

Differentiability of Product Measures*

Bernd Heidergott

Haralambie Leahu †

Abstract

In this paper, we study cost functions over a finite collection of random variables. For this type of models, a calculus of differentiation is developed that allows to obtain a closed-form expression for derivatives, where “differentiation” has to be understood in the weak sense. The techniques for establishing the results is new and establish an interesting link between functional analysis and gradient estimation. By establishing a product rule of weak analyticity, Taylor series approximations of finite products can be established. 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 technique are provided.

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 $1 \leq i \leq n$) depending on some parameter θ (with $\theta \in \Theta = (a, b) \subset \mathbb{R}$ for $a < b$), then one is concerned with the following type of models

$$J_g(\theta) \stackrel{\text{def}}{=} \mathbb{E}_\theta[g(X_n, \dots, X_1)] = \int_{\mathbb{S}_1} \cdots \int_{\mathbb{S}_n} g(s_n, \dots, s_1) \mu_{n,\theta}(ds_n) \cdots \mu_{1,\theta}(ds_1), \quad (1)$$

for g a cost function 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, transient waiting times in queueing networks or insurance models over a finite number of claims.

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 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 to obtain a closed-form expression for $dJ_g(\theta)/d\theta$. Here, “differentiation” has to be understood in the weak sense,

***Acknowledgement:** This research is supported by the Technology Foundation STW, applied science division of NWO and the technology programme of the Dutch Ministry of Economic Affairs.

†Vrije Universiteit, Department of Econometrics and Operations Research, De Boelelaan 1105, 1081 HV Amsterdam. e-mail: {bheidergott,hleahu}@feweb.vu.nl

see Section 5 for a formal definition. The concept of weak differentiation of measures was introduced by Pflug [15] (see also the monograph [16]) and has been extended to the more general concept of \mathcal{D} -differentiation in [9].

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 measurable spaces $(\mathbb{S}, \mathcal{S})$ and $(\mathbb{T}, \mathcal{T})$, respectively. In order 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 for establishing 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 $\mu_\theta \times \nu_\theta$.

Like in conventional analysis, the main work in establishing product rules for weak differentiability, is 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, best to our knowledge, hasn't been established in the literature so far.

The contributions of the paper are:

1. We establish the theory of limits of sequences of signed measures. In particular, we will show that weak convergence of a sequence of signed measures does not implies weak convergence of the negative and positive parts, respectively.
2. A full account on weak differentiability of finite products of probability measures is provided. The techniques for establishing the results is new and establish an interesting link between functional analysis and gradient estimation. Moreover, the theory of weak differentiation is modular in the sense that sufficient conditions for weak differentiability of product measures are expressed in terms of (easier to analyze) properties of weak differentiability of the individual probability measures.
3. By establishing a product rule of weak analyticity, Taylor series approximations of finite products can be established. 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.

The paper is organized as follows. For ease of reference, we introduce in Section 2 the basic notation that will be used throughout the paper. Section 3 is devoted to signed measures and their limits. In Section 3.3, \mathcal{D}_v -spaces are introduced. Functional analysis on \mathcal{D}_v -spaces will be addressed in Section 4. Weak differentiability of probability measures is discussed in Section 5 and that of products of probability measures is discussed in Section 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 technique (PERT) will be provided in Section 8. Some technical material is provided in the Appendix.

2 General Notation

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

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

$$\mathcal{C}(\mathbb{S}) \stackrel{\text{def}}{=} \{g \in \mathcal{C}(\mathbb{S}) : g(s) \geq 0, \forall s \in \mathbb{S}\}.$$

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

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

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

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

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

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

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

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

3 Weak Convergence of Measures

Weak convergence of probability measures was originally introduced in [1]. In this section we extend the concept and give a topology on \mathcal{M} , by means of convergence of sequences of arbitrary measures. We start by a brief overview of signed measures and in Section 3.2 we define the concept of weak convergence on \mathcal{M} .

3.1 Signed Measures

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

$$\forall E \in \mathcal{S} : \mu(E) = \mu^+(E) - \mu^-(E). \quad (2)$$

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

Two positive measures are called *orthogonal* if they have disjoint support; more formally, $\mu_1, \mu_2 \in \mathcal{M}^+$ are orthogonal if there exists a set $A \in \mathcal{S}$ 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 $\mu \in \mathcal{M}$ one can define *the variation measure* $|\mu| \in \mathcal{M}^+$ and the *total variation* $\|\mu\|_{TV}$, of μ as follows:

$$\forall E \in \mathcal{S} : |\mu|(E) = \sup_{A \in \mathcal{S}, A \subseteq E} |\mu(A)|$$

and

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

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 further details on measure theory we refer to [2].

3.2 Weak Convergence on \mathcal{M}

The following definition introduces the concept of *weak convergence* on \mathcal{M} .

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

$$\forall g \in \mathcal{D} : \lim_{n \rightarrow \infty} \int g(s) \mu_n(ds) = \int g(s) \mu(ds). \quad (4)$$

We write $\mu_n \xrightarrow{\mathcal{D}} \mu$ (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 $\mathcal{D} = \mathcal{C}_B$; see [1]. The following example illustrates the dependence of \mathcal{D} -convergence of a sequence of measures $\{\mu_n\}_n$ on the choice of \mathcal{D} .

Example 1. On $\mathbb{S} = [0, \infty)$ let us consider the family of probability measures

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

for $\theta \in (0, 2)$, where $C_1 = 2$ and

$$\forall \theta \neq 1 : C_\theta = \frac{2 \sin(\pi\theta)}{\pi\theta(1-\theta)}.$$

¹In general the weak limit μ is not unique.

