MATHEMATICS OF OPERATIONS RESEARCH
Vol. 35, No. 1, February 2010, pp. 27–51
ISSN 0364-765X | EISSN 1526-5471 | 10 | 3501 | 0027

DOI 10.1287/moor.1090.0422

Weak Differentiability of Product Measures

Bernd Heidergott
Department of Econometrics and Operations Research, Vrije Universiteit Amsterdam,
1081 HV Amsterdam, The Netherlands, bheidergott@feweb.vu.nl

Haralambie Leahu
Department of Mathematics and Computer Science, Technische Universiteit Eindhoven,
5600 MB Eindhoven, The Netherlands, h.leahu@tue.nl

In this paper, we study cost functions over a finite collection of random variables. For these types of models, a calculus of differentiation is developed that allows us to obtain a closed-form expression for derivatives where "differentiation" has to be understood in the weak sense. The technique for proving the results is new and establishes an interesting link between functional analysis and gradient estimation. The key contribution of this paper is a product rule of weak differentiation. In addition, a product rule of weak analyticity is presented that allows for Taylor series approximations of finite products measures. In particular, from characteristics of the individual probability measures, a lower bound (i.e., domain of convergence) can be established for the set of parameter values for which the Taylor series converges to the true value. Applications of our theory to the ruin problem from insurance mathematics and to stochastic activity networks arising in project evaluation review techniques are provided.

Key words: weak derivatives; gradient estimation; Taylor series approximations
MSC2000 subject classification: Primary: 60A10; secondary: 28A15
OR/MS subject classification: Primary: Probability: distribution; secondary: probability: stochastic models applications
History: Received October 6, 2007; revised October 12, 2008. Published online in *Articles in Advance* December 8, 2009.

1. Introduction. A wide range of probabilistic models in the area of manufacturing, transportation, finance, and communication can be modeled by studying cost functions over a finite collection of random variables. More specifically, letting $\mu_{i,\theta}$ be a probability measure on some state space $\mathbb{S}_i$ (for $1 \le i \le n$) depending on some parameter θ (with $\theta \in \Theta = (a, b) \subset \mathbb{R}$ for $a < b$), one is concerned with the following type of models:

$$J_g(\theta) \stackrel{\text{def}}{=} \mathbb{E}_\theta[g(X_n, \ldots, X_1)] = \int_{S_1} \cdots \int_{S_n} g(s_n, \ldots, s_1)\, \mu_{n,\theta}(ds_n) \cdots \mu_{1,\theta}(ds_1) \tag{1}$$

for g a cost function defined on the product space $\mathbb{S}_1 \times \cdots \times \mathbb{S}_n$, where X_i is distributed according to $\mu_{i,\theta}$. This class of models contains, for example, insurance models over a finite number of claims or transient waiting times in queueing networks.

In performance analysis, one is not only interested in evaluating $J_g(\theta)$ but also in sensitivity analysis and optimization, which requires evaluating $dJ_g(\theta)/d\theta$. In general, $J_g(\theta)$ cannot be obtained in a closed form and $dJ_g(\theta)/d\theta$ can only be evaluated with the help of advanced mathematical techniques.

In this paper, we will provide a calculus of differentiation for finite products of measures that allows us to obtain a closed-form expression for $dJ_g(\theta)/d\theta$. Here, "differentiation" has to be understood in the weak sense (see §5 for a formal definition). The concept of weak differentiation of measures was introduced in Pflug [28] (see also the monograph (Pflug [27])) and has been extended to the more general concept of $\mathcal{D}$-differentiation in Heidergott and Vázquez-Abad [13].

Weak differentiation has been successfully applied to deriving unbiased gradient estimators for cost functions over Markov chains (see Heidergott and Hordijk [11], Heidergott and Vázquez-Abad [12, 13], Heidergott et al. [17], Pflug [28]). Applications to gradient estimation for finite models can be found in Heidergott et al. [15, 18, 19]. For results on bounds on perturbations, we refer to Heidergott et al. [16]. However, some important fundamental issues that are of importance in applications of weak derivatives have not been addressed in the literature so far. This paper will close that gap as it provides the theory of weak differentiability of finite products of probability measures in a clean mathematical setting. In particular, the following issues will be addressed:

- The relation between the cost functions g for which the derivative in (1) exists and the distributions used in the model.
- The relation between strong (i.e., norm) differentiability and weak differentiability
- Preserving weak differentiability under finite products of measures for the types of cost functions (such as continuous or measurable) that are of importance in applications.
- Writing weak derivatives as the difference between appropriate measures. This is not only of importance for gradient estimation but can also be used in other ways. For example, we show how bounds for higher-order

derivatives can be obtained by a simple stochastic order argument, which in turn yield bounds for the remainder term of a Taylor expansion.

- Establishing Taylor series approximations of finite products by deriving a product rule of weak analyticity. In particular, from characteristics of the individual probability measures, a lower bound can be established for the set of parameter values for which the Taylor series converges to the true value.

Taylor series expansions of finite models as in (1) have been studied in the literature on $(\max, +)$-linear systems (see Baccelli et al. [1], Hasenfuß [8]), where continuity of the cost function is assumed. See also Heidergott [9]. The theory developed in this paper extends these results as (i) the continuity assumption on the cost function can be dropped, and (ii) Taylor approximations can be obtained when only finitely many derivatives exist (i.e., analyticity fails).

The main technical difficulty in establishing a calculus of weak differentiation is the following. Consider two probability measures, say, μ_θ and ν_θ, living on measurable spaces $(\mathbb{S}, \mathscr{S})$ and $(\mathbb{T}, \mathscr{T})$, respectively. To come up with a product rule of differentiation for the product measure $\mu_\theta \times \nu_\theta$, one has to be able to conclude from properties of μ_θ on $\mathbb{S}$ and ν_θ on $\mathbb{T}$ differentiability properties of the product measure living on the product space $\mathbb{S} \times \mathbb{T}$. Hence, it is clear that to establish a product rule, one has to study the relations between the function spaces on $\mathbb{S}$, $\mathbb{T}$ and $\mathbb{S} \times \mathbb{T}$. As we will show in this paper, techniques from functional analysis can be made fruitful for this. More precisely, if functional spaces are equipped with the v-norm and product spaces with the product v-norm, then Banach space theory can be used to bound the effect of perturbing θ in $\mu_\theta \times \nu_\theta$.

As in conventional analysis, the main work in establishing a product rule for weak differentiability lies in establishing the fact that the product of weakly continuous probability measures is again weakly continuous. Such results are built on limit theory for sequences of signed measures. To this end, we will develop in this paper a limit theory for signed measures, which to the best of our knowledge has not been established in the literature so far. In particular, we will show that weak convergence of a sequence of signed measures does not imply weak convergence of the negative and positive parts, respectively.

The paper is organized as follows. For ease of reference, we introduce in §2 the basic notation that will be used throughout the paper. Section 3 is devoted to signed measures and their limits. In particular, in §3.3, the important concept of a $\mathscr{D}_v$-space is introduced. Functional analysis on $\mathscr{D}_v$-spaces will be addressed in §4. Weak differentiability of probability measures is discussed in §5 and that of products of probability measures in §6. Section 7 is devoted to weak analyticity. Applications of our theory to the ruin problem from insurance mathematics and to stochastic activity networks arising in project evaluation review techniques (PERT) will be provided in §8. Some technical material is provided in Appendices A and B.

2. General notation. Throughout the paper, we consider a separable metric space $(\mathbb{S}, \rho)$ and we denote by $\mathscr{M} = \mathscr{M}(\mathscr{S})$ the linear space of all finite signed, regular measures on the measurable space $(\mathbb{S}, \mathscr{S})$, where $\mathscr{S}$ denotes the Borel field on $\mathbb{S}$. We also introduce the following notation:

- $\mathscr{C}(\mathbb{S})$ denotes the space of real-valued continuous mappings on $\mathbb{S}$.
- $\mathscr{C}_B(\mathbb{S}) \subset \mathscr{C}(\mathbb{S})$ denotes the subspace of continuous and bounded mappings.
- $\mathscr{C}^+(\mathbb{S}) \subset \mathscr{C}(\mathbb{S})$ denotes the subset of positive mappings, i.e.,

$$\mathscr{C}^+(\mathbb{S}) \stackrel{\text{def}}{=} \{g \in \mathscr{C}(\mathbb{S})\colon g(s) \geq 0, \quad \forall s \in \mathbb{S}\}.$$

- $\mathscr{F}(\mathbb{S})$ denotes the space of real-valued measurable mappings on $\mathbb{S}$.
- $\mathscr{F}_B(\mathbb{S}) \subset \mathscr{F}(\mathbb{S})$ denotes the subspace of bounded mappings.
- $\mathscr{M}^+(\mathscr{S}) \subset \mathscr{M}(\mathscr{S})$ denotes the cone of positive measures, i.e.,

$$\mathscr{M}^+(\mathscr{S}) \stackrel{\text{def}}{=} \{\mu \in \mathscr{M}(\mathscr{S})\colon \mu(E) \geq 0, \quad \forall E \in \mathscr{S}\}.$$

- $\mathscr{M}^1(\mathscr{S}) \subset \mathscr{M}(\mathscr{S})$ denotes the set of probability measures, i.e.,

$$\mathscr{M}^1(\mathscr{S}) \stackrel{\text{def}}{=} \{\mu \in \mathscr{M}^+(\mathscr{S})\colon \mu(\mathbb{S}) = 1\}.$$

- $\mathscr{L}^1(\mathbb{S}, \mu)$ is the set of all integrable functions w.r.t. $\mu \in \mathscr{M}(\mathscr{S})$ (note that $g \in \mathscr{L}^1(\mathbb{S}, \mu)$ if $|g| \in \mathscr{L}^1(\mathbb{S}, \mu)$). For $\mathscr{P} \subset \mathscr{M}(\mathscr{S})$, we denote by $\mathscr{L}^1(\mathbb{S}, \mathscr{P})$ the set of all integrable functions w.r.t. each μ in $\mathscr{P}$; in formula:

$$\mathscr{L}^1(\mathbb{S}, \mathscr{P}) \stackrel{\text{def}}{=} \bigcap_{\mu \in \mathscr{P}} \mathscr{L}^1(\mathbb{S}, \mu).$$

For $p \geq 1$, we denote by $\mathscr{L}^p(\mathbb{S}, \mathscr{P})$ the set of all real-valued functions g on S that are p-integrable; in symbols, $\forall\, p \geq 1\colon g \in \mathscr{L}^p(\mathbb{S}, \mathscr{P}) \Leftrightarrow |g|^p \in \mathscr{L}^1(\mathbb{S}, \mathscr{P})$.

We shall omit specifying the underlying set $\mathbb{S}$, or Borel field $\mathscr{S}$, when no confusion occurs.

3. Weak convergence of measures. Weak convergence of probability measures was originally introduced in Billingsley [2]. In this section, we extend the concept and give a topology on $\mathscr{M}$ by means of a convergence of sequences of arbitrary measures. We start by a brief overview on signed measures in §3.1. In §3.2, we define the concept of weak convergence on $\mathscr{M}$.

3.1. Signed measures. In the following, we state some standard facts about signed measures. Any signed measure $\mu \in \mathscr{M}$ can be written as the difference between two positive measures. More precisely, there exist $\mu^+, \mu^- \in \mathscr{M}^+$ such that

$$\forall E \in \mathscr{S}\colon \ \mu(E) = \mu^+(E) - \mu^-(E). \tag{2}$$

Note that the presentation in (2) is not unique.

Two positive measures are called *orthogonal* if they have disjoint support. More formally, $\mu_1, \mu_2 \in \mathscr{M}^+$ are orthogonal if there exists a set $A \in \mathscr{S}$ such that $\mu_1(\mathbb{S} \setminus A) = \mu_2(A) = 0$. Uniqueness of the presentation in (2) can be achieved if μ^+ and μ^- are orthogonal. In this case, (2) is called *Hahn-Jordan decomposition*.

For any $\mu \in \mathscr{M}$, one can define *the variation measure* $|\mu| \in \mathscr{M}^+$ and the *total variation* $\|\mu\|_{TV}$ of μ as follows:

$$\forall E \in \mathscr{S}\colon \ |\mu|(E) = \sup_{A \in \mathscr{S},\, A \subset E} |\mu(A)|,$$

and

$$\|\mu\|_{TV} = |\mu|(\mathbb{S}) = \sup_{A \in \mathscr{S}} |\mu(A)|. \tag{3}$$

It is worth noting that the Hahn-Jordan decomposition "minimizes" the sum $\mu^+ + \mu^-$, meaning that the variation measure as defined in (3) satisfies $|\mu| = \mu^+ + \mu^-$ for μ^+ and μ^- orthogonal. For further details on measure theory, we refer to Cohn [3].

3.2. Weak convergence on $\mathscr{M}$. The following definition introduces the concept of *weak convergence* on $\mathscr{M}$.

Definition 3.1. A sequence $\{\mu_n\}_n \subset \mathscr{M}$ is said to be *weakly $\mathscr{D}$-convergent* for some $\mathscr{D} \subset \mathscr{L}^1(\{\mu_n\colon n \in \mathbb{N}\})$, or weakly convergent for short, if there exists $\mu \in \mathscr{M}$ such that

$$\forall g \in \mathscr{D}\colon \ \lim_{n\to\infty} \int g(s)\mu_n(ds) = \int g(s)\mu(ds). \tag{4}$$

We write $\mu_n \overset{\mathscr{D}}{\Longrightarrow} \mu$ (or $\mu_n \Rightarrow \mu$ when no confusion occurs) and we call μ a weak limit[1] of the sequence $\{\mu_n\}_n$.

Note that classical weak convergence of measures is recovered through $\mathscr{D} = \mathscr{C}_B$ (see Billingsley [2]). The following example illustrates the dependence of $\mathscr{D}$-convergence of a sequence of measures $\{\mu_n\}_n$ on the choice of $\mathscr{D}$.

Example 3.1. On $\mathbb{S} = [0, \infty)$, let us consider the family of probability measures

$$\forall x \geq 0\colon \ \mu_\theta(dx) = C_\theta \frac{x^\theta}{(1+x)^3}\, dx$$

for $\theta \in (0, 2)$, where

$$C_\theta = \begin{cases} 2, & \theta = 1 \\[2mm] \dfrac{2\sin(\pi\theta)}{\pi\theta(1-\theta)}, & \text{otherwise.} \end{cases}$$

Provided that $\mathscr{D} = \mathscr{C}_B$, one can easily show that $\mu_\theta \overset{\mathscr{D}}{\Longrightarrow} \mu_1$ for $\theta \uparrow 1$. However, for g being the identity mapping, the limit

$$\lim_{\theta \uparrow 1} \int x\, \mu_\theta(dx)$$

fails to be finite. Hence, for $\mathscr{D} = \mathscr{C}$, μ_θ is not $\mathscr{D}$-convergent.

A natural question that rises in the study of limits of signed measures is whether $\mu_n \overset{\mathscr{D}}{\Longrightarrow} \mu$ implies that $\mu_n^+ \overset{\mathscr{D}}{\Longrightarrow} \mu^+$ and $\mu_n^- \overset{\mathscr{D}}{\Longrightarrow} \mu^-$. The following example shows that this is generally not the case.

[1] In general, the weak limit μ is not unique.

EXAMPLE 3.2. Let us consider the sequence

$$\mu_n = \begin{cases} \delta_{\frac{1}{n}} + \delta_{(n+1/n)} - \delta_1, & \text{for } n \text{ even,} \\ \delta_{\frac{1}{n}}, & \text{for } n \text{ odd,} \end{cases}$$

where δ_x denotes the Dirac distribution that assigns mass to x. Then, $\mu_n \overset{\mathscr{C}_B}{\Longrightarrow} \delta_0$ as n tends to ∞, but $\mu^+_{2k+1} \overset{\mathscr{C}_B}{\Longrightarrow} \delta_0$ and $\mu^+_{2k} \overset{\mathscr{C}_B}{\Longrightarrow} \delta_0 + \delta_1$ as k tends to ∞.

3.3. $\mathscr{D}_v$-spaces. Let $\mathscr{D}(\mathbb{S})$ be a linear space such that $\mathscr{C}_B(\mathbb{S}) \subset \mathscr{D}(\mathbb{S}) \subset \mathscr{F}(\mathbb{S})$. For $v \in \mathscr{C}^+(\mathbb{S})$, denote the set of mappings in $\mathscr{D}(\mathbb{S})$ that are bounded by a multiple of v by $\mathscr{D}_v(\mathbb{S})$:

$$\mathscr{D}_v(\mathbb{S}) \overset{\text{def}}{=} \{g \in \mathscr{D}(\mathbb{S}) \mid \exists\, c > 0\colon |g(s)| \leq c \cdot v(s),\ \forall\, s \in \mathbb{S}\}. \tag{5}$$

The minimal c for which inequality (5) holds true is the so-called *v-norm* (to be formally introduced in the next section). Note that $\mathscr{D}_v(\mathbb{S})$ is a linear subspace of $\mathscr{D}(\mathbb{S})$. Moreover, $\mathscr{C}_B(\mathbb{S}) \subset \mathscr{D}_v(\mathbb{S})$ provided that[2]

$$\inf\{v(s)\colon s \in \mathbb{S}\} > 0.$$

A typical choice for $\mathscr{D}_v(\mathbb{S})$ is provided in the following example.

EXAMPLE 3.3. Let $v(x) = e^x$ for $x \in \mathbb{S} = [0, \infty)$. Because for every polynomial P it holds that $\lim_{x\to\infty} e^{-x}P(x) = 0$, it turns out that the space $\mathscr{D}_v([0,\infty))$ contains all (finite) polynomials. However, the polynomials are not the only elements of $\mathscr{D}_v$ because, for instance, the mapping $x \mapsto \ln(1+x)$ also belongs to $\mathscr{D}_v$.

