \to \mathbb{C}$ and $\varphi_Y : \mathbb{R}^m \to \mathbb{C}$ be the characteristic functions of X and Y, respectively. Then,*

$$\varphi_Y(z) = e^{i\langle z,b\rangle}\varphi_X(A^T z), \qquad z \in \mathbb{R}^m.$$

Proof. cf. [Fer02a], Satz 4.6 (3).

□

Lemma A.14 *Let X be a random variable having a Poisson distribution with parameter λ. The characteristic function $\varphi_X : \mathbb{R} \to \mathbb{C}$ of X is given by*

$$\varphi_X(t) = \mathbb{E}e^{itX} = \exp\{\lambda(e^{it} - 1)\}.$$

Proof. cf. [Mül91], page 301.

□

References

[AHE84] Ata Al-Hussaini and Robert J. Elliot. Convergence of the empirical distribution to the poisson process. *Stochastics*, 13:299–308, 1984.

[Bau92] Heinz Bauer. *Maß- und Integrationstheorie*. De Gruyter Lehrbuch. Walter de Gruyter & Co., 2nd edition, 1992.

[Bil95] Patrick Billingsley. *Probability and Measure*. Wiley series in probability and mathematical statistics. John Wiley & Sons, Inc., 3rd edition, 1995.

[Bil99] Patrick Billingsley. *Convergence of Probability Measures*. Wiley Series in Probability and Statisitics. John Wiley & Sons, Inc., 2nd edition, 1999.

[Don51] Monroe. D. Donsker. An invariance principle for certain probability limit theorems. *Mem. Amer. Math. Soc.*, 6, 1951.

[Don52] Monroe D. Donsker. Justification and extension of Doob's heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Stat.*, 23:277–281, 1952.

[Doo49] Joseph L. Doob. Heuristic approach to the Kolmogorov-Smirnov theorems. *Ann. Math. Stat.*, 20:393–403, 1949.

[EK86] Stewart N. Ethier and Thomas G. Kurtz. *Markov Processes: Characterization and Convergence,*. Wiley, 1986.

[Els02] Jürgen Elstrodt. *Maß- und Integrationstheorie*. Grundwissen Mathematik. Springer, 3rd edition, 2002.

[Fer02a] Dietmar Ferger. *Lecture "Lineare Modelle"*. Technische Universität Dresden, Wintersemester 2001/2002.

[Fer02b] Dietmar Ferger. *Lecture "Asymptotische Statistik 1"*. Technische Universität Dresden, Sommersemester 2002.

[Fer03a] Dietmar Ferger. *Lecture "Asymptotische Statistik 2"*. Technische Universität Dresden, Wintersemester 2002/2003.

[Fer03b] Dietmar Ferger. On the minimizing point of the incorrect centered empirical process and its limit distribution in nonregular experiments. *Dredner Schriften zur Mathematischen Stochastik*, February 2003.

[JS02] Jean Jacod and Albert N. Shiryaev. *Limit Theorems for Stochastic Processes*. Springer, 2nd edition, 2002.

[KLS80] Yu. M. Kabanov, R. Sh. Liptser, and A. N. Shiryaev. Some limit theorems for simple point processes (a martingare approach). *Stochastics*, 3:203–216, 1980.

[Kol56] N. A. Kolmogorov. On Skorokhod convergence. *Theory Probab. Appl.*, 1:215–222, 1956.

[Lin73] Torgny Lindvall. Weak convergence of probability measures and random functions in the function space $D[0, \infty)$. *Journal of Applied Probability*, 10:109–121, 1973.

[ŁR85] Zbigniew A. Łagodowski and Zdzisław Rychlik. Weak convergence of probability measures on the function space $D_d[0, \infty)$. *Bull. Pol. Acad. Sci., Math.*, 33:219–223, 1985.

[Mül91] P. H. Müller, editor. *Wahrscheinlichkeitsrechnung und mathematische Statistik: Lexikon der Stochastik*. Akademie-Verlag Berlin, 5th edition, 1991.

[Pol84] David Pollard. *Convergence of Stochastic Processes*. Springer, 1984.

[Pro56] Yu. V. Prokhorov. Convergence of random processes and limit theorems in probability theory. *Theory Probab. Appl.*, 1:157–214, 1956.

[Sko56] A. V. Skorokhod. Limit theorems for stochastic processes. *Theory Probab. Appl.*, 1:261–290, 1956.

[Sto63] Charles Stone. Weak convergence of stochastic processes defined on a semi-infinite time interval. *Proc. Amer. Math. Soc.*, 14:694–696, 1963.

[SW86] Galen R. Shorack and Jon A. Wellner. *Empirical Processes with Applications to Statistics*. Wiley Series in Probability and Mathematical Statistics. John Wiley & Sons, 1986.

[vdWW96] Aad W. van der Waart and Jon A. Wellner. *Weak Convergence and Empirical Processes*. Springer Series in Statistics. Springer, 1996.

[Wel02] Jon A. Wellner. *Lecture "Advanced Theory of Statistical Inference"*. University of Washington, Autumn Quater 2002.

[Whi70] Ward Whitt. Weak convergence of probability measures on the function space C[0,1]. *Ann. Math. Stat.*, 41:939–944, 1970.

[Whi71] Ward Whitt. Weak convergence of probability measures on the function space $D[0,\infty)$. Technical report, Yale University, 1971.

[Whi02] Ward Whitt. *Stochastic-Process Limits*. Springer Series in Operations Research. Springer, 2002.

Erklärung

Hiermit erkläre ich, dass ich die am heutigen Tag eingereichte Diplomarbeit zum Thema “Weak Convergence of the Empirical Process and the Rescaled Empirical Distribution Function in the Skorokhod Product Space” unter Betreuung von Herrn Prof. Dr. rer. nat. habil. D. Ferger selbständig erarbeitet, verfasst und Zitate kenntlich gemacht habe. Andere als die angegebenen Hilfsmittel wurden von mir nicht benutzt.

Datum Unterschrift