For $\theta \uparrow 1$, one can easily check that $\mu_\theta \xrightarrow{\mathcal{D}} \mu_1$ holds true when $\mathcal{D} = \mathcal{C}_B$ but it does not hold true if the identity mapping on \mathbb{S} belongs to \mathcal{D} . That is because

$$\lim_{\theta \uparrow 1} \int x \mu_\theta(dx) = \infty.$$

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

Example 2. Let us consider the sequence

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

where δ_x denotes the Dirac distribution which assigns mass to x . Then $\mu_n \xrightarrow{\mathcal{C}_B} \delta_0$, for $n \rightarrow \infty$ but $\mu_{2k+1}^+ \xrightarrow{\mathcal{C}_B} \delta_0$ and $\mu_{2k}^+ \xrightarrow{\mathcal{C}_B} \delta_0 + \delta_1$, for $k \rightarrow \infty$.

3.3 \mathcal{D}_v -Spaces

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

$$\mathcal{D}_v(\mathbb{S}) \stackrel{\text{def}}{=} \{g \in \mathcal{D}(\mathbb{S}) \mid \exists c > 0 : |g(s)| \leq c \cdot v(s), \forall s \in \mathbb{S}\}. \quad (5)$$

The minimal c for which the inequality in (5) holds true is the so-called v -norm (to be introduced in next section). Note that $\mathcal{D}_v(\mathbb{S})$ is a linear subspace of $\mathcal{D}(\mathbb{S})$ and $\mathcal{C}_B(\mathbb{S}) \subset \mathcal{D}_v(\mathbb{S})$ provided that²

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

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

Example 3. Let $v(x) = e^x$, for $x \in \mathbb{S} = [0, \infty)$. Since for every polynomial P it holds that $\lim_{x \rightarrow \infty} e^{-x} P(x) = 0$ it turns out that the space $\mathcal{D}_v([0, \infty))$ contains all (finite) polynomials. However, the polynomials are not the only elements of \mathcal{D}_v since, for instance, the mapping $x \mapsto \ln(1+x)$ also belongs to \mathcal{D}_v .

Remark 1. If the sequence $\{\mu_n\}_n \subset \mathcal{M}^1$ converges weakly to μ in the classical way, i.e., $\mu_n \xrightarrow{\mathcal{C}_B} \mu$, then

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

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

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

see, e.g., [1]. One can easily show that uniform integrability of v implies uniform integrability of all continuous $g \in \mathcal{D}_v$. Hence, we conclude that, if $\mu_n \xrightarrow{\mathcal{C}_B} \mu$ and v is uniformly integrable w.r.t. $\{\mu_n\}_n$ then $\mu_n \xrightarrow{\mathcal{D}_v} \mu$, provided that $\mathcal{D}(\mathbb{S}) \subset \mathcal{C}(\mathbb{S})$.

²This condition is typically assumed in the literature in order to ensure the embedding $\mathcal{C}_b \subset \mathcal{D}_v$ and consequently the uniqueness of the \mathcal{D}_v -limit. However, as detailed in the Appendix, the assumption is not crucial since we can still speak about "uniqueness of limit" in a sensible way; This is precisely formulated in Lemma 3 in the Appendix.

4 Normed Spaces

This section relates the theory put forward so far to classical functional analysis. Precisely, the objects we use throughout this paper, i.e., functions and measures, can be treated as elements in some appropriate spaces. Usually, these spaces have nice structural properties which allow for a more detailed analysis. Section 4.1 deals with functional normed spaces and Section 4.2 addresses spaces of measures. Eventually, Section 4.3 extends the results from the previous sections to product spaces.

4.1 Functional Spaces

We present a formal definition of the \mathcal{D}_v -space, introduced in (5), in terms of the v -norm (to be introduced). We show by means of an example that in some common situations the resulting \mathcal{D}_v -space becomes a Banach space when endowed with the appropriate norm.

For $v \in \mathcal{C}^+(\mathbb{S})$ one introduces the so-called v -norm on $\mathcal{F}(\mathbb{S})$, as follows:

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

In particular, for each $g \in \mathcal{F}$ it holds that³:

$$\forall s \in \mathbb{S} : |g(s)| \leq \|g\|_v \cdot v(s). \quad (6)$$

Example 4. Let \mathcal{D}_v be defined as in Example 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 $\|Q\|_v = e^{-1}$ since:

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

Remark 2. The v -norm is also known as weighted supremum norm in the literature. An early reference is [13]. The v -norm is frequently used in Markov decision analysis. First traces date back to the early eighties, see [3] and the revised version which was published as [4]. It was originally used in analysis of Blackwell optimality; see [4], and [10] 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, [14] and [11]. For the use of v -norm in the theory of measure-valued differentiation of Markov chains, see [7].

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

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

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

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

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

Definition 2. Let $\mathcal{D}(\mathbb{S}) \subset \mathcal{F}(\mathbb{S})$ be a linear space and $v \in \mathcal{C}^+(\mathbb{S})$. The pair $(\mathcal{D}(\mathbb{S}), v)$ is called a **Banach base** on \mathbb{S} if:

- (i) $\mathcal{D}(\mathbb{S})$ is a linear space such that $\mathcal{C}_B(\mathbb{S}) \subset \mathcal{D}(\mathbb{S})$.
- (ii) The set $[\mathcal{D}(\mathbb{S})]_v$ 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 $v \in \mathcal{C}^+(\mathbb{S})$ the functional spaces $[\mathcal{C}(\mathbb{S})]_v$, $[\mathcal{F}(\mathbb{S})]_v$ and $[\mathcal{L}^p(\mathbb{S}, \{\mu_\theta : \theta \in \Theta\})]_v$ are Banach bases.