REMARK 3.1. If the sequence $\{\mu_n\}_n \subset \mathscr{M}^1$ converges weakly to μ in the classical sense, i.e., $\mu_n \overset{\mathscr{C}_B}{\Longrightarrow} \mu$, then

$$\lim_{n\to\infty} \int v(s)\mu_n(ds) = \int v(s)\mu(ds)$$

is equivalent to the uniform integrability of v w.r.t. the sequence $\{\mu_n\}_n$, i.e.,

$$\lim_{\alpha\to\infty} \sup_n \int |v(s)| \cdot \mathbb{I}_{\{s\colon |v(s)| \geq \alpha\}}(s)\mu_n(ds) = 0.$$

See, e.g., Billingsley [2]. One can easily show that uniform integrability of v implies uniform integrability of all continuous $g \in \mathscr{D}_v$. Hence, we conclude that if $\mu_n \overset{\mathscr{C}_B}{\Longrightarrow} \mu$ and v is uniformly integrable w.r.t. $\{\mu_n\}_n$, then $\mu_n \overset{\mathscr{D}_v}{\Longrightarrow} \mu$ provided that $\mathscr{D}(\mathbb{S}) \subset \mathscr{C}(\mathbb{S})$.

4. Normed spaces. This section relates the theory put forward so far to classical functional analysis. Section 4.1 deals with functional normed spaces and §4.2 addresses spaces of measures. Eventually, §4.3 extends the results from the previous sections to product spaces.

4.1. Functional spaces. For $v \in \mathscr{C}^+(\mathbb{S})$, one introduces the so-called *v-norm* on $\mathscr{F}(\mathbb{S})$ as follows:

$$\|g\|_v \overset{\text{def}}{=} \sup_{s\in\mathbb{S}} \frac{|g(s)|}{v(s)} = \inf\{c > 0\colon |g(s)| \leq c \cdot v(s),\ \forall\, s \in \mathbb{S}\}.$$

In particular, for each $g \in \mathscr{F}$, it holds that[3]

$$\forall\, s \in \mathbb{S}\colon |g(s)| \leq \|g\|_v \cdot v(s). \tag{6}$$

EXAMPLE 4.1. Let $\mathscr{D}_v$ be defined as in Example 3.3. For $P(x) = 1 + x$, for $x \geq 0$ we have $P(x) \leq e^x$ for all $x \geq 0$ and

$$\sup_{x\geq 0} P(x)e^{-x} = \lim_{x\downarrow 0} (1+x)e^{-x} = 1.$$

Hence, $\|P\|_v = 1$. On the other hand, if $Q(x) = x$, then $\|Q\|_v = e^{-1}$ because

$$\sup_{x\geq 0} xe^{-x} = e^{-1}.$$

[2] This condition is typically assumed in the literature to ensure the embedding $\mathscr{C}_b \subset \mathscr{D}_v$ and, consequently, the uniqueness of the $\mathscr{D}_v$-limit. However, as detailed in Appendix A, the assumption is not crucial because we can still speak about "uniqueness of limit" in a sensible way. This is precisely formulated in Lemma A.1 (Appendix A).

[3] Note that the inequality in (6) still holds true if $\|g\|_v = \infty$.

REMARK 4.1. The v-norm is also known as weighted supremum norm in the literature. An early reference is Lipman [25]. The v-norm is frequently used in Markov decision analysis. First traces date back to the early 1980s (see Dekker and Hordijk [4] and the revised version that was published as Dekker and Hordijk [5]). The v-norm was originally used in analysis of Blackwell optimality; see Dekker and Hordijk [5]. Also see Hordijk and Yushkevich [20] for a recent publication on this topic. Since then, it has been used in various forms under different names in many subsequent papers. See, for example, Meyn and Tweedie [26] and Kartashov [21]. For the use of v-norm in the theory of measure-valued differentiation of Markov chains, see Heidergott and Hordijk [10].

Let $\mathscr{D}(\mathbb{S}) \subset \mathscr{F}(\mathbb{S})$. We now introduce the set of elements of $\mathscr{D}(\mathbb{S})$ with finite v-norm, denoted by $[\mathscr{D}(\mathbb{S})]_v$, as follows:

$$[\mathscr{D}(\mathbb{S})]_v \stackrel{\text{def}}{=} \{g \in \mathscr{D}(\mathbb{S}): \|g\|_v < \infty\}. \tag{7}$$

The set $\mathscr{D}(\mathbb{S})$ in the definition of $[\mathscr{D}(\mathbb{S})]_v$ is called the *base set* of $[\mathscr{D}(\mathbb{S})]_v$. Note that the set $\mathscr{D}_v(\mathbb{S})$ defined in (5) can be written as $[\mathscr{D}(\mathbb{S})]_v$ and $[\mathscr{C}]_v = \mathscr{C}_B$, for $v \in \mathscr{C}_B$. Moreover, if $v \equiv 1$, then the v-norm coincides with the supremum norm $\|\cdot\|_\infty$ on $\mathscr{C}_B$.

As it will turn out, powerful results on convergence, continuity, and differentiability of product measures can be established if the base set in (7) is such that $[\mathscr{D}]_v$ becomes a Banach space when endowed with the appropriate v-norm. This gives rise to the following definition.

DEFINITION 4.1. Let $\mathscr{D}(\mathbb{S}) \subset \mathscr{F}(\mathbb{S})$ and let $v \in \mathscr{C}^+(\mathbb{S})$. The pair $(\mathscr{D}(\mathbb{S}), v)$ is called a *Banach base* on $\mathbb{S}$ if:
(i) $\mathscr{D}(\mathbb{S})$ is a linear space such that $\mathscr{C}_B(\mathbb{S}) \subset \mathscr{D}(\mathbb{S})$.
(ii) The set $[\mathscr{D}(\mathbb{S})]_v$ endowed with the v-norm is a Banach space.

In the following, we present Banach bases that are of importance in applications. In particular, it is shown that for $v \in \mathscr{C}^+(\mathbb{S})$, the functional spaces $[\mathscr{C}(\mathbb{S})]_v$, $[\mathscr{F}(\mathbb{S})]_v$, and $[\mathscr{L}^p(\mathbb{S}, \{\mu_\theta: \theta \in \Theta\})]_v$ are Banach bases.

EXAMPLE 4.2. The *continuity* paradigm: $\mathscr{D} = \mathscr{C}$. Taking $v \in \mathscr{C}^+$, we obtain $[\mathscr{C}]_v$ as the set of all continuous mappings bounded by v. It can be shown that $(\mathscr{C}, v)$ is a Banach base on $\mathbb{S}$. Indeed, note that the mapping[4] $\Phi: [\mathscr{C}(\mathbb{S})]_v \to \mathscr{C}_B(\mathbb{S}_v)$ defined as

$$\forall s \in \mathbb{S}_v: (\Phi g)(s) = \frac{g(s)}{v(s)}, \tag{8}$$

where $\mathbb{S}_v$ denotes the support[5] of v, establishes a linear bijection between two normed spaces, and the inverse $\Phi^{-1}: \mathscr{C}_B(\mathbb{S}_v) \to [\mathscr{C}(\mathbb{S})]_v$ is given by

$$\forall s \in \mathbb{S}: (\Phi^{-1} f)(s) = \begin{cases} f(s) \cdot v(s), & s \in \mathbb{S}_v, \\ 0, & s \notin \mathbb{S}_v. \end{cases}$$

Furthermore, Φ is an isometry as it satisfies

$$\forall g \in [\mathscr{C}(\mathbb{S})]_v: \|\Phi g\|_\infty = \|g\|_v.$$

Because $\mathscr{C}_B(\mathbb{S}_v)$ is a Banach space when equipped with the supremum-norm, $[\mathscr{C}(\mathbb{S})]_v$ inherits the same property (see Semadeni [31]).

The *measurability* paradigm: $\mathscr{D} = \mathscr{F}$. Taking $v \in \mathscr{C}^+$, we obtain $[\mathscr{F}]_v$ as the set of all measurable mappings bounded by v. Again, the mapping Φ: $[\mathscr{F}(\mathbb{S})]_v \to \mathscr{F}_B(\mathbb{S}_v)$ defined by (8) is an isometry and, using the same argument as in the continuity paradigm, we conclude that $(\mathscr{F}, v)$ is a Banach base on $\mathbb{S}$.

The $\mathscr{L}^p$*-integrability* paradigm: Let $\{\mu_\theta: \theta \in \Theta\} \subset \mathscr{M}^1$ and $v \in \mathscr{C}^+ \cap \mathscr{L}^p(\{\mu_\theta: \theta \in \Theta\})$ for some $p \geq 1$, s.t. $\mu_\theta(\mathbb{S} \setminus \mathbb{S}_v) = 0$, for all $\theta \in \Theta$. By considering the isometry Φ: $[\mathscr{L}^p(\mathbb{S}, \{\mu_\theta: \theta \in \Theta\})]_v \to \mathscr{L}^\infty(\mathbb{S}_v, \{\mu_\theta: \theta \in \Theta\})$,[6] we conclude that $(\mathscr{L}^p(\{\mu_\theta: \theta \in \Theta\}), v)$ is a Banach base on $\mathbb{S}$.

4.2. Spaces of measures. In functional analysis, signed measures often appear as continuous linear functionals on functional spaces. More precisely, by *Riesz's representation theorem*, a space of measures can be seen as the *topological dual* of a certain Banach space of functions. Throughout this section, we aim to exploit this fact to derive new results, using specific tools from Banach space theory.

[4] The assumption $v \in \mathscr{C}$ guarantees that Φg is a continuous mapping, provided that g is continuous.

[5] That is the set where v does not vanish. In formula: $\mathbb{S}_v = \{s \in \mathbb{S}: v(s) \neq 0\}$.

[6] By writing $\mathscr{L}^\infty(\mathbb{S}_v, \{\mu_\theta: \theta \in \Theta\})$, we commit a slight abuse of notation as the measure μ_θ is defined on $\mathbb{S}$ and not on $\mathbb{S}_v$. However, because for all $\theta \in \Theta$ the measure μ_θ does not assign any mass outside $\mathbb{S}_v$, the notation is justified.

For $v \in \mathscr{C}^+(\mathbb{S})$, let us consider the following space of measures:

$$\mathscr{M}_v \stackrel{\text{def}}{=} \left\{\mu \in \mathscr{M}\colon v \in \mathscr{L}^1(\mu)\right\},$$

and note that $[\mathscr{F}]_v \subset \mathscr{L}^1(\{\mu_\theta\colon \theta \in \Theta\})$ is equivalent to $\{\mu_\theta\colon \theta \in \Theta\} \subset \mathscr{M}_v$. For $\mu \in \mathscr{M}_v$, consider its Hahn-Jordan decomposition $\mu = \mu^+ - \mu^-$ and define the *weighted total variation norm* of μ w.r.t. v (for short, v-norm) as follows:

$$\|\mu\|_v = \int v(s)|\mu|(ds) = \int v(s)\mu^+(ds) + \int v(s)\mu^-(ds). \tag{9}$$

Note that by using the v-norm, the space $\mathscr{M}_v$ can be alternatively described as

$$\mathscr{M}_v = \{\mu \in \mathscr{M}\colon \|\mu\|_v < \infty\}.$$

For $\mu \in \mathscr{M}_v$, $T_\mu\colon [\mathscr{D}]_v \to \mathbb{R}$ defined as $T_\mu(g) = \int g(s)\mu(ds)$ is a linear mapping of the Banach space $[\mathscr{D}]_v$ onto $\mathbb{R}$ whose operator norm is given by

$$\|T_\mu\|_v \stackrel{\text{def}}{=} \sup\{|T_\mu(g)|\colon \|g\|_v \le 1\} = \sup\left\{\left|\int g(s)\mu(ds)\right|\colon \|g\|_v \le 1\right\}.$$

It is easy to check that the operator norm of T_μ coincides with the v-norm of μ and if $v \equiv 1$ one recovers the total variation norm as introduced in (3). In particular, the *Cauchy-Schwartz inequality* holds for v-norms. In formula

$$\forall\, g \in [\mathscr{D}]_v, \quad \forall\, \mu \in \mathscr{M}_v\colon \left|\int g(s)\mu(ds)\right| \le \|g\|_v \cdot \|\mu\|_v. \tag{10}$$

The v-norm induces *strong convergence* on $\mathscr{M}_v$ in the obvious way: We say that the sequence $\{\mu_n\}_n$ converges in v-norm (strongly) to some $\mu \in \mathscr{M}_v$ if

$$\lim_{n\to\infty} \|\mu_n - \mu\|_v = \lim_{n\to\infty} \sup_{\|g\|_v \le 1} \left|\int g(s)\mu_n(ds) - \int g(s)\mu(ds)\right| = 0.$$

Note that v-norm convergence implies $[\mathscr{D}]_v$-weak convergence. For example, it is known that convergence of μ_n toward μ in total variation norm implies that (4) holds for the set $\mathscr{D} = \mathscr{C}_B$. For general v, this is a consequence of the Cauchy-Schwartz inequality. Indeed, if μ_n converges in v-norm to μ, letting $\nu = \mu_n - \mu$ in (10), it follows that (4) holds true for all $g \in [\mathscr{D}]_v$. In words, "strong convergence implies weak convergence," which justifies the terms "weak" and "strong." The converse, however, is not true, as detailed in Example 4.3.

Example 4.3. Consider the convergent sequence $\{x_n\}_n \subset \mathbb{R}$ having limit $x \in \mathbb{R}$. Then, the sequence of corresponding Dirac distributions $\{\delta_{x_n}\}_n \subset \mathscr{M}$ is weakly $\mathscr{C}_B$-convergent to δ_x. However, strong convergence does not hold because

$$\|\delta_{x_n} - \delta_x\|_{TV} = \sup_{g\in\mathscr{C},\, |g|\le 1} |g(x_n) - g(x)| = 2 \ne 0, \quad \forall\, n \in \mathbb{N}.$$

Provided that $(\mathscr{D}, v)$ is a Banach base, the Banach-Steinhaus theorem (see, for example, Dunford [7]) can be applied to a sequence $\{\mu_n\}_n$ of measures, which allows us to deduce v-norm boundedness of $\{\mu_n\}_n$ on $[\mathscr{D}]_v$ from $[\mathscr{D}]_v$-convergence of μ_n. The precise statement is provided in Lemma 4.1.

Lemma 4.1. *Let $(\mathscr{D}, v)$ be a Banach base and $\{\mu_n\}_n \subset \mathscr{M}_v$ be a $[\mathscr{D}]_v$-convergent sequence with finite limit μ, i.e., $\|\mu\|_v < \infty$. Then, it holds that*

$$\sup_{n\in\mathbb{N}} \|\mu_n\|_v < \infty.$$

Proof. Under the assumption of Lemma 4.1, the set $\{\mu_n \mid n \in \mathbb{N}\}$ is bounded in the weak sense, i.e., for each $g \in [\mathscr{D}]_v$, the set $\{\int g\, d\mu_n \mid n \in \mathbb{N}\}$ is bounded in $\mathbb{R}$. The claim then follows from the Banach-Steinhaus theorem (see, for example, Dunford [7]). □

4.3. Product spaces. Recall that $\mathbb{S}$ denotes a separable complete metric space endowed with its Borel field $\mathscr{S}$. Let $\mathscr{T}$ be another separable complete metric space endowed with its Borel field $\mathscr{T}$. We denote by $\sigma(\mathscr{S} \times \mathscr{T})$ the σ-field generated by the product $\mathscr{S} \times \mathscr{T}$ on $\mathbb{S} \times \mathbb{T}$. Let $(\mathscr{D}(\mathbb{S}), v)$ and $(\mathscr{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. The product of $(\mathscr{D}(\mathbb{S}), v)$ and $(\mathscr{D}(\mathbb{T}), u)$, denoted by $(\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T}), v \otimes u)$, is defined as follows:

$$\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T}) = \{g\colon \mathbb{S} \times \mathbb{T} \to \mathbb{R}\colon g(s, \cdot) \in \mathscr{D}(\mathbb{T}),\ g(\cdot, t) \in \mathscr{D}(\mathbb{S}),\ \forall\, s \in \mathbb{S},\ t \in \mathbb{T}\}, \tag{11}$$

and

$$v \otimes u\colon \mathbb{S} \times \mathbb{T} \to \mathbb{R}\colon (v \otimes u)(s, t) = v(s) \cdot u(t), \quad \forall\, s \in \mathbb{S}, \quad t \in \mathbb{T}. \tag{12}$$

Condition (11) imposes no restriction in applications, which is illustrated in Example 4.4.

EXAMPLE 4.4. We revisit the Banach bases introduced in Example 4.2.