Example 5. The continuity paradigm: $\mathcal{D} = \mathcal{C}$. Taking $v \in \mathcal{C}^+$ we obtain $[\mathcal{C}]_v$ as the set of all continuous mappings bounded by v . It can be shown that (\mathcal{C}, v) is a Banach base on \mathbb{S} . Indeed, the mapping⁴ $\Phi : [\mathcal{C}(\mathbb{S})]_v \rightarrow \mathcal{C}_B(\mathbb{S}_v)$ defined as:

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

where \mathbb{S}_v denotes the support⁵ of v , establishes a linear bijection between two normed spaces and the inverse $\Phi^{-1} : \mathcal{C}_B(\mathbb{S}_v) \rightarrow [\mathcal{C}(\mathbb{S})]_v$ 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:

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

Since $\mathcal{C}_B(\mathbb{S}_v)$ is a Banach space when equipped with the supremum-norm, $[\mathcal{C}(\mathbb{S})]_v$ inherits the same property; see [19].

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

The \mathcal{L}^p -integrability paradigm: Let $\{\mu_\theta : \theta \in \Theta\} \subset \mathcal{M}^1$ and $v \in \mathcal{C}^+ \cap \mathcal{L}^p(\{\mu_\theta : \theta \in \Theta\})$, 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 $\Phi : [\mathcal{L}^p(\mathbb{S}, \{\mu_\theta : \theta \in \Theta\})]_v \rightarrow \mathcal{L}^\infty(\mathbb{S}_v, \{\mu_\theta : \theta \in \Theta\})$ ⁶ we conclude that $(\mathcal{L}^p(\{\mu_\theta : \theta \in \Theta\}), v)$ is a Banach base on \mathbb{S} .

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

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

⁶By writing $\mathcal{L}^\infty(\mathbb{S}_v, \{\mu_\theta : \theta \in \Theta\})$ we commit a slight abuse of notation, as the measure μ_θ is defined on \mathbb{S} and not on \mathbb{S}_v . However, since for all $\theta \in \Theta$ the measure μ_θ does not assign any mass outside \mathbb{S}_v , μ_θ coincides with its trace on \mathbb{S}_v so that the notation is justified.

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 in order to derive new results, using specific tools from Banach space theory.

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

$$\mathcal{M}_v \stackrel{\text{def}}{=} \{\mu \in \mathcal{M} : v \in \mathcal{L}^1(\mu)\}$$

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

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

Note that, using the v -norm, the space \mathcal{M}_v can be alternatively described as:

$$\mathcal{M}_v = \{\mu \in \mathcal{M} : \|\mu\|_v < \infty\}.$$

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

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

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:

$$\forall g \in [\mathcal{D}]_v, \forall \mu \in \mathcal{M}_v : \left| \int g(s)\mu(ds) \right| \leq \|g\|_v \cdot \|\mu\|_v. \quad (10)$$

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

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

Note that v -norm-convergence implies $[\mathcal{D}]_v$ -weak convergence. For example, it is known that convergence of μ_n towards μ in total variation norm implies that (4) holds for the set $\mathcal{D} = \mathcal{C}_B$. 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 $g \in [\mathcal{D}]_v$. In words, 'strong convergence implies weak convergence', which justifies the terms 'weak' and 'strong'. The converse is, however, not true as detailed in the following example.

Example 6. Consider the convergent sequence $\{x_n\}_n \subset \mathbb{R}$ having limit $x \in \mathbb{R}$. It is known that the sequence of corresponding Dirac distributions $\{\delta_{x_n}\}_n \subset \mathcal{M}$ is weakly \mathcal{C}_B -convergent to δ_x . However, strong convergence

does not hold since

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

Provided that (\mathcal{D}, v) is a Banach base, the Banach-Steinhaus-Theorem can be applied to a sequence $\{\mu_n\}_n$ of measures which allows to deduce v -norm boundedness of $\{\mu_n\}_n$ on $[\mathcal{D}]_v$ from $[\mathcal{D}]_v$ -converges of μ_n . The precise statement is provided in the following lemma.

Lemma 1. *Let (\mathcal{D}, v) be a Banach base and $\{\mu_n\}_n \subset \mathcal{M}_v$ be a $[\mathcal{D}]_v$ -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 in the lemma, the set $\{\mu_n | n \in \mathbb{N}\}$ is bounded in the weak sense, i.e., for each $g \in [\mathcal{D}]_v$, the set $\{\int g d\mu_n | n \in \mathbb{N}\}$ is bounded in \mathbb{R} . The claim then follows from the Banach-Steinhaus Theorem. \square

4.3 Product Spaces

Let \mathbb{S}, \mathbb{T} be separable complete metric spaces endowed with Borel fields \mathcal{S} and \mathcal{T} , respectively. Let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on \mathbb{S} and \mathbb{T} , respectively. The product of $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$, denoted by $(\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}), v \otimes u)$, is defined as follows:

$$\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}) = \{g : \mathbb{S} \times \mathbb{T} \rightarrow \mathbb{R} : g(s, \cdot) \in \mathcal{D}(\mathbb{T}), g(\cdot, t) \in \mathcal{D}(\mathbb{S}), \forall s \in \mathbb{S}, t \in \mathbb{T}\}, \quad (11)$$

and

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

Condition (11) imposes no restriction in applications which is illustrated in the following example.