- Let $g \in \mathscr{C}(\mathbb{S} \times \mathbb{T})$, then $g(s, \cdot) \in \mathscr{C}(\mathbb{T})$ for all $s \in \mathbb{S}$ and $g(\cdot, t) \in \mathscr{C}(\mathbb{S})$ for all $t \in \mathbb{T}$. In addition, it holds that

$$\mathscr{C}(\mathbb{S} \times \mathbb{T}) \subset \mathscr{C}(\mathbb{S}) \otimes \mathscr{C}(\mathbb{T}). \tag{13}$$

- Let $g \in \mathscr{F}(\mathbb{S} \times \mathbb{T})$, then $g(s, \cdot) \in \mathscr{F}(\mathbb{T})$ for all $s \in \mathbb{S}$ and $g(\cdot, t) \in \mathscr{F}(\mathbb{S})$ for all $t \in \mathbb{T}$. Moreover, it holds that

$$\mathscr{F}(\mathbb{S} \times \mathbb{T}) \subset \mathscr{F}(\mathbb{S}) \otimes \mathscr{F}(\mathbb{T}). \tag{14}$$

- Let $g \in \mathscr{L}^p(\mathbb{S} \times \mathbb{T}, \{\mu_\theta \times \nu_\theta \colon \theta \in \Theta\})$ for some $p \geq 1$, then $g(s, \cdot) \in \mathscr{L}^p(\mathbb{T}, \{\nu_\theta \colon \theta \in \Theta\})$ for all $s \in \mathbb{S}$ and $g(\cdot, t) \in \mathscr{L}^p(\mathbb{S}, \{\mu_\theta \colon \theta \in \Theta\})$ for all $t \in \mathbb{T}$ (for a proof, use Fubini's theorem). Moreover, it holds that

$$\mathscr{L}^p(\mathbb{S} \times \mathbb{T}, \{\mu_\theta \times \nu_\theta \colon \theta \in \Theta\}) \subset \mathscr{L}^p(\mathbb{S}, \{\mu_\theta \colon \theta \in \Theta\}) \otimes \mathscr{L}^p(\mathbb{T}, \{\nu_\theta \colon \theta \in \Theta\}). \tag{15}$$

The next result shows that products of Banach bases are again Banach bases, where the above definitions are extended to the general case in the obvious way.

THEOREM 4.1. *Let $(\mathscr{D}(\mathbb{S}_i), v_i,)$ be Banach bases for $1 \leq i \leq k$. Then, $(\mathscr{D}(\mathbb{S}_1) \otimes \cdots \otimes \mathscr{D}(\mathbb{S}_k), v_1 \otimes \cdots \otimes v_k)$ is a Banach base on $\mathbb{S}_1 \times \cdots \times \mathbb{S}_k$. Moreover, if $g \in [\mathscr{D}(\mathbb{S}_1) \otimes \cdots \otimes \mathscr{D}(\mathbb{S}_k)]_{v_1 \otimes \cdots \otimes v_k}$, then*

$$\forall\, 1 \leq i \leq k \colon g(s_1, \ldots, s_{i-1}, \cdot, s_{i+1}, \ldots, s_k) \in [\mathscr{D}(\mathbb{S}_i)]_{v_i}$$

for all $s_j \in \mathbb{S}_j$, $1 \leq j \leq k$, and $j \neq i$.

PROOF. The proof follows by finite induction with respect to k and we only provide a proof for the case $k = 2$. More precisely, we prove the following: Let $(\mathscr{D}(\mathbb{S}), v)$ and $(\mathscr{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. Then, $(\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T}), v \otimes u)$ is a Banach base on the product space $\mathbb{S} \times \mathbb{T}$. Moreover, if $g \in [\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v \otimes u}$, then $g(s, \cdot) \in \mathscr{D}(\mathbb{T})$ for all $s \in \mathbb{S}$ and $g(\cdot, t) \in \mathscr{D}(\mathbb{S})$ for all $t \in \mathbb{T}$.

We verify the conditions in Definition 4.1. It is immediate that $\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})$ is a linear space, satisfying

$$\mathscr{C}_B(\mathbb{S} \times \mathbb{T}) \subset \mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T}) \subset \mathscr{F}(\mathbb{S} \times \mathbb{T}).$$

For the second part, one proceeds as follows. First, let $g \in [\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v \otimes u}$. It holds that

$$\sup_{t \in \mathbb{T}} \frac{\|g(\cdot, t)\|_v}{u(t)} = \sup_{t \in \mathbb{T}} \sup_{s \in \mathbb{S}} \frac{|g(s, t)|}{v(s) \cdot u(t)} \leq \sup_{(s, t)} \frac{|g(s, t)|}{v(s) \cdot u(t)} = \|g\|_{v \otimes u} < \infty. \tag{16}$$

Thus, for all $t \in \mathbb{T}$, we have $\|g(\cdot, t)\|_v \leq \|g\|_{v \otimes u} \cdot u(t) < \infty$, which means that $g(\cdot, t) \in [\mathscr{D}(\mathbb{S})]_v$. By symmetry, we obtain $g(s, \cdot) \in [\mathscr{D}(\mathbb{T})]_u$ for all $s \in \mathbb{S}$.

Next, we show that $[\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v \otimes u}$ is a Banach space w.r.t. $v \otimes u$-norm. To this end, let $\{g_n\}_n$ be a Cauchy sequence in $[\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v \otimes u}$. That is, for each $\epsilon > 0$, there exists a rank $n_\epsilon \geq 1$ such that for all $j, k \geq n_\epsilon$, it holds that $\|g_j - g_k\|_{v \otimes u} \leq \epsilon$. Inserting now $g = g_j - g_k$ in (16), one obtains

$$\forall\, t \in \mathbb{T}, j, k \geq n_\epsilon \colon \|g_j(\cdot, t) - g_k(\cdot, t)\|_v \leq \|g_j - g_k\|_{v \otimes u} \cdot u(t) \leq \epsilon \cdot u(t).$$

Hence, for all $t \in \mathbb{T}$, $\{g_n(\cdot, t)\}_n$ is a Cauchy sequence in the Banach space $[\mathscr{D}(\mathbb{S})]_v$ and thus is convergent to some limit $\bar{g}^t \in [\mathscr{D}(\mathbb{S})]_v$. Using again a symmetry argument, we deduce that for all $s \in \mathbb{S}$, the sequence $\{g_n(s, \cdot)\}_n$ converges to some ${}^s\bar{g} \in [\mathscr{D}(\mathbb{T})]_u$. Hence, for fixed $(s, t) \in \mathbb{S} \times \mathbb{T}$, we have ${}^s\bar{g}(t) = \lim_{n \to \infty} g_n(s, t) = \bar{g}^t(s)$ and we define $\bar{g} \in \mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})$ as follows:

$$\bar{g}(s, t) = {}^s\bar{g}(t) = \bar{g}^t(s). \tag{17}$$

Finally, we prove that $\bar{g}$ is the $v \otimes u$-norm limit of the sequence $\{g_n\}_n$, which, in particular, will show that $\bar{g} \in [\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v \otimes u}$. Choosing $\epsilon > 0$ and $n_\epsilon \geq 1$ s.t. for all $j, k \geq n_\epsilon$, we have $\|g_j - g_k\|_{v \otimes u} < \epsilon$. More explicitly,

$$\forall\, j, k \geq n_\epsilon, \quad s \in \mathbb{S}, \quad t \in \mathbb{T} \colon |g_j(s, t) - g_k(s, t)| < \epsilon \cdot v(s) u(t).$$

Because $\bar{g}$ is the pointwise limit of the sequence $\{g_n\}_n$, letting $k \to \infty$ yields

$$\forall\, j \geq n_\epsilon, \quad s \in \mathbb{S}, \quad t \in \mathbb{T} \colon |g_j(s, t) - \bar{g}(s, t)| \leq \epsilon \cdot v(s) u(t),$$

i.e., $\|g_j - \bar{g}\|_{v \otimes u} \leq \epsilon$ for $j \geq n_\epsilon$. Because ϵ was arbitrarily chosen, it follows that $\lim_{n \to \infty} \|g_n - \bar{g}\|_{v \otimes u} = 0$. In addition, for $j = n_\epsilon$, we have

$$\|\bar{g}\|_{v \otimes u} \leq \|\bar{g} - g_{n_\epsilon}\|_{v \otimes u} + \|g_{n_\epsilon}\|_{v \otimes u} \leq \epsilon + \|g_{n_\epsilon}\|_{v \otimes u} < \infty,$$

i.e., $\bar{g} \in [\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v \otimes u}$, which completes the proof. $\square$

For $\mu \in \mathcal{M}(\mathcal{S})$, $\nu \in \mathcal{M}(\mathcal{T})$, we denote their product by $\mu \times \nu \in \mathcal{M}(\sigma(\mathcal{S} \times \mathcal{T}))$. We conclude this section with a technical result (to be used later on) that shows that the product measure is (strongly) continuous with respect to its components.

Lemma 4.2. *For $\mu \in \mathcal{M}(\mathcal{S})$ and $\nu \in \mathcal{M}(\mathcal{T})$, it holds that*

$$\|\mu \times \nu\|_{v\otimes u} \le \|\mu\|_v \, \|\nu\|_u. \tag{18}$$

In particular, if $\mu \in \mathcal{M}_v(\mathcal{S})$ and $\nu \in \mathcal{M}_u(\mathcal{T})$, then $\mu \times \nu \in \mathcal{M}_{v\otimes u}(\sigma(\mathcal{S} \times \mathcal{T}))$.

Proof. Let $\mu = \mu^+ - \mu^-$ and $\nu = \nu^+ - \nu^-$ be the Hahn-Jordan decompositions of μ and ν, respectively. Then,

$$\mu \times \nu = (\mu^+ \times \nu^+ + \mu^- \times \nu^-) - (\mu^+ \times \nu^- + \mu^- \times \nu^+)$$

is a decomposition of $\mu \times \nu$ and the minimality property[7] of Hahn-Jordan decomposition ensures that

$$(\mu \times \nu)^+ \le \mu^+ \times \nu^+ + \mu^- \times \nu^-; \quad (\mu \times \nu)^- \le \mu^+ \times \nu^- + \mu^- \times \nu^+.$$

Thus, according to (9), it holds that (use Fubini for the equality below)

$$\|\mu \times \nu\|_{v\otimes u} \le \int (v \otimes u)(s, z)\left[(\mu^+ + \mu^-) \times (\nu^+ + \nu^-)\right](ds, dz) = \|\mu\|_v \, \|\nu\|_u,$$

which establishes (18). $\square$

5. Differentiability. In this section, we discuss two concepts of differentiability of probability measures based on the types of convergence on $\mathcal{M}$ presented in §§3 and 4 (weak and strong). Particular attention will be paid to weak differentiation because it is a less restrictive condition, though still nice results can be obtained. We conclude the section with a brief note on the class of *truncated distributions* that arise frequently in applications.

5.1. The concept of measure-valued differentiation.

Definition 5.1. Let $(\mathcal{D}, v)$ be a Banach base on $\mathbb{S}$. We say that the mapping $\mu_*\colon \Theta \to \mathcal{M}_v$ is *weakly $[\mathcal{D}]_v$-differentiable* at θ (or μ_θ is weakly differentiable, for short) if there exists $\mu'_\theta \in \mathcal{M}_v$ such that

$$g \in [\mathcal{D}]_v\colon \lim_{\xi \to 0} \frac{1}{\xi}\left(\int g(s)\mu_{\theta+\xi}(ds) - \int g(s)\mu_\theta(ds)\right) = \int g(s)\mu'_\theta(ds). \tag{19}$$

If the left-hand side of the above equation equals zero for all $g \in [\mathcal{D}]_v$, then we say that the weak $[\mathcal{D}]_v$-derivative of μ_θ is not significant. Moreover, if μ_θ is $[\mathcal{D}]_v$-differentiable, then any triplet $(c_\theta, \mu_\theta^+, \mu_\theta^-)$ with $c_\theta \in \mathbb{R}$ and $\mu_\theta^\pm \in \mathcal{M}^1$ satisfying

$$\forall\, g \in [\mathcal{D}]_v\colon \int g(s)\mu'_\theta(ds) = c_\theta\left(\int g(s)\mu_\theta^+(ds) - \int g(s)\mu_\theta^-(ds)\right)$$

is called a weak $[\mathcal{D}]_v$-derivative of μ_θ, and we write in slight abuse of notation $\mu'_\theta = (c_\theta, \mu_\theta^+, \mu_\theta^-)$. If μ'_θ is not significant, we set $\mu'_\theta = (1, \mu_\theta, \mu_\theta)$.

Higher-order derivatives can be introduced in the same way. More precisely, we say that μ_θ is n-times weakly $[\mathcal{D}]_v$-differentiable at θ (or μ_θ is n-times weakly $[\mathcal{D}]_v$-differentiable, for short) if there exist $\mu_\theta^{(n)} \in \mathcal{M}_v$ such that

$$\forall\, g \in [\mathcal{D}]_v\colon \frac{d^n}{d\theta^n}\int g(s)\mu_\theta(ds) = \int g(s)\mu_\theta^{(n)}(ds).$$

Consequently, we denote a nth order $[\mathcal{D}]_v$-derivative by $(c_\theta^{(n)}, \mu_\theta^{(n,+)}, \mu_\theta^{(n,-)})$ with $c_\theta^{(n)} \in \mathbb{R}$ and $\mu_\theta^{(n,\pm)} \in \mathcal{M}^1$.

Remark 5.1. Differentiability of probability measures in the weak sense as defined in Definition 5.1 was introduced by Pflug for $\mathcal{D} = \mathcal{C}_B$; see Pflug [28] for an early reference and the monograph (Pflug [27]) for a thorough treatment of $\mathcal{C}_B$-derivatives. Other early traces are Kushner and Vázquez-Abad [22, 23]. Heidergott and Vàzquez-Abad [13] extended this concept to general $\mathcal{D}$-differentiability and showed that $\mathcal{D}$-derivatives yield efficient unbiased gradient estimators. A recent result in this line of research shows that $\mathcal{D}$-derivative gradient estimators can outperform single-run estimators such as infinitesimal perturbation analysis (see Heidergott et al. [18]).

[7] If $\mu = \mu^+ - \mu^-$ is the Hahn-Jordan decomposition of μ, and $\mu = \kappa^+ - \kappa^-$ is another decomposition of μ such that $\kappa^+, \kappa^- \in \mathcal{M}^+$, then we have $\kappa^+ - \mu^+, \kappa^- - \mu^- \in \mathcal{M}^+$.

Remark 5.2. For any $[\mathscr{D}]_v$-differentiable probability measure μ_θ, an instance of the nth order $[\mathscr{D}]_v$-derivative can be obtained via the Hahn-Jordan decomposition of $\mu_\theta^{(n)}$ (see §3). It is worth noting that weak derivatives can be computed in a straightforward way if it holds that $\mu_\theta(dx) = f(x,\theta)\cdot\mu(dx)$, $\forall\,\theta\in\Theta$, i.e., if μ_θ has a density $f(\cdot,\theta)$ with respect to μ. Then, for each $n\geq 1$ we have

$$\forall\, g\in[\mathscr{D}]_v\colon\ \frac{d^n}{d\theta^n}\int g(x)f(x,\theta)\mu(dx) = \int g(x)\frac{d^n}{d\theta^n}f(x,\theta)\mu(dx), \tag{20}$$

provided that $f(x,\cdot)$ is n-times differentiable at θ for all $x\in S$ and interchanging differentiation and integral is justified. Thus,

$$\mu_\theta^{(n)}(dx) = \frac{d^n f(x,\theta)}{d\theta^n}\cdot\mu(dx),$$

and a weak derivative can be easily computed by considering the positive and the negative parts of $d^n f(\cdot,\theta)/d\theta^n$.

We illustrate the concept of weak differentiability with three basic families of distributions. More examples can be found in Pflug [27].

Example 5.1. Let $\mathbb{S}=[0,\infty)$ with the usual topology and $\Theta=[a,b]\subset[0,\infty)$. Choose $\mu_\theta(dx)=\theta\exp(-\theta x)\cdot\lambda(dx)$, where λ denotes the Lebesgue measure on $\mathbb{S}$. Moreover, let $v_p(s)=1+s^p$ for some $p\in\mathbb{N}$.

Then, for all $n,p\geq 1$, μ_θ is n-times $\mathscr{D}_{v_p}$-differentiable. Higher-order derivatives can be computed by differentiating the density $f(x,\theta)=\theta\exp(-\theta x)$ in the classical sense (see Remark 5.2). More specifically, one obtains, for $n\geq 1$,

$$\mu_\theta^{(n)}(dx) = (-1)^n x^{n-1}\exp(-\theta x)(\theta x - n)\lambda(dx).$$

Furthermore, if we denote by $\gamma(n,\theta)$ the *gamma*-(n,θ)-distribution, i.e., the convolution of n exponential distributions with rate θ, then we have

$$\mu_\theta^{(n)} = \begin{cases}\left(\dfrac{n!}{\theta^n},\gamma(n,\theta),\gamma(n+1,\theta)\right), & \text{for } n \text{ odd,}\\[2ex] \left(\dfrac{n!}{\theta^n},\gamma(n+1,\theta),\gamma(n,\theta)\right), & \text{for } n \text{ even.}\end{cases},\quad n\geq 1.$$

Example 5.2. Let $\mathbb{S}=[0,\infty)$. Denote by ψ_θ the uniform distribution on the interval $(0,\theta)$ for $\theta\in(0,b]$, $b>0$ and denote by δ_θ the Dirac distribution with point mass θ. Take as $\mathscr{D}$ the set $\mathscr{C}(\mathbb{S})$. Because the density $\theta^{-1}\mathbb{I}_{(0,\theta)}(x)$ is not differentiable w.r.t. θ, we calculate the weak derivative directly. Thus, by definition, for each g continuous at θ we have

$$\int g(s)\psi_\theta'(ds) = \lim_{\xi\to 0}\frac{1}{\xi}\left(\frac{1}{\theta+\xi}\int_0^{\theta+\xi} g(s)ds - \frac{1}{\theta}\int_0^\theta g(s)ds\right),$$

which yields

$$\forall\, g\in\mathscr{C}(\mathbb{S})\colon\ \int g(s)\psi_\theta'(ds) = \frac{1}{\theta}g(\theta) - \frac{1}{\theta^2}\int_0^\theta g(s)ds.$$

Thus, $\psi_\theta'=(1/\theta)\delta_\theta-(1/\theta)\psi_\theta$ or, in triplet representation, $\psi_\theta'=(\theta^{-1},\delta_\theta,\psi_\theta)$. Higher-order derivatives of ψ_θ do not exist. This stems from the fact that δ_θ fails to be weakly $\mathscr{D}$-differentiable for any sensible set $\mathscr{D}$.

Example 5.3. Let $\mathbb{S}=\{x_1,x_2\}$ with the discrete topology $\Theta=[0,1)$ and set

$$\forall\,\theta\in\Theta\colon\ \mu_\theta = (1-\theta)\cdot\delta_{x_1} + \theta\cdot\delta_{x_2},$$

where δ_x denotes the Dirac distribution with total mass at point x. To avoid trivialities, we assume $x_1\neq x_2$. Then, it holds for each $g\colon\mathbb{S}\to\mathbb{R}$ that

$$\frac{d}{d\theta}\int g(x)\mu_\theta(dx) = \frac{d}{d\theta}((1-\theta)g(x_1)+\theta g(x_2)) = g(x_2)-g(x_1).$$

Obviously, this means that $\mu_\theta'=\delta_{x_2}-\delta_{x_1}$ so that $c_\theta=1$, $\mu_\theta^+=\delta_{x_2}$ and $\mu_\theta^-=\delta_{x_1}$. Moreover, higher-order derivatives exist but are not significant in this situation and we set $\mu_\theta^{(n)}=(1,\mu_\theta,\mu_\theta)$ for $n\geq 2$.

Strong differentiability is introduced in an obvious way by replacing in (19) weak convergence with the strong one. More precisely, μ_θ is called *strongly differentiable*, with derivative μ'_θ, if

$$\lim_{\xi\to 0}\left\|\frac{\mu_{\theta+\xi}-\mu_\theta}{\xi}-\mu'_\theta\right\|_v=0.$$

Note that the "strong" derivative μ'_θ is also called *Fréchet derivative* on the space $\mathcal{M}_v$ in the literature.

Strong v-norm differentiability of μ_θ implies weak $[\mathcal{D}]_v$-differentiability. The converse is, however, not true, which stems from the fact that

$$\lim_{\xi\to 0}\frac{1}{\xi}\left|\left(\int g(s)\mu_{\theta+\xi}(ds)-\int g(s)\mu_\theta(ds)\right)-\int g(s)\mu'_\theta(ds)\right|=0,\quad \forall\, g\in[\mathcal{D}]_v \tag{21}$$

does not, in general, imply that

$$\lim_{\xi\to 0}\sup_{\|g\|_v\le 1}\left|\frac{1}{\xi}\left(\int g(s)\mu_{\theta+\xi}(ds)-\int g(s)\mu_\theta(ds)\right)-\int g(s)\mu'_\theta(ds)\right|=0. \tag{22}$$

Example 5.4 illustrates this fact.

Example 5.4. Consider the uniform distribution ψ_θ on $(0,\theta)$. In Example 5.2, we have shown that ψ_θ is weakly $\mathcal{D}$-differentiable for $\mathcal{D}=\mathcal{C}$, and that its weak derivative satisfies

$$\psi'_\theta=\frac{1}{\theta}\delta_\theta-\frac{1}{\theta}\psi_\theta.$$

Hence, (21) holds true for $\psi_\theta=\mu_\theta$ for $\xi\neq 0$. Let $v(s)=s^p$. Then, it holds that

$$\sup_{\|g\|_v\le 1}\left|\frac{1}{\xi}\left(\int g(s)\psi_{\theta+\xi}(ds)-\int g(s)\psi_\theta(ds)\right)-\int g(s)\psi'_\theta(ds)\right|\ge\theta^{p-1},$$

which violates (22). The uniform distribution on $(0,\theta)$ is weakly but not strongly differentiable.

Weak and strong properties of measure-valued mappings are related as follows.

Theorem 5.1. *Let $\mu_*\colon\Theta\to\mathcal{M}_v$ be a $[\mathcal{D}]_v$-continuous measure-valued mapping such that μ_θ is $[\mathcal{D}]_v$-differentiable. Then, for each closed neighborhood V of 0 such that $\theta+\xi\in\Theta$ for each $\xi\in V$, there exists some $M>0$ such that*

$$\forall\,\xi\in V\colon\ \|\mu_{\theta+\xi}-\mu_\theta\|_v\le M|\xi|.$$

In words, μ_θ is v-norm locally Lipschitz continuous.

A sufficient condition for strong differentiability is provided in Theorem 5.2.

Theorem 5.2. *If $\mu_*\colon\Theta\to\mathcal{M}$ is weakly $[\mathcal{D}]_v$-differentiable on Θ such that μ'_* is v-norm continuous, then μ_* is strongly v-norm differentiable on Θ.*

For a proof of the above two results, we refer to Heidergott et al. [16].

Note that Theorem 5.2 implies that the exponential distribution in Example 5.1 is strongly differentiable. There is a trade-off between $\mathcal{D}$ (resp. $[\mathcal{D}]_v$) and the class of probability measures that are $\mathcal{D}$ (resp. $[\mathcal{D}]_v$) differentiable. To see this, consider the Banach bases introduced in Example 4.2. Roughly speaking, $[\mathcal{F}]_v$-differentiability is the most restrictive condition because it requires that $\mu_\theta(A)$ is differentiable for all $A\in\mathcal{S}$; this is the definition used by Kushner and Vázquez-Abad in Kushner and Vázquez-Abad [22]. For instance, the uniform distribution fails to be $[\mathcal{F}]_v$-differentiable (recall that continuity of the test function at θ is required) whereas the exponential and the Bernoulli distribution are. Indeed, if ψ_θ denotes the uniform distribution on the interval $(0,\theta)$ defined in Example 5.2 and we let $A=[0,x]$ for some $x>0$, then we have

$$\psi_\theta(A)=\frac{1}{\theta}\min\{x,\theta\},$$

which is not differentiable at $\theta=x$. On the other hand, for $v\equiv 1$, $[\mathcal{C}]_v$-differentiability is the least restrictive condition because it only requires weak convergence; this is the derivative introduced by Pflug in Pflug [28]. The uniform, the exponential, and the Bernoulli distribution are $[\mathcal{C}]_{v\equiv 1}$-differentiable. Only the Dirac distribution in θ fails to be $[\mathcal{C}]_{v\equiv 1}$-differentiable.

5.2. A note on truncated distributions. The class of truncated distributions is a more general example of weakly but not strongly differentiable distributions, and it is interesting especially because of the form of their weak derivative. In particular, it will turn out that the uniform distribution presented in Example 5.2 belongs to this class.

Let X be a real-valued random variable (r.v.) and let $-\infty \le a < b \le \infty$ be such that $\mathbb{P}(\{a < X < b\}) > 0$. By a *truncation* μ of the distribution of X, we mean the conditional distribution of X on the event $\{a < X < b\}$; in formula:

$$\forall A\colon \mu(A) \stackrel{\text{def}}{=} \frac{\mathbb{P}(A \cap \{a < X < b\})}{\mathbb{P}(\{a < X < b\})}.$$

If X has a probability density ρ, then the mapping

$$\forall x \in \mathbb{R}\colon f(x) \stackrel{\text{def}}{=} \frac{\rho(x)}{\int_a^b \rho(s)\,ds} \cdot \mathbb{I}_{(a,\,b)}(x) \tag{23}$$

is the probability density of a truncated distribution.

Remark 5.3. Note that f as defined by (23) is still a probability density if we only require that ρ is a nonnegative integrable function on (a, b) and not necessarily a density on $\mathbb{R}$.

Example 5.5. In the following, we provide several examples.

(i) Letting $\rho(x) = x$, $a = 0$, and $b < \infty$ in (23), one recovers the uniform distribution on $(0, b)$, cf. Example 5.2.

(ii) Letting $\rho(x) = x^{-\beta}$ for some $\beta > 1$, $a > 0$, and $b = \infty$ in (23), one obtains the Pareto distribution with density

$$f(x) = \beta a^{\beta} x^{-(\beta+1)} \mathbb{I}_{(a,\,\infty)}(x).$$

(iii) For $\rho(x) = e^{-\lambda x}$ for some $\lambda > 0$ and $b = \infty$, one obtains the shifted exponential distribution[8] with density

$$f(x) = e^{-\lambda(x-a)} \mathbb{I}_{(a,\,\infty)}(x).$$

Truncated distributions arise naturally in applications. Indeed, consider a constant ζ modeling a planned traveling time in a transportation network. Then, it is quite common to add a normally distributed noise, say Z, to ζ to model some intrinsic randomness (see Heidergott and Vázquez-Abad [14]). However, traveling times are bounded from below by the minimal traveling that is physical possible. Denote this lower bound by θ. Note that θ is a design parameter as it depends on safety regulations, the type of rolling stock, and the conditions of the track. It is important to ensure that $\mathbb{P}(\zeta + Z < \theta) = 0$ (so that the perturbed traveling times are still feasible) and that one considers a truncated version of Z. In other words, the distribution of $\zeta + Z$ is conditioned on the event $\zeta + Z \in (\theta, \infty)$ for $\zeta > \theta > 0$. In the setting of this section, the truncated density (23) is considered with $a = \theta$ and $b = \infty$; more formally, a parametric family of left-sided truncated distributions μ_θ is introduced with density given

$$f_\theta(x) = \frac{\rho(x)}{\int_\theta^\infty \rho(x)\,dx} \mathbb{I}_{(\theta,\,\infty)}(x). \tag{24}$$

The remainder of this section is devoted to computation of the weak derivative of $\mu_\theta(dx) = f_\theta(x)\,dx$ in accordance with Definition 5.1. To this end, let $v \in \mathscr{C}^+(\mathbb{R})$ be such that

$$\int v(x)\rho(x)\,dx < \infty.$$

For $g \in [\mathscr{C}]_v$, proceed as follows:

$$\frac{d}{d\theta} \frac{\int_\theta^\infty g(x)\rho(x)\,dx}{\int_\theta^\infty \rho(x)\,dx} = \frac{\rho(\theta)\int_\theta^\infty g(x)\rho(x)\,dx}{\left(\int_\theta^\infty \rho(x)\,dx\right)^2} - \frac{g(\theta)\rho(\theta)}{\int_\theta^\infty \rho(x)\,dx} = \frac{\rho(\theta)}{\int_\theta^\infty \rho(x)\,dx}\left(\int g(x)\mu_\theta(dx) - \int g(x)\delta_\theta(dx)\right).$$

Consequently, we can write the derivative as $\mu'_\theta = c_\theta(\mu_\theta - \delta_\theta)$, where

$$c_\theta \stackrel{\text{def}}{=} \frac{\rho(\theta)}{\int_\theta^\infty \rho(x)\,dx}.$$

Hence, the derivative of a left-sided truncated distribution can be represented as the rescaled difference between the original truncated distribution and the Dirac distribution; higher-order derivatives do not exist because the Dirac distribution is not differentiable (see Example 5.2).

[8] The fact can be also derived from the memoryless property of the exponential distribution.

6. Differentiability of product measures. In this section, we will establish sufficient conditions for (higher-order) weak differentiability of product measures. For the ease of reading, we will first provide an analysis of the product of two probability measures (see §6.1). The results for general products of probability measures are presented in §6.2.

6.1. Products of two probability measures. In this section, we will establish sufficient conditions for weak differentiability of the product of two probability measures. As it will turn out, the product of weakly differentiable probability measures is again weakly differentiable, provided that the functional spaces are Banach bases. The precise statement is given in Theorem 6.1. Recall that $\sigma(\mathcal{S}\times\mathcal{T})$ denotes the σ-field generated by the product $\mathcal{S}\times\mathcal{T}$ on $\mathbb{S}\times\mathbb{T}$.

THEOREM 6.1. *Let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. Assume further that $\mu_\theta \in \mathcal{M}_1(\mathbb{S})$ is $[\mathcal{D}(\mathbb{S})]_v$-differentiable and $\nu_\theta \in \mathcal{M}_1(\mathbb{T})$ is $[\mathcal{D}(\mathbb{T})]_u$-differentiable and denote their weak derivatives by μ'_θ and ν'_θ, respectively. Then, the product measure $\mu_\theta\times\nu_\theta \in \mathcal{M}_1(\sigma(\mathcal{S}\times\mathcal{T}))$ is $[\mathcal{D}(\mathbb{S})\otimes\mathcal{D}(\mathbb{T})]_{v\otimes u}$-differentiable and it holds that* (*compare with classical analysis*)

$$(\mu_\theta\times\nu_\theta)' = (\mu'_\theta\times\nu_\theta) + (\mu_\theta\times\nu'_\theta).$$

PROOF. For ξ such that $\theta+\xi\in\Theta$, set

$$\bar{\mu}_\xi = \frac{\mu_{\theta+\xi}-\mu_\theta}{\xi} - \mu'_\theta, \quad \bar{\nu}_\xi = \frac{\nu_{\theta+\xi}-\nu_\theta}{\xi} - \nu'_\theta.$$

By hypothesis, we have $\bar{\mu}_\xi \overset{[\mathcal{D}]_v}{\Longrightarrow} \mathbf{0}$ and $\bar{\nu}_\xi \overset{[\mathcal{D}]_u}{\Longrightarrow} \mathbf{0}$, where $\mathbf{0}$ denotes the null measure. In this notation, the conclusion is equivalent to

$$\xi\cdot(\bar{\mu}_\xi+\mu'_\theta)\times(\bar{\nu}_\xi+\nu'_\theta) + \mu_\theta\times\bar{\nu}_\xi + \bar{\mu}_\xi\times\nu_\theta \overset{[\mathcal{D}]_{v\otimes u}}{\Longrightarrow} \mathbf{0} \quad \text{as } \xi\to 0. \tag{25}$$

Hence, we show that each term on the left side of (25) converges weakly to $\mathbf{0}$. For the first term, applying the Cauchy-Schwartz inequality (10) together with Lemma 4.2 yields

$$\left|\xi\int g(s,t)\big((\bar{\mu}_\xi+\mu'_\theta)\times(\bar{\nu}_\xi+\nu'_\theta)\big)(ds,dt)\right| \le |\xi|\cdot\|g\|_{v\otimes u}\cdot\|\bar{\mu}_\xi+\mu'_\theta\|_v\cdot\|\bar{\nu}_\xi+\nu'_\theta\|_u. \tag{26}$$

Because $\bar{\mu}_\xi+\mu'_\theta \overset{[\mathcal{D}]_v}{\Longrightarrow} \mu'_\theta$ and $\bar{\nu}_\xi+\nu'_\theta \overset{[\mathcal{D}]_u}{\Longrightarrow} \nu'_\theta$, applying Lemma 4.1 yields

$$\sup_{\xi\in V}\|\bar{\mu}_\xi+\mu'_\theta\|_v < \infty \qquad \text{and} \qquad \sup_{\xi\in V}\|\bar{\nu}_\xi+\nu'_\theta\|_u < \infty$$

for a neighborhood V of 0. Letting now $\xi\to 0$ in (26), the conclusion follows. For the second term in (25), note that

$$\int g(s,t)(\mu_\theta\times\bar{\nu}_\xi)(ds,dt) = \int\int g(s,t)\mu_\theta(ds)\,\bar{\nu}_\xi(dt) = \int H_\theta(g,t)\bar{\nu}_\xi(dt),$$

where $H_\theta(g,t) = \int g(s,t)\mu_\theta(ds)$ for all t and for all g. Theorem 4.1 implies that $(\mathcal{D}(\mathbb{S})\otimes\mathcal{D}(\mathbb{T}),\ v\otimes u)$ is a Banach base. Hence, applying the Chauchy-Schwartz inequality yields

$$\frac{|H_\theta(g,t)|}{u(t)} \le \frac{\|g(\cdot,t)\|_v}{u(t)}\|\mu_\theta\|_v \le \|g\|_{v\otimes u}\|\mu_\theta\|_v, \quad \forall\, t\in\mathbb{T},$$

where the second inequality follows from the second part of Theorem 4.1. Consequently, $H_\theta(g,\cdot)\in[\mathcal{D}(\mathbb{T})]_u$ for $g\in[\mathcal{D}(\mathbb{S})\otimes\mathcal{D}(\mathbb{T})]_{v\otimes u}$. We have assumed ν_θ is $[\mathcal{D}(\mathbb{T})]_u$-differentiable, which yields that $\bar{\nu}_\xi \overset{[\mathcal{D}]_u}{\Longrightarrow} \mathbf{0}$. Hence,

$$\lim_{\xi\to 0}\int H_\theta(g,t)\bar{\nu}_\xi(dt) = 0,$$

which shows that the second term in (25) converges weakly to $\mathbf{0}$. The third term can be treated in a similar way, which concludes the proof. □

REMARK 6.1. It is worth noting that the conditions on the functional spaces in Theorem 6.1 are typically satisfied in applications (see Example 4.2).

REMARK 6.2. Choosing $\mathscr{D}(\mathbb{S}) = \mathscr{C}(\mathbb{S})$, $\mathscr{D}(\mathbb{T}) = \mathscr{C}(\mathbb{T})$, $v \equiv 1$, and $u \equiv 1$ in Theorem 6.1, we conclude from (13) that "weak $\mathscr{C}_B$-differentiability is preserved by the product measure." This is asserted in Pflug [27] though no proof is given. In the same vein, taking (13) and (14) into account, we conclude that weak differentiability is preserved by the product measure in both the continuity and measurability paradigm see (Example 4.2).