Example 7. *We revisit the Banach bases introduced in Example 5.*

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

$$\mathcal{C}(\mathbb{S} \times \mathbb{T}) \subset \mathcal{C}(\mathbb{S}) \otimes \mathcal{C}(\mathbb{T}). \quad (13)$$

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

$$\mathcal{F}(\mathbb{S} \times \mathbb{T}) \subset \mathcal{F}(\mathbb{S}) \otimes \mathcal{F}(\mathbb{T}), \quad (14)$$

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

$$\mathcal{L}^p(\mathbb{S} \times \mathbb{T}, \{\mu_\theta \times \nu_\theta : \theta \in \Theta\}) \subset \mathcal{L}^p(\mathbb{S}, \{\mu_\theta : \theta \in \Theta\}) \otimes \mathcal{L}^p(\mathbb{T}, \{\nu_\theta : \theta \in \Theta\}). \quad (15)$$

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 1. Let $(\mathcal{D}(\mathbb{S}_i), v_i)$ be Banach bases on \mathbb{S}_i , respectively, for $1 \leq i \leq k$. Then $(\mathcal{D}(\mathbb{S}_1) \otimes \cdots \otimes \mathcal{D}(\mathbb{S}_k), v_1 \otimes \cdots \otimes v_k)$ is a Banach base on $\mathbb{S}_1 \times \cdots \times \mathbb{S}_k$. Moreover, if $g \in (\mathcal{D}(\mathbb{S}_1) \otimes \cdots \otimes \mathcal{D}(\mathbb{S}_k), v_1 \otimes \cdots \otimes v_k)$, then for $1 \leq i \leq k$ it holds that

$$g(s_1, \dots, s_{i-1}, \cdot, s_{i+1}, \dots, s_k) \in \mathcal{D}(\mathbb{S}_i),$$

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 $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on \mathbb{S} and \mathbb{T} , respectively, then $(\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}), w)$ is a Banach base on the product space $\mathbb{S} \times \mathbb{T}$; moreover, if $g \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$, then $g(s, \cdot) \in \mathcal{D}(\mathbb{T})$ for all $s \in \mathbb{S}$, and $g(\cdot, t) \in \mathcal{D}(\mathbb{S})$ for all $t \in \mathbb{T}$.

We verify the conditions in Definition 2. It's immediate that $\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})$ is a linear space, satisfying:

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

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

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

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 $g(\cdot, t) \in [\mathcal{D}(\mathbb{S})]_v$. By symmetry, we obtain $g(s, \cdot) \in [\mathcal{D}(\mathbb{T})]_u$, for all $s \in \mathbb{S}$.

Next, we show that $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$ is a Banach space w.r.t. $v \otimes u$ -norm. To this end, let $\{g_n\}_n$ be a Cauchy sequence in $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$. That means that for each $\epsilon > 0$, there exist 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:

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

Hence, for all $t \in \mathbb{T}$, $\{g_n(\cdot, t)\}_n$ is a Cauchy sequence in the Banach space $[\mathcal{D}(\mathbb{S})]_v$, thus convergent to some limit $\bar{g}(\cdot, t) \in [\mathcal{D}(\mathbb{S})]_v$. Using again a symmetry argument we deduce that $\bar{g}(s, \cdot) \in [\mathcal{D}(\mathbb{T})]_u$, for all $s \in \mathbb{S}$. Hence, we conclude that $\bar{g} \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$.

Finally, we show that \bar{g} is the $v \otimes u$ -norm limit of the sequence $\{g_n\}_n$. Choose $\epsilon > 0$ 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:

$$|g_j(s, t) - g_k(s, t)| < \epsilon \cdot v(s)u(t),$$

for all $j, k \geq n_\epsilon$ and for all s, t . Letting now $k \rightarrow \infty$ yields:

$$|g_j(s, t) - \bar{g}(s, t)| \leq \epsilon \cdot v(s)u(t),$$

for all $j \geq n_\epsilon$ and for all s, t , which is equivalent to $\|g_j - \bar{g}\|_{v \otimes u} \leq \epsilon$ for all $j \geq n_\epsilon$. Since ϵ was chosen arbitrarily, this proves the claim. \square

For $\mu \in \mathcal{M}(\mathcal{S})$, $\nu \in \mathcal{M}(\mathcal{T})$ we denote their product by $\mu \times \nu \in \mathcal{M}(\sigma(\mathcal{S} \times \mathcal{T}))$ ⁷. We conclude this section with a result which provides an upper bound for the $v \otimes u$ -norm of the product measure $\mu \times \nu$. In particular, it shows that the product measure is (strongly) continuous with respect to its components.