Inspired by the resemblance of Theorem 6.1 with classical analysis, we proceed to establish the "Leibnitz-Newton" product rule, which extends Theorem 6.1 to higher-order derivatives. The precise statement is as follows:

THEOREM 6.2. *Let $(\mathscr{D}(\mathbb{S}), v)$ and $(\mathscr{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. If μ_θ is n-times $[\mathscr{D}(\mathbb{S})]_v$-differentiable and if ν_θ is n-times $[\mathscr{D}(\mathbb{T})]_u$-differentiable, then the product measure $\mu_\theta \times \nu_\theta \in \mathscr{M}(\sigma(\mathscr{S} \times \mathscr{T}))$ is n-times $[\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v\otimes u}$-differentiable and it holds that*

$$(\mu_\theta \times \nu_\theta)^{(n)} = \sum_{j=0}^{n} \binom{n}{j} \left(\mu_\theta^{(j)} \times \nu_\theta^{(n-j)}\right).$$

PROOF. We proceed by induction over $n \geq 1$. For $n = 1$, the assertion reduces to Theorem 6.1. Assume now that the conclusion holds true for $n \geq 1$. Then,

$$(\mu_\theta \times \nu_\theta)^{(n+1)} = \left(\sum_{j=0}^{n} \binom{n}{j} \left(\mu_\theta^{(j)} \times \nu_\theta^{(n-j)}\right)\right)' = \sum_{j=0}^{n} \binom{n}{j} \left(\mu_\theta^{(j)} \times \nu_\theta^{(n-j)}\right)'.$$

Applying Theorem 6.1 to evaluate the derivatives on the right-hand side, the proof follows as in conventional analysis from basic algebraic calculations. □

6.2. General products. In this section, we address differentiability of n-fold products of probability measures. Theorem 6.3 presents the general formula of the weak differential calculus.

THEOREM 6.3. *For $1 \leq i \leq k$, let $(\mathscr{D}(\mathbb{S}_i), v_i)$ be Banach bases on $\mathbb{S}_i$ such that $\mu_{i,\theta}$ is n-times $[\mathscr{D}(\mathbb{S}_i)]_{v_i}$-differentiable. Then, $\Pi_\theta \stackrel{\text{def}}{=} \mu_{1,\theta} \times \cdots \times \mu_{k,\theta}$ is n-times $[\mathscr{D}(\mathbb{S}_1) \otimes \cdots \otimes \mathscr{D}(\mathbb{S}_k)]_{v_1\otimes\cdots\otimes v_k}$-differentiable on $\mathbb{S}_1 \times \cdots \times \mathbb{S}_k$ and it holds that*

$$\Pi_\theta^{(n)} = \sum_{\tilde{j} \in \mathscr{J}(k,n)} \binom{n}{j_1, \ldots, j_k} \cdot (\mu_{1,\theta})^{(j_1)} \times \cdots \times (\mu_{k,\theta})^{(j_k)}, \tag{27}$$

where

$$\mathscr{J}(k,n) \stackrel{\text{def}}{=} \{\tilde{j} = (j_1, \ldots, j_k)\colon 0 \leq j_i \leq n,\ j_1 + \cdots + j_k = n\}$$

for $k, n \geq 1$.

PROOF. The proof follows from Theorem 4.1 and Theorem 6.2 via finite induction. □

An instance of a derivative of the product measure $\Pi_\theta^{(n)}$ in Theorem 6.3 can be obtained by inserting the appropriate $\mathscr{D}_{v_i}$ derivatives for the measures $\mu_{i,\theta}^{(j_i)}$ and rearranging terms in (27). To present the result, we introduce the following notations. For $\tilde{j} = (j_1, \ldots, j_k) \in \mathscr{J}(k,n)$, we denote by $\xi(\tilde{j})$ the number of nonzero elements of the vector $\tilde{j}$, and by $\mathscr{I}(\tilde{j})$ the set of vectors $\eta \in \{-1, 0, +1\}^k$ such that $\eta_i \neq 0$ if and only if $j_i \neq 0$ and such that the product of all nonzero elements of η equals one, i.e., there is an even number of "-1." For $\eta \in \mathscr{I}(\tilde{j})$, we denote by $\bar{\eta}$ the vector obtained from η by changing the sign of the nonzero element at the highest position. The precise statement is as follows.

COROLLARY 6.1. *Under the conditions put forward in Theorem 6.3, let μ_θ^i have* m*th order $\mathscr{D}_{v_i}$-derivative*

$$\mu_\theta^{(m)} = \left(c_{i,\theta}^{(m)}, \mu_{i,\theta}^{(m,+)}, \mu_{i,\theta}^{(m,-)}\right)$$

for $m \geq 0$ with $c_{i,\theta}^{(0)} = 1$ and $\mu_{i,\theta}^{(0,0)} = \mu_{i,\theta}$. For $n \geq 1$, an instance $\left(\Gamma_\theta^{(n)}, \Pi_\theta^{(n,+)}, \Pi_\theta^{(n,-)}\right)$ of $\Pi_\theta^{(n)}$ is given by

$$\Gamma_\theta^{(n)} = \sum_{\tilde{j} \in \mathscr{J}(k,n)} 2^{\xi(\tilde{j})-1} \binom{n}{j_1, \ldots, j_k} \prod_{i=1}^{k} c_{i,\theta}^{(j_i)},$$

$$\Pi_\theta^{(n,+)} = \sum_{\tilde{j} \in \mathscr{J}(k,n)} \binom{n}{j_1, \ldots, j_k} \frac{\prod_{i=1}^{k} c_{i,\theta}^{(j_i)}}{\Gamma_\theta^{(n)}} \sum_{\eta \in \mathscr{I}(\tilde{j})} \mu_{1,\theta}^{(j_1,\eta_1)} \times \mu_{2,\theta}^{(j_2,\eta_2)} \times \cdots \times \mu_{k,\theta}^{(j_k,\eta_k)},$$

$$\Pi_\theta^{(n,-)} = \sum_{\tilde{j} \in \mathscr{J}(k,n)} \binom{n}{j_1, \ldots, j_k} \frac{\prod_{i=1}^{k} c_{i,\theta}^{(j_i)}}{\Gamma_\theta^{(n)}} \sum_{\eta \in \mathscr{I}(\tilde{j})} \mu_{1,\theta}^{(j_1,\bar{\eta}_1)} \times \mu_{2,\theta}^{(j_2,\bar{\eta}_2)} \times \cdots \times \mu_{k,\theta}^{(j_k,\bar{\eta}_k)},$$

where, for convenience, we identify

$$\mu_{i,\theta}^{(j_k,+1)} = \mu_{i,\theta}^{(j_k,+)},\ \mu_{i,\theta}^{(j_k,-1)} = \mu_{i,\theta}^{(j_k,-)},\ \mu_{i,\theta}^{(0,0)} = \mu_{i,\theta}.$$

Example 6.1. Consider the Banach base $\mathscr{C}_B(\mathbb{S}) = ((\mathscr{C}(\mathbb{S}), v)$ for $v \equiv 1$. Denote the k-fold product of μ_θ by $\Pi_\theta(k)$. Suppose that μ_θ has $\mathscr{C}_B(\mathbb{S})$-derivative $(c_\theta, \mu_\theta^+, \mu_\theta^-)$. Then, by Theorem 6.3, $\Pi_\theta(n)$ is $\mathscr{C}_B(\mathbb{S}^n)$-differentiable and an instance of a $\mathscr{C}_B(\mathbb{S}^n)$-derivative can be obtained from Corollary 6.1. This yields, for any $g \in \mathscr{C}_B(\mathbb{S}^n)$,

$$\frac{d}{d\theta}\int g(s)\Pi_\theta(n, ds) = c_\theta \sum_{j=1}^{n} \biggl(\int g(s,t,u)\Pi_\theta(n-j, ds) \times \mu_\theta^+(dt) \times \Pi_\theta(j-1, du) - \int g(s,t,u)\Pi_\theta(n-j, ds) \times \mu_\theta^-(dt) \times \Pi_\theta(j-1, du)\biggr).$$

Consider the performance function $J_g(\theta)$ defined in (1) with $\mu_{i,\theta} = \mu_\theta$. Let X_θ^+ have distribution μ_θ^+ and let X_θ^- have distribution μ_θ^-. Then, the above derivative representation reads in terms of random variables

$$\frac{d}{d\theta}J_g(\theta) = c_\theta \sum_{j=1}^{n} \mathbb{E}_\theta[g(X_n, \dots, X_{j+1}, X_\theta^+, X_{j-1}, \dots, X_1) - g(X_n, \dots, X_{j+1}, X_\theta^-, X_{j-1}, \dots, X_1)] \tag{28}$$

for any $g \in \mathscr{C}_B(\mathbb{S}^n)$. For example, the expression

$$c_\theta \sum_{j=1}^{n} \bigl(g(X_n, \dots, X_{j+1}, X_\theta^+, X_{j-1}, \dots, X_1) - g(X_n, \dots, X_{j+1}, X_\theta^-, X_{j-1}, \dots, X_1)\bigr)$$

provides an unbiased estimator for the stochastic gradient $(d/(d\theta)J_g(\theta))$.

7. Weak analyticity. In this section, we introduce the concept of weak $[\mathscr{D}]_v$-analyticity for probability measures and we provide results for the radius of convergence of the corresponding Taylor series and weak analyticity of product measures.

Definition 7.1. Let $(\mathscr{D}, v)$ be a Banach base on $\mathbb{S}$. We call the measure-valued mapping $\mu_*\colon \Theta \to \mathscr{M}_v$ weakly $[\mathscr{D}]_v$-analytic at θ or we say that μ_θ is weakly $[\mathscr{D}]_v$-analytic if

- all higher-order $[\mathscr{D}]_v$-derivatives of μ_θ exist,
- there exists a neighborhood V of θ such that, for all ξ satisfying $\theta + \xi \in V$, it holds that

$$\forall\, g \in [\mathscr{D}]_v\colon \int g(s)\mu_{\theta+\xi}(ds) = \sum_{n=0}^{\infty} \frac{\xi^n}{n!} \cdot \int g(s)\mu_\theta^{(n)}(ds). \tag{29}$$

If μ_θ is n times weakly differentiable for $n \geq 0$, the expression $\mathbf{T}_n(\mu, \theta, \cdot)$ defined as

$$\forall\, \xi \in \mathbb{R}\colon \mathbf{T}_n(\mu, \theta, \xi) := \sum_{k=0}^{n} \frac{\xi^k}{k!} \cdot \mu_\theta^{(k)} \tag{30}$$

will be called the nth order Taylor polynomial of μ_* in θ. The nth order Taylor polynomial $\mathbf{T}_n(\mu, \theta, \xi)$ defined by (30) is, in fact, an element in $\mathscr{M}_v$ and defines a linear functional on $[\mathscr{D}]_v$. Therefore, (29) is equivalent to

$$\forall\, \xi;\ \theta + \xi \in V\colon \mathbf{T}_n(\mu, \theta, \xi) \overset{[\mathscr{D}]_v}{\Longrightarrow} \mu_{\theta+\xi}. \tag{31}$$

Moreover, because all higher-order derivatives of μ_θ exist, it follows by Theorem 5.1 that, for each $n \geq 1$, $\mu_\theta^{(n)}$ is strongly continuous and, by Theorem 5.2, we conclude that $\mu_\theta^{(n-1)}$ is strongly differentiable. In particular, it follows that if μ_θ is weakly analytic, then it is strongly differentiable of any order $n \geq 1$.

For fixed $g \in [\mathscr{D}]_v$, the maximal set $D_\theta(g, \mu)$ for which the equality in (29) holds is called the *domain of convergence of the Taylor series*. Note that the domain of convergence $D_\theta(g, \mu)$ of the series in (29) depends on g. Our next result provides a set $D_\theta^v(\mu) \subset \Theta$, where the Taylor series in (29) converges for all $g \in [\mathscr{D}]_v$. The precise statement is as follows.

THEOREM 7.1. *Let $(\mathscr{D}, v)$ be a Banach base on $\mathbb{S}$ such that μ_θ is $[\mathscr{D}]_v$-analytic. Then, for each $g \in [\mathscr{D}]_v$, the Taylor series in (29) converges for all ξ such that $|\xi| < R^v_\theta(\mu)$, where $R^v_\theta(\mu)$ is given by*

$$\frac{1}{R^v_\theta(\mu)} = \limsup_{n\in\mathbb{N}} \left(\frac{\|\mu^{(n)}_\theta\|_v}{n!} \right)^{1/n}. \tag{32}$$

In particular, the set $D^v_\theta(\mu) := \Theta \cap (\theta - R^v_\theta(\mu), \theta + R^v_\theta(\mu))$ satisfies

$$\forall\, g \in [\mathscr{D}]_v\colon\ D^v_\theta(\mu) \subset D_\theta(g, \mu).$$

PROOF. We apply the Cauchy-Hadamard theorem. It follows that the radius of convergence $R_\theta(g, \mu)$ of the Taylor series in (29) is given by

$$\frac{1}{R_\theta(g, \mu)} = \limsup_{n\in\mathbb{N}} \left(\frac{\left|\int g(s)\mu^{(n)}_\theta(ds)\right|}{n!} \right)^{1/n},$$

i.e., the series converges for $|\xi| < R_\theta(g, \mu)$ and it suffices to show that

$$\forall\, g \in [\mathscr{D}]_v\colon\ R^v_\theta(\mu) \le R_\theta(g, \mu). \tag{33}$$

This follows from the Cauchy-Schwartz inequality. To see this, note that

$$\left| \int g(s)\mu^{(n)}_\theta(ds) \right|^{\frac{1}{n}} \le \left(\|g\|_v \cdot \|\mu^{(n)}_\theta\|_v \right)^{1/n},$$

which together with the fact that $\lim_{n\to\infty} \sqrt[n]{\|g\|_v} = 1$ for $g \in [\mathscr{D}]_v$ concludes the proof. □

The nonnegative number $R^v_\theta(\mu)$ is called the $[\mathscr{D}]_v$*-radius of convergence* of μ_θ and the set $D^v_\theta(\mu)$ is called *the* $[\mathscr{D}]_v$*-domain of convergence* of μ_θ. Note that, in general, this is not the maximal set for which the series converges for all $g \in [\mathscr{D}]_v$ because the inequality in (33) may be strict.

The $[\mathscr{D}]_v$-domain of convergence $D^v_\theta(\mu)$ plays an important role in applications. Though Theorem 7.1 shows that, for $|\xi| < R^v_\theta(\mu)$, the sequence of Taylor polynomials $\mathbf{T}_n(\mu, \theta, \xi)$ converges weakly as $n \to \infty$, Theorem 7.2 will show that this convergence is, in fact strong.

THEOREM 7.2. *Let $(\mathscr{D}, v)$ be a Banach base on $\mathbb{S}$ such that μ_θ is $[\mathscr{D}]_v$-analytic with $[\mathscr{D}]_v$-radius of convergence $R^v_\theta(\mu)$. Then,*

$$\forall\, \xi;\ |\xi| < R^v_\theta(\mu)\colon\ \lim_{n\to\infty} \|\mathbf{T}_n(\mu, \theta, \xi) - \mu_{\theta+\xi}\|_v = 0.$$

PROOF. By hypothesis, we have

$$\|\mathbf{T}_n(\mu, \theta, \xi) - \mu_{\theta+\xi}\|_v = \left\| \sum_{k=n+1}^{\infty} \frac{\xi^k}{k!} \cdot \mu^{(k)}_\theta \right\|_v \le \sum_{k=n+1}^{\infty} \frac{|\xi|^k}{k!} \left\|\mu^{(k)}_\theta\right\|_v. \tag{34}$$

Let ξ be such that $|\xi| < R^v_\theta(\mu)$ and choose $\epsilon > 0$ such that $|\xi| + \epsilon < R^v_\theta(\mu)$. Because

$$\frac{1}{R^v_\theta(\mu) - \epsilon} > \frac{1}{R^v_\theta(\mu)} = \limsup_{n\in\mathbb{N}} \left(\frac{\|\mu^{(n)}_\theta\|_v}{n!} \right)^{1/n},$$

it follows that there exists some $n_\epsilon \ge 1$ such that

$$\forall\, k \ge n_\epsilon\colon\ \left(\frac{\|\mu^{(k)}_\theta\|_v}{k!} \right)^{1/k} < \frac{1}{R^v_\theta(\mu) - \epsilon}.$$

Consequently, we conclude from (34) that, for each $n \ge n_\epsilon$, it holds that

$$\|\mathbf{T}_n(\mu, \theta, \xi) - \mu_{\theta+\xi}\|_v \le \sum_{k=n+1}^{\infty} \left(\frac{|\xi|}{R^v_\theta(\mu) - \epsilon} \right)^k \frac{R^v_\theta(\mu) - \epsilon}{R^v_\theta(\mu) - \epsilon - |\xi|} \left(\frac{|\xi|}{R^v_\theta(\mu) - \epsilon} \right)^{n+1} \tag{35}$$

because, by assumption, $|\xi| < R^v_\theta(\mu) - \epsilon$. Therefore, the conclusion follows by letting $n \to \infty$ in (35). □

EXAMPLE 7.1. Let us revisit Example 5.1 and consider the exponential distribution with rate θ denoted by μ_θ. We aim to determine the $[\mathscr{D}]_v$-radius of convergence of μ_θ for $v(x) = 1 + x$, $\forall\, x \geq 0$. Recall that an instance of the nth order derivative $\mu_\theta^{(n)}$ is given by

$$\mu_\theta^{(n)} = \begin{cases} \left(\dfrac{n!}{\theta^n}, \gamma(n, \theta), \gamma(n+1, \theta)\right), & \text{for } n \text{ odd}, \\ \left(\dfrac{n!}{\theta^n}, \gamma(n+1, \theta), \gamma(n, \theta)\right), & \text{for } n \text{ even}, \end{cases}$$

where

$$\gamma(n, \theta) \cdot dx = \frac{\theta^n \cdot x^{n-1}}{(n-1)!} e^{-\theta x} \cdot dx.$$