Lemma 2. *For $\mu \in \mathcal{M}(\mathcal{S})$ and $\nu \in \mathcal{M}(\mathcal{T})$ it holds that*

$$\|\mu \times \nu\|_{v \otimes u} \leq \|\mu\|_v \|\nu\|_u. \quad (17)$$

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 property⁸ of Hahn-Jordan decomposition ensures that:

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

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

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

which establishes (17). □

5 Differentiability

In this section we discuss two concepts of differentiability of probability measures. Both types of convergence on \mathcal{M} presented in Section 3 and Section 4 (weak and strong) induce corresponding types of differentiability. However, particular attention will be paid to weak differentiation because it is a less restrictive condition, while 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 Differentiation of Measures

Definition 3. *Let (\mathcal{D}, v) be a Banach base on \mathbb{S} . We say that the mapping $\mu : \Theta \rightarrow \mathcal{M}_v$, 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:*

$$g \in [\mathcal{D}]_v : \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left(\int g(s) \mu_{\theta+\epsilon}(ds) - \int g(s) \mu_\theta(ds) \right) = \int g(s) \mu'_\theta(ds). \quad (18)$$

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

⁷Here $\sigma(\mathcal{S} \times \mathcal{T})$ denotes the σ -field generated by the product $\mathcal{S} \times \mathcal{T}$ on $\mathbb{S} \times \mathbb{T}$.

⁸If $\mu = \mu^+ - \mu^-$ 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 $\kappa^+ - \mu^+, \kappa^- - \mu^- \in \mathcal{M}^+$.

$\mu_\theta^\pm \in \mathcal{M}^1$, satisfying:

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

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 if we note that (18) in the above definition is equivalent to

$$\forall g \in [\mathcal{D}]_v : \frac{d}{d\theta} \int g(s)\mu_\theta(ds) = \int g(s)\mu'_\theta(ds).$$

Hence, 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:

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

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

Remark 3. *Differentiability of probability measures in the weak sense as defined in Definition 3 was introduced by Pflug for $\mathcal{D} = \mathcal{C}_B$; see [15] for an early reference and the monograph [16] for a thorough treatment of \mathcal{C}_B -derivatives. Other early traces are [20, 12]. Heidergott and Vázquez-Abad [9] 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 like infinitesimal perturbation analysis; see [6].*

Remark 4. *For any $[\mathcal{D}]_v$ -differentiable probability measure μ_θ an instance of the n^{th} order $[\mathcal{D}]_v$ -derivative can be obtained via the Hahn-Jordan decomposition of $\mu_\theta^{(n)}$; see Section 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:*

$$\forall g \in [\mathcal{D}]_v : \frac{d^n}{d\theta^n} \int g(x)f(x, \theta)\mu(dx) = \int g(x) \frac{d^n}{d\theta^n} f(x, \theta)\mu(dx), \quad (19)$$

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 [16].

Example 8. *Let $\mathbb{S} = [0, \infty)$ with the usual topology, $\Theta = [a, b] \subset [0, \infty)$ and 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 \mathcal{D}_{v_p} -differentiable. Higher-order derivatives can be computed by differentiating the density $f(x, \theta) = \theta \exp(-\theta x)$ in classical sense, see Remark 4. 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:

$$\mu_\theta^{(n)} = \begin{cases} (\frac{n!}{\theta^n}, \gamma(n, \theta), \gamma(n+1, \theta)), & \text{for } n \text{ odd;} \\ (\frac{n!}{\theta^n}, \gamma(n+1, \theta), \gamma(n, \theta)), & \text{for } n \text{ even.} \end{cases}, n \geq 1.$$

Example 9. Let $\mathbb{S} = [0, \infty)$, and 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 \mathcal{D} the set $\mathcal{C}(\mathbb{S})$. Since the density $\theta^{-1} \mathbb{1}_{(0, \theta)}(x)$ is not differentiable w.r.t. θ , we calculate the weak derivative directly. Thus, by definition, for each g continuous at θ , we have:

$$\int g(s) \psi'_\theta(ds) = \lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left(\frac{1}{\theta + \epsilon} \int_0^{\theta + \epsilon} g(s) ds - \frac{1}{\theta} \int_0^\theta g(s) ds \right),$$

which yields:

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

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 \mathcal{D} -differentiable for any sensible set \mathcal{D} . Indeed, $\int g(s) \delta_\theta(ds)$ is differentiable at θ only if g is differentiable at θ . This however would impose quite strong restrictions on the performance measures to be analyzed.

Example 10. Let $\mathbb{S} = \{x_1, x_2\}$, with the discrete topology, $\Theta = [0, 1)$ and set for each $\theta \in \Theta$, $\mu_\theta = (1 - \theta)\delta_{x_1} + \theta\delta_{x_2}$, 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 $g : \mathbb{S} \rightarrow \mathbb{R}$ that:

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

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, as it can be easily seen 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 weak convergence by strong one, in (18). More precisely, μ_θ is called *strongly differentiable*, with derivative μ'_θ , if

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

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

Strong 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

$$\lim_{\epsilon \rightarrow 0} \frac{1}{\epsilon} \left| \left(\int g(s) \mu_{\theta+\epsilon}(ds) - \int g(s) \mu_\theta(ds) \right) - \int g(s) \mu'_\theta(ds) \right| = 0, \forall g \in [\mathcal{D}]_v \quad (20)$$

does, in general, not imply that

$$\lim_{\epsilon \rightarrow 0} \sup_{\|g\|_v \leq 1} \left| \frac{1}{\epsilon} \left(\int g(s) \mu_{\theta+\epsilon}(ds) - \int g(s) \mu_\theta(ds) \right) - \int g(s) \mu'_\theta(ds) \right| = 0. \quad (21)$$

The following example illustrates this fact.

Example 11. Consider the uniform distribution ψ_θ on $(0, \theta)$. In Example 9 we have shown that ψ_θ is weakly \mathcal{D} -differentiable, for $\mathcal{D} = \mathcal{C}$ and its weak derivative satisfies:

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

Hence, (20) holds true for $\psi_\theta = \mu_\theta$, for $\epsilon \neq 0$. Let $v(s) = s^p$, then it holds that

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

which violates (21). The uniform distribution on $(0, \theta)$ is thus $[\mathcal{D}]_v$ -differentiable but fails to be strongly differentiable.

The following theorem provides sufficient conditions for strong differentiability. More precisely, it is shown that weak differentiability together with strong continuity of μ'_θ implies strong differentiability.

Theorem 2. If $\mu : \Theta \rightarrow \mathcal{M}$ is weakly $[\mathcal{D}]_v$ -differentiable on Θ such that μ' is v -norm continuous, then μ is strongly v -norm differentiable on Θ .

Proof. Let $\varrho > 0$ be arbitrary and choose $\epsilon > 0$ such that:

$$\|\mu'_{\theta+\eta} - \mu'_\theta\|_v < \varrho, \forall \eta \in (-\epsilon, \epsilon).$$

Applying the Mean Value Theorem for $\theta \mapsto \int g(s) \mu_\theta(ds)$ yields:

$$\forall g \in [\mathcal{D}]_v : \int g(s) \mu_{\theta+\epsilon}(ds) - \int g(s) \mu_\theta(ds) = \epsilon \int g(s) \mu'_{\theta+\eta_g}, \quad (22)$$

for some $\eta_g \in (-\epsilon, \epsilon)$ depending on g . Hence, for all $g \in [\mathcal{D}]_v$ it holds that:

$$\begin{aligned} \left| \int g(s) \mu_{\theta+\epsilon}(ds) - \int g(s) \mu_\theta(ds) - \epsilon \int g(s) \mu'_\theta(ds) \right| \\ = |\epsilon| \cdot \left| \int g(s) \mu'_{\theta+\eta_g} - \int g(s) \mu'_\theta \right| \leq |\epsilon| \cdot \|\mu'_{\theta+\eta_g} - \mu'_\theta\|_v. \end{aligned}$$

Taking the supremum w.r.t. $\|g\|_v \leq 1$ in the above inequality, we conclude from (22) that:

$$\|\mu_{\theta+\epsilon} - \mu_\theta - \epsilon \cdot \mu'_\theta\|_v \leq \varrho |\epsilon|.$$

Dividing both sides by $|\epsilon|$ and letting $\epsilon \rightarrow 0$, proves the claim. \square

Note that Theorem 2 implies that the exponential distribution in Example 8 is strongly differentiable.

There is a tradeoff 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 5. Roughly speaking, $[\mathcal{F}]_v$ -differentiability is the most restrictive condition since it requires that $\mu_\theta(A)$ is differentiable for all $A \in \mathcal{S}$. 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 9 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 since it only requires weak convergence. 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 9 belongs to this class.

Let X be a real-valued random variable and let $-\infty \leq a < b \leq \infty$ 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:

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

If X has a probability density ρ , then the mapping

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

is the probability density of a truncated distribution.

Remark 5. Note that f as defined by (23) is still a probability density if we only require that ρ is a non-negative integrable function on (a, b) and not necessarily a density on \mathbb{R} .

Example 12. 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 9.
- (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 distribution⁹ with density

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

⁹The fact can be also derived from the memoryless property of the exponential distribution.

Truncated distributions arise naturally in applications. Indeed, consider a constant a modeling a traveling time in a transportation network. Then it is quite common to add a normal distributed noise, say Y , to a in order to model some intrinsic randomness; see [8]. Since it is important to ensure that $\mathbb{P}(Y + a < 0) = 0$ (so that traveling times stay larger than zero), one considers a truncated version of Y . In other words, the distribution of $Y + a$ is conditioned on the event (θ, ∞) for $\theta > 0$ small. 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

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

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 3. To this end, let $v \in \mathcal{C}^+(\mathbb{R})$ be such that

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

and for $g \in [\mathcal{C}]_v$ proceed as follows:

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

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 re-scaled difference between the original truncated distribution and the Dirac distribution, while higher-order derivatives do not exist, since the Dirac distribution is not differentiable; see Example 9.

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 Section 6.1. The results for general products of probability measures are presented in Section 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 measure. 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 the following theorem.

Theorem 3. 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 ϵ such that $\theta + \epsilon \in \Theta$, set:

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

We have assumed that $\bar{\mu}_\epsilon \xrightarrow{[\mathcal{D}]_v} \mathbf{0}$ and $\bar{\nu}_\epsilon \xrightarrow{[\mathcal{D}]_u} \mathbf{0}$, where $\mathbf{0}$ denotes the null measure. Simple algebra shows that the proof of the claim follows from:

$$\epsilon \cdot (\bar{\mu}_\epsilon + \mu'_\theta) \times (\bar{\nu}_\epsilon + \nu'_\theta) + \mu_\theta \times \bar{\nu}_\epsilon + \bar{\mu}_\epsilon \times \nu_\theta \xrightarrow{[\mathcal{D}]_{v \otimes u}} \mathbf{0} \quad \text{as } \epsilon \rightarrow 0. \quad (25)$$

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 2 yields

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

Since $\bar{\mu}_\epsilon + \mu'_\theta \xrightarrow{[\mathcal{D}]_v} \mu'_\theta$ and $\bar{\nu}_\epsilon + \nu'_\theta \xrightarrow{[\mathcal{D}]_u} \nu'_\theta$, applying Lemma 1 yields:

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

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

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

where $H_\theta(g, t) = \int g(s, t) \mu_\theta(ds)$ for all t and for all g . Theorem 1 implies that $(\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T}), v \otimes u)$ is a Banach base. Hence, applying Cauchy-Schwartz Inequality yields

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

where the second inequality follows from the second part of Theorem 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 $\bar{\nu}_\epsilon \xrightarrow{[\mathcal{D}]_u} \mathbf{0}$. Hence,

$$\lim_{\epsilon \rightarrow 0} \int H_\theta(g, t) \bar{\nu}_\epsilon(dt) \rightarrow 0,$$

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. \square

Remark 6. It is worth noting that the condition on the functional spaces in Theorem 3 are typically satisfied in applications; see Example 5.