Consequently, the v-norm $\|\mu_\theta^{(n)}\|_v$ satisfies

$$\left|\int v(x) \mu_\theta^{(n)}(dx)\right| \leq \|\mu_\theta^{(n)}\|_v \leq \frac{n!}{\theta^n} \int v(x) \gamma(n+1, \theta)(dx) + \frac{n!}{\theta^n} \int v(x) \gamma(n, \theta)(dx).$$

Elementary computation shows that, for $p \geq 1$, we have

$$\int x^p\, \gamma(n, \theta)(dx) = \frac{\theta^n}{(n-1)!} \int x^{n+p-1} e^{-\theta x} dx = \frac{1}{\theta^p} \cdot \frac{(n+p-1)!}{(n-1)!}.$$

Hence, for $v(x) = 1 + x$, we obtain the following bounds:

$$\frac{1}{\theta^{n+1}} \leq \frac{\|\mu_\theta^{(n)}\|_v}{n!} \leq \frac{2n + 2\theta + 1}{\theta^{n+1}}.$$

Finally, we obtain

$$\frac{1}{R_\theta^v(\mu)} = \limsup_{n \in \mathbb{N}} \left(\frac{\|\mu_\theta^{(n)}\|_v}{n!}\right)^{1/n} = \frac{1}{\theta}.$$

To show analyticity, we have to show that (29) holds true for $|\Delta| < \theta$. First, we note that the density $f(x, \theta)$ of μ_θ is analytical (in the classical sense) in θ, i.e.,

$$\forall\, x > 0, \quad \forall\, \Delta \in \mathbb{R}\colon f(x, \theta + \Delta) = \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} \frac{d^k}{d\theta^k} f(x, \theta).$$

Hence, (29) is equivalent to

$$\forall\, g \in [\mathscr{F}]_v\colon \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} \int g(x) \frac{d^k}{d\theta^k} f(x, \theta) dx = \int g(x) \sum_{k=0}^{\infty} \frac{\Delta^k}{k!} \frac{d^k}{d\theta^k} f(x, \theta) dx.$$

Fix $g \in [\mathscr{F}]_v$. To apply the dominated convergence theorem, it suffices to show that the function

$$F_\theta(x) \stackrel{\text{def}}{=} \sum_{k=0}^{\infty} \left| g(x) \frac{\Delta^k}{k!} \frac{d^k}{d\theta^k} f(x, \theta) \right|$$

is integrable. Computing the derivatives of $f(x, \theta)$ (see Example 5.1) yields

$$F_\theta(x) \leq |g(x)| \sum_{k=0}^{\infty} \frac{|\Delta|^k}{k!} (\theta x^k + k x^{k-1}) e^{-\theta x} \leq \|g\|_v (\theta + \Delta) v(x) e^{-(\theta - |\Delta|)x}.$$

Because the right-hand side above is obviously integrable for $|\Delta| < \theta$, we conclude that, for $v(x) = 1 + x$, the exponential distribution μ_θ is weakly $[\mathscr{F}]_v$-analytical for $\theta > 0$ and its radius of convergence is $R_\theta^v(\mu) = \theta$. Moreover, this is still true if we replace v by any finite polynomial.

In the following, we will investigate the error of the nth Taylor polynomial $\mathbf{T}_n(\mu, \theta, \xi)$, i.e., we will establish a bound for

$$\|\mathbf{T}_n(\mu, \theta, \xi) - \mu_{\theta+\xi}\|_v. \tag{36}$$

Theorem 7.2 shows that the expression in (36) converges pointwise to zero w.r.t. $\xi \in (-R_\theta^v(\mu), R_\theta^v(\mu))$ as $n \to \infty$, provided that μ_θ is $[\mathscr{D}]_v$-analytical with $[\mathscr{D}]_v$-radius $R_\theta^v(\mu)$. Unfortunately, the result is of rather theoretical value because it can hardly be used to derive a bound on the error term in (36). Theorem 7.3, based on Taylor's theorem, provides more practical bounds under some additional assumptions.

Theorem 7.3. *Let $(\mathscr{D}, v)$ be a Banach base on $\mathbb{S}$. If $\mu_*: \Theta \to \mathscr{M}_v$ is $(n+1)$-times $[\mathscr{D}]_v$-differentiable on $[\theta - b, \theta + b] \subset \Theta$ for some $b > 0$ such that $\mu_*^{(n+1)}$ is weakly $[\mathscr{D}]_v$-continuous on $(\theta - b, \theta + b)$, then*

$$\forall \xi \in (\theta - b, \theta + b)\colon M_n(\xi) \stackrel{\text{def}}{=} \sup_{\zeta \in (\theta - \xi, \theta + \xi)} \left\| \mu_\zeta^{(n+1)} \right\|_v < \infty. \tag{37}$$

In addition, the following error estimate holds true:

$$\forall g \in [\mathscr{D}]_v\colon \left| \int g(s)\mu_{\theta+\xi}(ds) - \sum_{k=0}^{n} \frac{\xi^k}{k!} \int g(s)\mu_\theta^{(k)}(ds) \right| \le \frac{|\xi|^{n+1}}{(n+1)!} \|g\|_v\, M_n(\xi). \tag{38}$$

In particular, the expression in (36) *can be bounded as follows*:

$$\|\mathbf{T}_n(\mu, \theta, \xi) - \mu_{\theta+\xi}\|_v \le \frac{|\xi|^{n+1}}{(n+1)!} M_n(\xi). \tag{39}$$

Proof. First, we show that (37) holds true. Weak continuity of $\mu_*^{(n+1)}$ on $(\theta - b, \theta + b)$ implies that, for each $g \in [\mathscr{D}]_v$, the mapping $\zeta \mapsto \int g(s)\mu_\zeta^{(n+1)}(ds)$ is continuous on the compact interval $[\theta - \xi, \theta + \xi]$ (hence, bounded) and by Banach-Steinhaus theorem (see, for example, Dunford [7]), one concludes (37).

Applying now Taylor's theorem to the real-valued mapping $\zeta \mapsto \int g(s)\mu_\zeta(ds)$, we conclude that, for each $\xi \in (\theta - b, \theta + b)$ and $g \in [\mathscr{D}]_v$, there exists some $\zeta_g \in (\theta - \xi, \theta + \xi)$ such that

$$\int g(s)\mu_{\theta+\xi}(ds) - \sum_{k=0}^{n} \frac{\xi^k}{k!} \int g(s)\mu_\theta^{(k)}(ds) = \frac{\xi^{n+1}}{(n+1)!} \int g(s)\mu_{\zeta_g}^{(n+1)}(ds).$$

By applying the Cauchy-Schwarz inequality in the right-hand side above, one concludes (38).

Finally, letting $\|g\|_v \le 1$ in (38) yields

$$\left| \int g(s)\mu_{\theta+\xi}(ds) - \sum_{k=0}^{n} \frac{\xi^k}{k!} \int g(s)\mu_\theta^{(k)}(ds) \right| \le \frac{|\xi|^{n+1}}{(n+1)!} M_n(\xi).$$

By taking the supremmum w.r.t. $g \in [\mathscr{D}]_v$ in the left-hand side above, one concludes (39), which completes the proof. □

Remark 7.1. Note that if the conditions of Theorem 7.3 are fulfilled for each $n \ge 1$ and if the constants $M_n(\xi)$ can be chosen such that

$$\lim_{n\to\infty} \frac{|\xi|^{n+1}}{(n+1)!} M_n(\xi) = 0,$$

then μ_θ is weakly $[\mathscr{D}]_v$-analytical.

Theorem 7.3 applies to probability measures that fail to be weakly analytical. We conclude this section with a probability measure that is only twice-continuous weakly differentiable. This measure serves as an example of a distribution that satisfies the conditions put forward in Theorem 7.3 but fails to satisfy the conditions put forward in Theorem 7.2.

Example 7.2. Let $T > 0$ be fixed. Consider the following family of probability distributions:

$$\forall \theta \in (0, T)\colon \mu_\theta(dx) := \frac{2\max\{x, \theta\}}{T^2 + \theta^2} \mathbb{I}_{(0,T)}(x)\,dx.$$

Denoting by ψ_θ the uniform distribution on $(0, \theta)$ introduced in Example 5.2, we have

$$\mu'_\theta = \frac{2\theta}{T^2 + \theta^2}(\psi_\theta - \mu_\theta), \qquad \mu''_\theta = \frac{2}{T^2 + \theta^2}(\delta_\theta - \mu_\theta) - \frac{8\theta^2}{(T^2 + \theta^2)^2}(\psi_\theta - \mu_\theta).$$

Furthermore, higher-order derivatives do not exist because the Dirac distribution δ_θ is not weakly differentiable.

7.1. Weak analyticity of the product measure. We conclude this section with a result for weak analyticity of product measures. More specifically, in classical analysis, it is well-known that the product of two analytical functions is again analytical. Theorem 7.4 establishes the counterpart of this fact for measure-valued mappings.

THEOREM 7.4. *Let $(\mathscr{D}(\mathbb{S}), v)$ and $(\mathscr{D}(\mathbb{T}), u)$ be Banach bases on $\mathbb{S}$ and $\mathbb{T}$, respectively. Let μ_θ be $[\mathscr{D}(\mathbb{S})]_v$-analytic with domain of convergence $D^v_\theta(\mu)$ and let ν_θ be $[\mathscr{D}(\mathbb{T})]_u$-analytic with domain of convergence $D^u_\theta(\nu)$. Then, $\mu_\theta \times \nu_\theta$ is $[\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v\otimes u}$-analytic and for each $g \in [\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]$ the domain of convergence $D_\theta(g, \mu \times \nu)$ satisfies*:

$$D^v_\theta(\mu) \cap D^u_\theta(\nu) \subset D_\theta(g, \mu \times \nu).$$

PROOF. For $\theta \in \Theta$, let us denote by $(\mu \times \nu)_\theta$ the product measure $\mu_\theta \times \nu_\theta$. Recall that $D^v_\theta(\mu) = \Theta \cap (\theta - R^v_\theta(\mu), \theta + R^v_\theta(\mu))$ with $R^v_\theta(\mu)$ as defined in (32). Similarly, $D^u_\theta(\nu) = \Theta \cap (\theta - R^u_\theta(\nu), \theta + R^u_\theta(\nu))$. Let $\rho = \min\{R^v_\theta(\mu), R^u_\theta(\nu)\}$ and choose $g \in [\mathscr{D}(\mathbb{S}) \otimes \mathscr{D}(\mathbb{T})]_{v\otimes u}$ arbitrarily. We show that, for $|\Delta| < \rho$, it holds that

$$\int g(s,t)(\mu \times \nu)_{\theta+\Delta}(ds,dt) = \lim_{n\to\infty} \sum_{k=0}^{n} \frac{\Delta^k}{k!} \int g(s,t)(\mu \times \nu)^{(k)}_\theta(ds,dt). \tag{40}$$

Let us consider now the linear mappings $T_n\colon [\mathscr{D}(\mathbb{S})]_v \to \mathbb{R}$, defined as

$$\forall\, n \geq 1\colon T_n(f) \stackrel{\text{def}}{=} \sum_{j=0}^{n} \frac{\Delta^j}{j!} \int f(s)\mu^{(j)}_\theta(ds)$$

and, for $t \in \mathbb{T}$ and $n \geq 1$, let

$$H_n(t) = T_n(g(\cdot,t)); \quad H(t) = \int g(s,t)\mu_{\theta+\Delta}(ds).$$

By hypothesis, $H(t) = \lim_{n\to\infty} H_n(t)$. In the following, we show that the dominated convergence theorem applies to the sequence $\{H_n\}$ when integrated w.r.t. ν.

First, note that, according to (16), it holds that $\|g(\cdot,t)\|_v \leq \|g\|_{v\otimes u} u(t)$. Hence, an application of the Cauchy-Schwartz inequality yields

$$|H_n(t)| = |T_n(g(\cdot,t))| \leq \|T_n\|_v \|g(\cdot,t)\|_v \leq \Big(\sup_n \|T_n\|_v\Big) \|g\|_{v\otimes u} u(t). \tag{41}$$

To show that $\sup_n \|T_n\|_v < \infty$, we note that weak analyticity of μ_θ implies that $\{T_n(f)\colon n \in \mathbb{N}\}$ is bounded for each $f \in [\mathscr{D}(\mathbb{S})]_v$, and we apply the Banach-Steinhaus theorem (see Remark 4.1).

Thus, $H_n \in [\mathscr{D}(\mathbb{T})]_u$ and because $u \in \mathscr{L}^1(\{\nu_\theta\colon \theta \in \Theta\})$, the dominated convergence theorem applies to the sequence $\{H_n\}_n$. Hence, interchanging limit with integration on the right-hand side of (40) is justified and yields

$$\int H(t)\nu_{\theta+\Delta}(dt) = \int \lim_{n\to\infty} H_n(t)\,\nu_{\theta+\Delta}(dt) = \lim_{n\to\infty} \int H_n(t)\nu_{\theta+\Delta}(dt). \tag{42}$$

Moreover, due to $[\mathscr{D}(\mathbb{T})]_u$-analyticity of ν_θ, the right-hand side in (42) equals to

$$\lim_{n\to\infty}\lim_{m\to\infty} \sum_{l=0}^{m} \frac{\Delta^l}{l!} \int H_n(t)\nu^{(l)}_\theta(dt).$$

Finally, inserting the expression of $H_n(t)$ in to the above expression, we conclude that the left-hand side of (40) equals to

$$\lim_{n\to\infty}\lim_{m\to\infty} \sum_{l=0}^{m}\sum_{j=0}^{n} \frac{\Delta^{j+l}}{j!\,l!} \int\int g(s,t)\mu^{(j)}_\theta(ds)\nu^{(l)}_\theta(dt). \tag{43}$$

According to Theorem 6.2, the right-hand side of (40) can be rewritten as

$$\lim_{k\to\infty} \sum_{0\leq j+l\leq k} \frac{\Delta^{j+l}}{j!\,l!} \int\int g(s,t)\mu^{(j)}_\theta(ds)\nu^{(l)}_\theta(dt). \tag{44}$$

The power series in (43) is convergent for $|\Delta| < \rho$. Hence, it is absolutely convergent so its limit is not affected by reshuffling terms (see Rudin [30]). It follows that the limits in (43) and (44) coincide and (40) holds true for $|\Delta| < \rho$. The fact that $D^v_\theta(\mu) \cap D^u_\theta(\nu) = \Theta \cap (\theta - \rho, \theta + \rho)$ concludes the proof. $\square$

8. Applications. In what follows, we present two applications—a first one where we provide a method to estimate the derivative of the probability of ruin in some simple insurance model using weak differentiation, and a second one where the expected completion time of a stochastic activity network is approximated analytically using weak analyticity.

8.1. A ruin problem. Let us consider the following example. An insurance company receives premiums from clients at some constant rate $r > 0$ while claims $\{Y_i\colon i \geq 1\}$ arrive according to a Poisson process with rate $\lambda > 0$. Let $\{X_i\colon i \geq 1\}$ denote the interarrival times of the Poisson process and let N_τ denote the number of claims recorded up to some fixed time horizon $\tau > 0$. Assume further that the values of claims are i.i.d. r.v. following a Pareto distribution π_θ, i.e.,

$$\pi_\theta(dx) = \frac{\beta\theta^\beta}{x^{\beta+1}} \mathbb{I}_{(\theta,\infty)}(x)\,dx$$

(see Example 5.5 (ii)) and that the values of the claims are independent of the Poisson process.

Let $V(0) \geq 0$ denote the initial credit of the insurance company. The credit (resp. debt) of the company right after the nth claim, denoted by $V(n)$, follows the recurrence relation:

$$\forall\, n \geq 0\colon\ V(n+1) = V(n) + rX_{n+1} - Y_{n+1}.$$

Ruin occurs before time τ if at least one $n \leq N_\tau$ exists such that $V(n) < 0$ (see Figure 1).

We are interested in estimating the derivative w.r.t. θ of the probability of ruin up to time τ. To this end, we denote by $\Re_\tau$ the event that *ruin occurs up to time* τ. Given the event $\{N_\tau = n\}$, $\Re_\tau$ can be written as follows:

$$\Re_\tau \cap \{N_\tau = n\} = \complement\left(\bigcap_{k=1}^{n}\{V(k) > 0\}\right) = \complement\left\{ r\cdot\sum_{i=1}^{j} X_i > \sum_{i=1}^{j} Y_i,\ \forall\, 1 \leq j \leq n \right\},$$

where $\complement A$ denotes the complement of A. Therefore, considering the sequence $\{g_n\colon n \geq 1\}$ with $g_n \in \mathcal{F}(\mathbb{R}^{2n})$ given by

$$g_n(x_1,\ldots,x_n,y_1,\ldots,y_n) = 1 - \prod_{j=1}^{n} \mathbb{I}_{\{r\cdot\sum_{i=1}^{j} x_i > \sum_{i=1}^{j} y_i\}}(x_1,\ldots,x_n,y_1,\ldots,y_n), \tag{45}$$

we can write

$$\forall\, n \geq 1\colon\ \mathbb{P}_\theta(\Re_\tau \cap \{N_\tau = n\}) = \mathbb{E}_\theta\left[\mathbb{I}_{\{N_\tau = n\}} g_n(X_1,\ldots,X_n,Y_1,\ldots,Y_n)\right], \tag{46}$$

where $\mathbb{E}_\theta$ denotes the expectation operator when the claims Y_i follow distribution π_θ and X_i is exponentially distributed with rate λ. Let μ denote the exponential distribution. As explained in §5.2, the truncated distribution π_θ is weakly $\mathscr{C}_B$-differentiable, satisfying

$$\pi'_\theta = \frac{\beta}{\theta}(\pi_\theta - \delta_\theta).$$

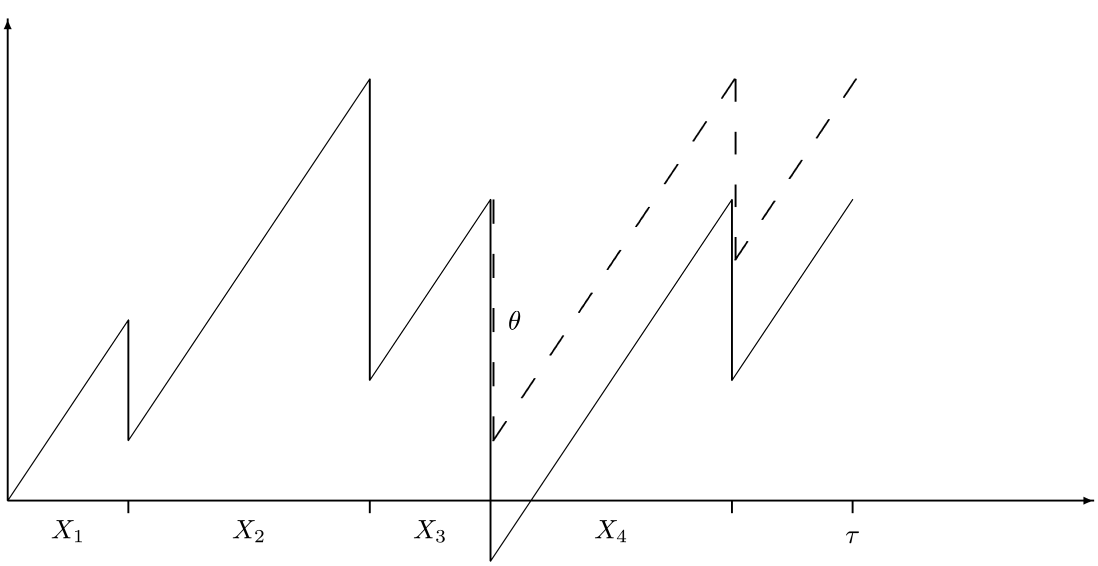

FIGURE 1. An occurrence of the event $\Re_\tau \setminus \Re_\tau^3$ and $N_\tau = 4$.
Note. The dashed line represents a version of the process where the value of the third claim is reduced.

Applying Theorem 6.3 with $v=1$ yields that the product measure $\mu \times \pi_\theta$ is weakly $\mathscr{C}_B(\mathbb{S}^2)$-differentiable with $(\mu \times \pi_\theta)' = \mu \times \pi'_\theta$ (for a proof, use the fact that μ is independent of θ). Applying Theorem 6.3 with $v=1$ again to the n-fold product of $\mu \times \pi_\theta$ yields that $(\mu \times \pi_\theta)^n$ is weakly $\mathscr{C}_B(\mathbb{S}^{2n})$-differentiable. Hence, for any $g \in \mathscr{C}_B(\mathbb{S}^{2n})$, the derivative of the $\int g d(\mu \times \pi_\theta)^n$ can be obtained in closed form (see Example 6.1 for the derivative expression). Note, however, that the sample performance g_n introduced for modeling the ruin probability is not an element of $\mathscr{C}_B(\mathbb{S}^{2n})$. Fortunately, because the discontinuities of g_n have measure zero, our derivative formulas apply to g_n as well. More formally, $(\mu \times \pi_\theta)^n$ is weakly $\mathscr{C}_B(\mathbb{S}^n) \cup \{g_n\}$-differentiable (see Appendix B for details). Hence, we arrive at

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{P}_\theta\left(\Re_\tau \cap \{N_\tau = n\}\right) &= \frac{d}{d\theta}\mathbb{E}_\theta\left[\mathbb{1}_{\{N_\tau=n\}} g_n(X_1,\ldots,X_n,Y_1,\ldots,Y_n)\right] \\
&= \frac{\beta}{\theta}\sum_{i=1}^{n}\mathbb{E}_\theta\Big[\mathbb{1}_{\{N_\tau=n\}} g_n(X_1,\ldots,X_n,Y_1,\ldots,Y_n) \\
&\qquad\qquad - \mathbb{1}_{\{N_\tau=n\}} g_n(X_1,\ldots,X_n,Y_1,\ldots,Y_{j-1},\theta,Y_{j+1},\ldots,Y_n)\Big] \\
&= \frac{\beta}{\theta}\sum_{i=1}^{n}\left(\mathbb{P}_\theta\left(\Re_\tau \cap \{N_\tau=n\}\right) - \mathbb{P}_\theta\left(\Re^i_\tau \cap \{N_\tau=n\}\right)\right),
\end{aligned}$$

where $\Re^i_\tau$ denotes the event that there is ruin up to time τ, when the value of the ith claim is replaced by the constant θ, i.e.,

$$\Re^i_\tau \cap \{N_\tau = n\} = \bigcup_{k=1}^{n}\{V^i(k) < 0\}.$$

Provided that interchanging limit with differentiation is allowed, we obtain

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{P}_\theta\left(\Re_\tau\right) &= \frac{d}{d\theta}\sum_{n=1}^{\infty}\mathbb{P}_\theta\left(\Re_\tau \cap \{N_\tau=n\}\right) && (47)\\
&= \sum_{n=1}^{\infty}\frac{d}{d\theta}\mathbb{P}_\theta\left(\Re_\tau \cap \{N_\tau=n\}\right) \\
&= \sum_{n=1}^{\infty}\frac{\beta}{\theta}\sum_{i=1}^{n}\left(\mathbb{P}_\theta\left(\Re_\tau \cap \{N_\tau=n\}\right) - \mathbb{P}_\theta\left(\Re^i_\tau \cap \{N_\tau=n\}\right)\right) \\
&= \frac{\beta}{\theta}\left(\mathbb{P}_\theta\left(\Re_\tau \cap \{N_\tau \ge n\}\right) - \mathbb{P}_\theta\left(\Re^n_\tau \cap \{N_\tau \ge n\}\right)\right). && (48)
\end{aligned}$$

Note that the nth remainder term of the series in (48) is bounded by

$$\sum_{k=n+1}^{\infty}\mathbb{P}_\theta(\{N_\tau \ge n\}) \le \sum_{k=n+1}^{\infty}\frac{(\lambda\tau)^k}{k!}.$$

Because the bound is independent of θ and converges to 0 as $n \to \infty$, it means that we deal with a uniformly convergent series of functions of θ. Interchanging limit with differentiation in (47) is thus justified.

Taking into account that $Y_n > \theta$ almost surely, a sample path analysis together with a monotonicity argument yields $\Re^n_\tau \subset \Re_\tau$. Moreover, the difference $\Re_\tau \setminus \Re^n_\tau$ represents the event that ruin occurs up to time τ but it does not occur anymore if one reduces the value of the nth claim by $Y_n - \theta$; a graphical representation of these facts can be found in Figure 1. Note that this event is incompatible with $\{N_\tau < n\}$, i.e., if the "reduced claim" comes after time τ. Hence, it holds that $\mathbb{P}_\theta\left((\Re_\tau \setminus \Re^n_\tau) \cap \{N_\tau < n\}\right) = 0$ so that (48) becomes

$$\frac{d}{d\theta}\mathbb{P}_\theta\left(\Re_\tau\right) = \frac{\beta}{\theta}\sum_{n=1}^{\infty}\mathbb{P}_\theta\left(\Re_\tau \setminus \Re^n_\tau\right). \tag{49}$$

Remark 8.1. Following the line of argument that leads from (47) to (48), we obtain

$$\begin{aligned}
\frac{d}{d\theta}\mathbb{P}_\theta(\Re_\tau) = \sum_{n=1}^{\infty}\frac{\beta}{\theta}\sum_{i=1}^{n}\mathbb{E}_\theta\Big[&\mathbb{1}_{\{N_\tau=n\}} g_n(X_1,\ldots,X_n,Y_1,\ldots,Y_n) \\
&- \mathbb{1}_{\{N_\tau=n\}} g_n(X_1,\ldots,X_n,Y_1,\ldots,Y_{j-1},\theta,Y_{j+1},\ldots,Y_n)\Big]
\end{aligned}$$

$$= \frac{\beta}{\theta}\mathbb{E}_\theta\bigg[N_\tau g_{N_\tau}(X_1,\ldots,X_{N_\tau},Y_1,\ldots,Y_{N_\tau}) \\ -\sum_{j=1}^{N_\tau} g_{N_\tau}(X_1,\ldots,X_{N_\tau},Y_1,\ldots,Y_{j-1},\theta,Y_{j+1},\ldots,Y_{N_\tau})\bigg]$$

(c.f. Example 6.1). The expression on the right-hand side provides an unbiased estimator for the derivative of the ruin probability. For details on the relationship between weak derivatives and unbiased estimators, we refer to Heidergott and Vázquez-Abad [13].

8.2. Stochastic activity networks. Stochastic activity networks (SAN) such as those arising in the PERT form an important class of models for systems and control engineering. Roughly, a SAN is a collection of activities, each with some (deterministic or random) duration, along with a set of precedence constraints that specify that activities begin only when certain others have finished. Such a network can be modeled as a directed acyclic weighted graph with one source, one sink node, and additive weight-function τ. A simple example is provided in Figure 2. The network has five nodes, labeled from one (source) to five (sink) and the edges denote the activities under consideration. The weights X_i, $1\le i\le 7$, denote the durations of the corresponding activities. For instance, activity six can only begin when both activities two and three have finished.

Let $\mathcal{P}$ denote the set of all paths from the source to the sink node. Should (some) durations be random variables, we assume them mutually independent. However, note that, in general, the path weights are not independent. The completion time, denoted by T, is defined as the weight of the "maximal" path:

$$T = \max\{\tau(\pi)\colon\ \pi\in\mathcal{P}\}.$$

For more details on SAN, we refer to Pich et al. [29]. For instance, in the above example, the set of paths from source node one to sink node five is

$$\mathcal{P} = \{(1,2,5);\,(1,2,4,5);\,(1,2,3,4,5);\,(1,3,4,5)\}.$$

Thus, the completion time in this case can be expressed as

$$T = \max\{X_1+X_5;\ X_1+X_4+X_7;\ X_1+X_3+X_6+X_7;\ X_2+X_6+X_7\}.$$

One of the most challenging problems in this area is to compute *the expected completion time*, i.e., $\mathbb{E}[T]$. Distribution free bounds for $\mathbb{E}[T]$ are provided in Devroye [6]. In the following paragraphs, we aim to establish a functional dependence between a particular parameter, e.g., the expected duration of some particular tasks, and the expected completion time of the system. Here, we propose a Taylor series approximation for a SAN with exponentially distributed service times, where the computation of higher-order derivatives relies on weak differentiation theory as presented in this paper.

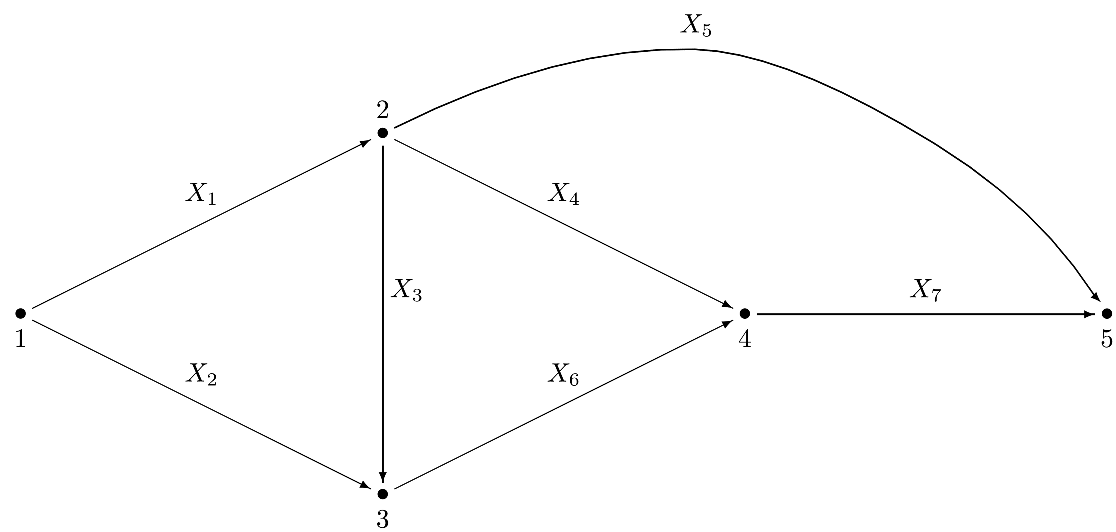

Figure 2. A stochastic activity network with source node 1 and sink node 5.

We start by considering $\mathbb{S} = [0, \infty)$ with the usual metric. v: $\mathbb{S} \to \mathbb{R}$ is defined as $v(x) = 1 + x$. Next, we define g^T: $\mathbb{S}^7 \to \mathbb{R}$ as

$$g^T(x_1, \ldots, x_7) \stackrel{\text{def}}{=} \max\{x_1 + x_5;\ x_1 + x_4 + x_7;\ x_1 + x_3 + x_6 + x_7;\ x_2 + x_6 + x_7\},$$

i.e., $T = g^T(X_1, \ldots, X_7)$ and

$$\mathbb{E}[T] = \int \cdots \int g^T(x_1, \ldots, x_7)\mu_1(dx_1)\cdots\mu_7(dx_7),$$

where we denote by μ_i the distribution of X_i for $1 \le i \le 7$. In accordance with Theorem 6.3, it holds that if μ_i is weakly differentiable with respect to some parameter θ for all $1 \le i \le 7$, then the distribution of T is weakly differentiable w.r.t. θ as well. Roughly speaking, this means that "the distribution of T is differentiable w.r.t. each μ_i."

Assume, for instance, that r.v. X_i, $1 \le i \le 7$, are independent and exponentially distributed with rates λ_i. We let $\lambda_1 = \lambda_3 = \theta$ vary and fix the other rates, i.e., λ_i is independent of θ for $i \in \{2, 4, 5, 6, 7\}$. By Example 7.1, the exponential distribution is weakly $[\mathcal{F}]_v$-analytical for $v(x) = 1 + x$ and the domain of convergence is given by $|\Delta| < \theta$. Because the distributions that are independent of θ are trivially weakly analytical, from Theorem 7.4 we conclude that the joint distribution of the vector $(X_1, \ldots, X_7)$ is weakly $[\mathcal{F}(S^7)]_{v\otimes\cdots\otimes v}$-analytical. Moreover, the radius of convergence of the Taylor series is equal to θ. Finally, we note that

$$|g^T(x_1, \ldots, x_7)| \le \prod_{i=1}^{7}(1 + x_i) = (v \otimes \cdots \otimes v)(x_1, \ldots, x_7),$$

i.e., g^T belongs to $[\mathcal{F}(S^7)]_{v\otimes\cdots\otimes v}$, the sevenfold product of the Banach base $(\mathcal{F}, v)$.

Next, we evoke Corollary 6.1 for computing the derivatives. Because only the derivatives of $\mu_{1,\theta}$ and $\mu_{3,\theta}$ are significant (inspired by Example 5.1 for $j, k \ge 0$), we consider a "modified" network where X_1 is replaced by the sum of j independent samples from an exponentially distributed r.v. with rate θ, and X_3 is replaced by the sum of k independent samples from the same distribution. All other durations remain unchanged, i.e., we replace the exponential distribution of X_1 and X_3 by the $\gamma(j, \theta)$ and $\gamma(k, \theta)$ distribution, respectively. Let $T_{j,k}$ denote the completion time of the modified SAN, i.e., $T_{1,1} = T$ and we agree that $T_{j,k} = 0$ if $j = 0$ or $k = 0$. With this notation, Theorem 6.3 yields

$$\forall\, n \ge 0\colon\ \frac{d^n}{d\theta^n}\mathbb{E}_\theta[T] = (-1)^n \frac{n!}{\theta^n} \sum_{i+j=n} \mathbb{E}_\theta[T_{i+1,j+1} - T_{i+1,j} - T_{i,j+1} + T_{i,j}], \tag{50}$$

and for each $n \ge 1$, we call

$$\mathbf{T}_n(\theta, \Delta) \stackrel{\text{def}}{=} \sum_{k=0}^{n}(-1)^k \left(\frac{\Delta}{\theta}\right)^k \sum_{i+j=k} \mathbb{E}_\theta[T_{i+1,j+1} - T_{i+1,j} - T_{i,j+1} + T_{i,j}] \tag{51}$$

the nth order Taylor polynomial for $\mathbb{E}_{\theta+\Delta}[T]$ at θ, where $\mathbb{E}_\theta$ denotes the expectation operator w.r.t. the product measure $\mu_{1,\theta} \times \mu_2 \times \mu_{3,\theta} \times \mu_4 \times \mu_5 \times \mu_6 \times \mu_7$. Bounds on the error term can be obtained by Theorem 7.3. However, as we will explain, in the case of the SAN model, bounds can be obtained in a more direct way. To see this, note that, by using a monotonicity argument, one can easily check that

$$\forall\, i, j \ge 0\colon\ \left|\mathbb{E}_\theta[T_{i+1,j+1} - T_{i+1,j} - T_{i,j+1} + T_{i,j}]\right| \le \mathbb{E}_\theta[X_1] + \mathbb{E}_\theta[X_3] = \frac{2}{\theta}. \tag{52}$$

Hence, a bound for the error of the nth order Taylor polynomial is given by

$$\begin{aligned}\forall\, |\Delta| < \theta\colon\ |\mathbb{E}_{\theta+\Delta}[T] - \mathbf{T}_k(\theta, \Delta)| &\le \frac{2}{\theta} \sum_{k=n+1}^{\infty} (k+1) \left(\frac{|\Delta|}{\theta}\right)^k \\ &= \frac{2}{\theta} \frac{(n+2) - (n+1)(|\Delta|/\theta)}{(1 - |\Delta|/\theta)^2} \left(\frac{|\Delta|}{\theta}\right)^{n+1} \\ &\le \frac{2(n+1)}{(\theta - |\Delta|)} \left(\frac{|\Delta|}{\theta}\right)^{n+1}.\end{aligned} \tag{53}$$

Example 8.1. To perform a numerical experiment, we considered the following rates:

$$\lambda_1 = \lambda_3 = \theta, \quad \lambda_6 = 1, \quad \lambda_2 = \lambda_4 = \frac{1}{2}, \quad \lambda_5 = \frac{1}{5}, \quad \lambda_7 = \frac{1}{3}.$$

Computation of the coefficients of the Taylor polynomial appears to be quite demanding and it is worth noting that the coefficients can, alternatively, be evaluated by simulation. Our calculations show that the Taylor polynomial $\mathbf{T}_3(1, \Delta)$ of order three approximates the true function $\mathbb{E}[T_{1+\Delta}]$ for $|\Delta| \le 0.4$ with a relative error, estimated according to (53), of less than 3.4%.