Remark 7. Choosing $\mathcal{D}(\mathbb{S}) = \mathcal{C}(\mathbb{S})$, $\mathcal{D}(\mathbb{T}) = \mathcal{C}(\mathbb{T})$, $v \equiv 1$ and $u \equiv 1$, in Theorem 3 we conclude from (13) that ‘weak \mathcal{C}_B -differentiability is preserved by the product measure.’ This is asserted in [16] but 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 5.

Inspired by the resemblance of Theorem 3 with classical analysis, we proceed to establish the ‘Leibnitz-Newton’ product rule which extends Theorem 3 to higher-order derivatives. The precise statement is, as follows:

Theorem 4. Let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on \mathbb{S} and \mathbb{T} , respectively. If μ_θ is n -times $[\mathcal{D}(\mathbb{S})]_v$ -differentiable and if ν_θ is n -times $[\mathcal{D}(\mathbb{T})]_u$ -differentiable, then the product measure $\mu_\theta \times \nu_\theta \in \mathcal{M}(\sigma(\mathbb{S} \times \mathbb{T}))$ is n -times $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$ -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 3. 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 3 to evaluate the derivatives on the right-hand side, the proof follows from basic algebraic calculations, just like in conventional analysis. \square

6.2 General Products

In this section we address differentiability of n -fold products of probability measures. Our next result presents the general formula of the weak differential calculus.

Theorem 5. For $1 \leq i \leq k$, let $(\mathcal{D}(\mathbb{S}_i), v_i)$ be Banach bases on \mathbb{S}_i such that $\mu_{i,\theta}$ is n -times $[\mathcal{D}(\mathbb{S}_i)]_{v_i}$ -differentiable. Then, $\Pi_\theta \stackrel{\text{def}}{=} \mu_{1,\theta} \times \dots \times \mu_{k,\theta}$ is n -times $[\mathcal{D}(\mathbb{S}_1) \otimes \dots \otimes \mathcal{D}(\mathbb{S}_k)]_{v_1 \otimes \dots \otimes v_k}$ -differentiable on $\mathbb{S}_1 \times \dots \times \mathbb{S}_k$ and it holds that:

$$\Pi_\theta^{(n)} = \sum_{\tilde{j} \in \mathcal{J}(k,n)} \binom{n}{j_1, \dots, j_k} \cdot (\mu_{1,\theta})^{(j_1)} \times \dots \times (\mu_{k,\theta})^{(j_k)}, \quad (27)$$

where

$$\mathcal{J}(k,n) \stackrel{\text{def}}{=} \{\tilde{j} = (j_1, \dots, j_k) : 0 \leq j_i \leq n, j_1 + \dots + j_k = n\},$$

for $k, n \geq 1$.

Proof. The proof follows from Theorem 1 and Theorem 4 via finite induction. \square

An instance of a derivative of the product measure $\Pi_\theta^{(n)}$, in Theorem 5, can be obtained by inserting the appropriate \mathcal{D}_{v_i} derivatives for the measures $\mu_{i,\theta}^{(j_i)}$ and rearranging terms in (27). In order to present the result

we introduce the following notations. For $\tilde{j} = (j_1, \dots, j_k) \in \mathcal{J}(k, n)$ we denote by $\xi(\tilde{j})$ the number of non-zero elements of the vector \tilde{j} and by $\mathcal{I}(\tilde{j})$ 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 non-zero elements of η equals one, i.e., there is an even number of -1 's. For $\eta \in \mathcal{I}(\tilde{j})$, we denote by $\bar{\eta}$ the vector obtained from η by changing the sign of the non-zero element at the highest position. The precise statement is as follows:

Corollary 1. *Under the conditions put forward in Theorem 5, let μ_θ^i have m^{th} -order \mathcal{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:

$$\begin{aligned} \Gamma_\theta^{(n)} &= \sum_{\tilde{j} \in \mathcal{J}(k,n)} 2^{\xi(\tilde{j})-1} \binom{n}{j_1, \dots, j_k} \prod_{i=1}^k c_{i,\theta}^{(j_i)}, \\ \Pi_\theta^{(n,+)} &= \sum_{\tilde{j} \in \mathcal{J}(k,n)} \binom{n}{j_1, \dots, j_k} \frac{\prod_{i=1}^k c_{i,\theta}^{(j_i)}}{\Gamma_\theta^{(n)}} \sum_{\eta \in \mathcal{I}(\tilde{j})} \mu_{1,\theta}^{(j_1, \eta_1)} \times \mu_{2,\theta}^{(j_2, \eta_2)} \times \dots \times \mu_{k,\theta}^{(j_k, \eta_k)}, \\ \Pi_\theta^{(n,-)} &= \sum_{\tilde{j} \in \mathcal{J}(k,n)} \binom{n}{j_1, \dots, j_k} \frac{\prod_{i=1}^k c_{i,\theta}^{(j_i)}}{\Gamma_\theta^{(n)}} \sum_{\eta \in \mathcal{I}(\tilde{j})} \mu_{1,\theta}^{(j_1, \bar{\eta}_1)} \times \mu_{2,\theta}^{(j_2, \bar{\eta}_2)} \times \dots \times \mu_{k,\theta}^{(j_k, \bar{\eta}_k)}, \end{aligned}$$