Appendix A. Uniqueness of the weak $\mathscr{D}_v$-limit. The $\mathscr{D}_v$-limit, as defined in (4), is, in general, not unique. Indeed, let us consider $\mathbb{S} = [0, \infty)$ endowed with the usual metric and $v(s) = s$ for all $s \in \mathbb{S}$. Denote by δ_0 the Dirac measure, i.e., δ_0 assigns mass one to point zero. Assume that μ is a $\mathscr{D}_v$-limit of the sequence $\{\mu_n\}_n \subset \mathscr{M}$. Because we have $g(0) = 0$ for $g \in \mathscr{D}_v$, it follows that $\mu + \alpha \cdot \delta_0 \in \mathscr{M}$ is also a $\mathscr{D}_v$-limit of the sequence $\{\mu_n\}_n$ for each $\alpha \in \mathbb{R}$ and the $\mathscr{D}_v$-limit fails to be unique. In words, (4) still holds true if one assigns a different mass to the "zero set" of v. Our next result will elucidate this issue. In particular, it shows that the set $\mathscr{D}_v$, likewise $\mathscr{C}_B$, is appropriate for introducing weak convergence.

Lemma A.1. *Let $v \in \mathscr{C}^+(\mathbb{S})$ and let $\mathbb{S}_v = \{s \in \mathbb{S}\colon v(s) > 0\}$. If $\mu, \nu \in \mathscr{M}(\mathbb{S})$ be such that $v \in \mathscr{L}^1(\{\mu, \nu\})$ and*

$$\forall g \in \mathscr{D}_v\colon \int g(s)\mu(ds) = \int g(s)\nu(ds), \tag{A1}$$

then the traces of μ and ν coincide on $\mathbb{S}_v$. That is,

$$\forall A \in \mathscr{S}\colon \mu(A \cap \mathbb{S}_v) = \nu(A \cap \mathbb{S}_v). \tag{A2}$$

Proof. Because $\mathscr{S}$ is the Borel field of $\mathbb{S}$, we may assume, without loss of generality, that $A \in \mathscr{S}$ is an arbitrary nonempty open set. For $\epsilon > 0$, consider the set:

$$A_\epsilon \stackrel{\text{def}}{=} \{s \in A\colon \rho(s, \complement A) \ge \epsilon^{-1}\} \subset A,$$

where, for $E \subset \mathbb{S}$, we denote $\complement E = \mathbb{S} \setminus E$ and $\rho(s, E) = \inf\{\rho(s, t)\colon t \in E\}$. Note that, for sufficiently large $\epsilon > 0$, A_ϵ is a nonempty closed set satisfying $A_\epsilon \cap \complement A = \varnothing$. Because A is an open set, $\complement A$ is closed and, according to Urysohn's lemma, there exists a continuous function $f_\epsilon\colon \mathbb{S} \to [0, 1]$ such that $f_\epsilon(x) = 1$ for $x \in A_\epsilon$ and $f_\epsilon(x) = 0$ for $x \in \complement A$. On the other hand, the family $\{A_\epsilon\}_{\epsilon>0} \subset \mathscr{F}$ is ascendent and $\cup_{\epsilon>0} A_\epsilon = A$. Hence, f_ϵ converges pointwise to $\mathbb{I}_A$ as $\epsilon \to \infty$.

Consider now, for each $\epsilon > 0$, the mapping $h_\epsilon \in \mathscr{C}^+(\mathbb{S})$ defined as

$$h_\epsilon(s) = \min\{f_\epsilon(s), \epsilon \cdot v(s)\}.$$

Obviously, $h_\epsilon \in \mathscr{D}_v$, and $h_\epsilon(s) = 0$ for $s \notin \mathbb{S}_v$ and all $\epsilon > 0$, and it holds that

$$\forall s \in \mathbb{S}\colon \lim_{\epsilon \uparrow \infty} h_\epsilon(s) = \mathbb{I}_{A \cap \mathbb{S}_v}(s).$$

Applying now the dominated convergence theorem yields:

$$\forall A \in \mathscr{S}\colon \mu(A \cap \mathbb{S}_v) = \lim_{\epsilon \uparrow \infty} \int h_\epsilon(s)\mu(ds) = \lim_{\epsilon \uparrow \infty} \int h_\epsilon(s)\nu(ds) = \nu(A \cap \mathbb{S}_v),$$

which concludes the proof of (A2). □

Remark A.1. If we denote by $\sim$ the equivalence relation on $\mathscr{M}$ given by $\mu \sim \nu$ if (A2) holds true, then Lemma A.1 shows that (A1) implies $\mu \sim \nu$. Going back to (4), we conclude that the $\mathscr{D}_v$-limit is uniquely determined up to this equivalence relation and the precise definition of the $\mathscr{D}_v$-limit would be in terms of the equivalence class of μ denoted by $[\mu]$. Note that because $g \in \mathscr{D}_v$ implies $g(s) = 0$ for $s \notin \mathbb{S}_v$, the behavior of μ outside $\mathbb{S}_v$ is not relevant for our analysis and we may, with slight abuse of notation, identify μ and $[\mu]$.

In fact, the algebraic dual space $\mathscr{D}_v^*$ of $\mathscr{D}_v$, i.e., the set of all linear functionals on $\mathscr{D}_v$, is $\mathscr{M}/\sim$, i.e., the quotient space of $\mathscr{M}$ w.r.t. equivalence relation $\sim$.

Appendix B. Regular convergence and setwise differentiation. According to Definition 5.1, weak $\mathscr{C}_B$-convergence can only handle continuous performance measures. In fact, the class of mappings g that satisfy Equation (19) is much larger and includes, for instance, indicator functions of so-called *continuity sets*. This fact is well-known for classical weak convergence of probability measures (see, e.g., the *Portmanteau theorem* in Billingsley [2]). We now show that a similar result holds true for weak differentiation of measures. To use the results known from classical weak convergence theory, we appeal to the concept of *regular convergence* of measures, which was introduced and studied in Leahu [24].

Definition B.1. We say that the sequence $\{\mu_n : n \in \mathbb{N}\}$ converges regularly in $\mathscr{M}_v$ if both sequences $[\mu_n]^{\pm}$ are weakly $[\mathscr{C}]_v$-convergent, where, for each $n \in \mathbb{N}$, we assume that

$$\mu_n = [\mu_n]^+ - [\mu_n]^-$$

is the Hahn-Jordan decomposition of μ_n. In addition, if $\mu_*: \Theta \to \mathscr{M}_v$, we say that μ_θ is regularly $[\mathscr{C}]_v$-continuous if $\mu_{\theta+\xi}$ converges regularly to μ_θ for $\xi \to 0$, and we say that μ_θ is regularly $[\mathscr{C}]_v$-differentiable if $(\mu_{\theta+\xi} - \mu_\theta)/\xi$ converges regularly for $\xi \to 0$.

Note that, if μ_θ is regularly $[\mathscr{C}]_v$-differentiable, then it is weakly $[\mathscr{C}]_v$-differentiable cf. Definition 5.1 and its weak derivative coincides with its "regular" derivative.

Example B.1. The Pareto distribution

$$\pi_\theta(dx) = \frac{\beta\theta^\beta}{x^{\beta+1}} \mathbb{I}_{(\theta,\infty)}(x)\,dx, \quad \theta > 0$$

is regularly $[\mathscr{C}]_B$-differentiable. To see this, one can use Theorem 2.2 in Leahu [24], which asserts that weak differentiability of μ_θ together with regular continuity of μ'_θ implies regular differentiability of μ_θ. In our case, the weak derivative of π_θ satisfies

$$\pi'_\theta = \frac{\beta}{\theta}(\pi_\theta - \delta_\theta).$$

Note that the Hahn-Jordan decomposition of π'_θ is given by $[\pi'_\theta]^+ = \beta\pi_\theta/\theta$ and $[\pi'_\theta]^- = \beta\delta_\theta/\theta$. Hence, π'_θ is regularly $[\mathscr{C}]_B$-continuous and it follows that π_θ is regularly $[\mathscr{C}]_B$-differentiable.

Theorem B.1, which was also proved in Leahu [24] (see Theorem 2.6), provides sufficient conditions for "setwise" differentiation. In other words, the result can be formulated as "regular differentiability implies setwise differentiability with respect to the class of continuity sets[9] of the weak derivative." The precise statement is as follows:

Theorem B.1. *If $\mu_{i,*}: \Theta \to \mathscr{M}^1_v$ are such that $\mu_{i,\theta}$ is regularly $[\mathscr{C}]_{v_i}$-differentiable with $[\mathscr{C}]_{v_i}$-derivative $(c_{i,\theta}, \mu^+_{i,\theta}, \mu^-_{i,\theta})$, for $1 \le i \le n$. Let Π_θ denote the product measure $\mu_{1,\theta} \times \cdots \times \mu_{n,\theta}$, and let $\vec{v}$ denote the tensor product $v_1 \otimes \cdots \otimes v_n$ (see (12) for a definition). For $g \in [\mathscr{F}]_{\vec{v}}$, denote by D_g the set of discontinuities of g. If*

$$\forall\, 1 \le i \le n: (\mu_{1,\theta} \times \cdots \times \mu^{\pm}_{i,\theta} \times \cdots \times \mu_{n,\theta})(D_g) = 0, \tag{B1}$$

then it holds that

$$\frac{d}{d\theta} \int g(x)\Pi_\theta(dx) = \int g(x)\Pi'_\theta(dx). \tag{B2}$$

Example B.2. Recall the situation in §8.1. Let μ denote the exponential distribution, let X be distributed according to μ, and let Y_i be distributed according to $\pi_{i,\theta}$ for $1 \le i \le j$. Let $Y^+_i = Y_i$ and $Y^-_i = \theta$, and take g as defined in (45). Then, condition (B1) in Theorem B.1 is equivalent to

$$\mathbb{P}_\theta\left(\left\{ r \cdot \sum_{i=1}^{j} X_i = \sum_{i=1}^{j} Y^{\pm}_i \right\}\right) = 0.$$

Because μ is a continuous distribution, the above equation holds. Therefore, $\Re_\tau$ is a continuity set and, by Theorem B.1, differentiability of $\mathbb{P}_\theta(\Re_\tau)$ in §8.1 follows.

Acknowledgments. The authors are grateful to Arie Hordijk for fruitful discussions. The research of the second author was supported by the Technology Foundation Stichting Technische Wetenschappen (STW), applied science division of Nederlandse Organisatie voor Wetenschappelijk Onderzoek (NOW) and the technology programme of the Ministry of Economic Affairs. The research was carried out while the second author was with Vrije Universiteit Amsterdam, Department of Econometrics and Operations Research, Amsterdam, The Netherlands.

[9] A set A is said to be a continuity set for the measure μ if $|\mu|(\partial(A)) = 0$, where ∂A denotes the boundary of A.

References

[1] Baccelli, F., S. Hasenfuß, V. Schmidt. 1997. Transient and stationary waiting times in $(\max, +)$-linear systems with Poisson input. *Queueing Systems: Theory and Applications (QUESTA)* **26** 301–342.
[2] Billingsley, P. 1966. *Weak Convergence of Probability Measures*. John Wiley and Sons, New York.
[3] Cohn, D. 1980. *Measure Theory*. Birkhäuser, Stuttgart, Germany.
[4] Dekker, R., A. Hordijk. 1983. Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. Report 83-36, Institute of Applied Mathematics and Computing Science, Leiden University, Leiden, The Netherlands.
[5] Dekker, R., A. Hordijk. 1988. Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. *Math. Oper. Res.* **13**(3) 395–421.
[6] Devroye, L. P. 1979. Inequalities for completion times of stochastic PERT networks. *Math. Oper. Res.* **4**(4) 441–447.
[7] Dunford, N. 1971. Linear operators. *Pure and Applied Mathematics*. Interscience Publishers, New York.
[8] Hasenfuß, S. 1997. Performance analysis of $(\max, +)$-linear systems with Taylor series expansions. Unpublished doctoral dissertation, Ulm University, Ulm, Germany.
[9] Heidergott, B. 2007. *Max Plus Linear Stochastic Systems and Perturbation Analysis*. Springer, Berlin.
[10] Heidergott, B., A. Hordijk. 2003. Taylor series expansions for stationary Markov chains. *Adv. Appl. Probab.* **35**(4) 1046–1070.
[11] Heidergott, B., A. Hordijk. 2004. Single-run gradient estimation via measure-valued differentiation. *IEEE Trans. Automatic Control* **49** 1843–1846.
[12] Heidergott, B., F. Vázquez-Abad. 2006. Measure-valued differentiation for random horizon problems. *Markov Processes Related Fields* **12** 509–536.
[13] Heidergott, B., F. Vázquez-Abad. 2008. Measure valued differentiation for Markov chains. *J. Optim. Theory Appl.* **136**(2) 187–209.
[14] Heidergott, B., F. Vázquez-Abad. 2009. Gradient estimation for a class of systems with bulk services: A problem in public transportation. *TOMACS*. Forthcoming.
[15] Heidergott, B., T. Fahrenhorst-Yuan, F. Vázquez-Abad. 2009. Perturbation analysis approach to phantom estimators for waiting times in the G/G/1 queue. *Discrete Event Dynam. System.* (February 20) http://www.springer.com/math/applications/journal/10626.
[16] Heidergott, B., A. Hordijk, H. Leahu. 2009. Strong bounds on perturbations. *Math. Methods Oper. Res.* **70**(1) 99–127.
[17] Heidergott, B., A. Hordijk, H. Weisshaupt. 2006. Measure-valued differentiation for stationary Markov chains. *Math. Oper. Res.* **31**(1) 154–172.
[18] Heidergott, B., F. Vázquez-Abad, W. Volk-Makarewicz. 2008. Sensitivity estimation for Gaussian systems. *Eur. J. Oper. Res.* **187**(1) 193–207.
[19] Heidergott, B., F. Vázquez-Abad, G. Pflug, T. Fahrenhorst-Yuan. 2009. Gradient estimation for discrete event systems by measure-valued differentiation. *TOMACS. ACM Trans. Model. Comput. Simulation*. Forthcoming.
[20] Hordijk, A., A. A. Yushkevich. 1999. Blackwell optimality in the class of all policies in Markov decision chains with a Borel state space and unbounded rewards. *Math. Methods Oper. Res.* **50**(3) 421–448.
[21] Kartashov, N. 1996. *Strong Stable Markov Chains*. VSP, Utrecht, The Netherlands.
[22] Kushner, H., F. Vázquez-Abad. 1992. Estimation of the derivative of a stationary measure with respect to a control parameter. *J. Appl. Probab.* **29** 343–352.
[23] Kushner, H., F. Vázquez-Abad. 1996. Stochastic approximations for systems of interest over an infinite time interval. *SIAM J. Control Optim.* **29** 712–756.
[24] Leahu, H. 2008. Measure-valued differentiation for finite products of measures: Theory and applications. Thela Thesis. No. 428, Tinbergen Institute Research Series, Tinbergen Institute, Rotterdam, The Netherlands.
[25] Lipman, S. 1974. On dynamic programming with unbounded rewards. *Management Sci.* **21**(11) 1225–1233.
[26] Meyn, S. P., R. L. Tweedie. 1993. *Markov Chains and Stochastic Stability*. Springer, London.
[27] Pflug, G. 1996. *Optimization of Stochastic Models*. Kluwer Academic, Boston.
[28] Pflug, G. 1988. Derivatives of probability measures—Concepts and applications to the optimization of stochastic systems. *Lecture Notes in Control and Information Science*, Vol. 103. Springer, Berlin, 252–274.
[29] Pich, M., C. Loch, A. de Meyer. 1996. On uncertainty, ambiguity and complexity in project management. *Management Sci.* **75**(2) 137–176.
[30] Rudin, W. 1976. *Principles of Mathematical Analysis*, 3rd ed. McGraw-Hill, Tokyo.
[31] Semadeni, Z. 1971. *Banach Spaces of Continuous Functions*. Polish Scientific Publishers, Warszaw, Poland.