where, for convenience, we identify

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

Example 13. *Consider the Banach base $\mathcal{C}_B(\mathbb{S}) = ((\mathcal{C}(\mathbb{S}), v)$ for $v = 1$. Denote the k -fold product of μ_θ by $\Pi_\theta(k)$. Suppose that μ_θ has $\mathcal{C}_B(\mathbb{S})$ -derivative $(c_\theta, \mu_\theta^+, \mu_\theta^-)$. Then, by Theorem 5, $\Pi_\theta(n)$ is $\mathcal{C}_B(\mathbb{S}^n)$ -differentiable and an instance of a $\mathcal{C}_B(\mathbb{S}^n)$ -derivative can be obtained from Corollary 1. This yields for any $g \in \mathcal{C}_B(\mathbb{S}^n)$*

$$\begin{aligned} &\frac{d}{d\theta} \int g(s) \Pi_\theta(n, ds) \\ &= c_\theta \sum_{j=1}^n \left(\int g(s, t, u) \Pi_\theta(n-j, ds) \times \mu_\theta^+(dt) \times \Pi_\theta(j-1, du) \right. \\ &\quad \left. - \int g(s, t, u) \Pi_\theta(n-j, ds) \times \mu_\theta^-(dt) \times \Pi_\theta(j-1, du) \right). \end{aligned}$$

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

$$\begin{aligned} \frac{d}{d\theta} J_g(\theta) &= c_\theta \sum_{j=1}^n \left(\mathbb{E}[g(X_n, \dots, X_{j+1}, X_\theta^+, X_{j-1}, \dots, X_1)] \right. \\ &\quad \left. - \mathbb{E}[g(X_n, \dots, X_{j+1}, X_\theta^-, X_{j-1}, \dots, X_1)] \right), \end{aligned}$$

for any $g \in \mathcal{C}_B(\mathbb{S}^n)$.

7 Weak Analyticity

In this section we introduce the concept of weak \mathcal{D}_v -analyticity for probability measures. Results regarding the radius of convergence of the Taylor series and weak analyticity of product measures are also provided.

Definition 4. Let (\mathcal{D}, v) be a Banach base on \mathbb{S} . The mapping $\mu : \Theta \rightarrow \mathcal{M}_v$ is called **weakly $[\mathcal{D}]_v$ -analytic** at θ , or **weakly $[\mathcal{D}]_v$ -analytic** for short, if

- all higher-order $[\mathcal{D}]_v$ -derivatives of μ_θ exist;
- there exists a neighborhood V of θ such that for all Δ s.t. $\theta + \Delta \in V$:

$$\forall g \in [\mathcal{D}]_v : \int g(s)\mu_{\theta+\Delta}(ds) = \sum_{n=0}^{\infty} \frac{\Delta^n}{n!} \cdot \int g(s)\mu_\theta^{(n)}(ds). \quad (28)$$

For fixed $g \in [\mathcal{D}]_v$, the maximal set $D_\theta(g, \mu)$ for which the equality in (28) 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 (28) depends in general on g . Our next result provides a set $D_\theta^v(\mu) \subset \Theta$ where the Taylor series in (28) converges for all $g \in [\mathcal{D}]_v$. The precise statement is as follows:

Theorem 6. Let (\mathcal{D}, v) be a Banach base on \mathbb{S} such that μ_θ is \mathcal{D}_v -analytic. Then for each $g \in [\mathcal{D}]_v$ the Taylor series in (28) converges for all Δ such that $|\Delta| < R_\theta^v(\mu)$, where $R_\theta^v(\mu)$ is given by:

$$\frac{1}{R_\theta^v(\mu)} = \limsup_{n \in \mathbb{N}} \left(\frac{\|\mu_\theta^{(n)}\|_v}{n!} \right)^{\frac{1}{n}}. \quad (29)$$

In particular, the set $D_\theta^v(\mu) \stackrel{\text{def}}{=} \Theta \cap (\theta - R_\theta^v(\mu), \theta + R_\theta^v(\mu))$ satisfies:

$$\forall g \in [\mathcal{D}]_v : D_\theta^v(\mu) \subset D_\theta(g, \mu).$$

Proof. According to the Cauchy-Hadamard Theorem, see, e.g., [18], the radius of convergence $R_\theta(g, \mu)$ of the Taylor series in (28) is given by:

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

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

$$R_\theta^v(\mu) \leq \inf\{R_\theta(g, \mu) : g \in [\mathcal{D}]_v\}. \quad (30)$$

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

$$\left| \int g(s)\mu_\theta^{(n)}(ds) \right|^{\frac{1}{n}} \leq \left(\|g\|_v \cdot \|\mu_\theta^{(n)}\|_v \right)^{\frac{1}{n}},$$

together with the fact that $\lim_{n \rightarrow \infty} \sqrt[n]{\|g\|_v} = 1$, for $g \in [\mathcal{D}]_v$, concludes the proof. \square

The non-negative number $R_\theta^v(\mu)$ is called the $[\mathcal{D}]_v$ -radius of convergence of μ_θ and the set $D_\theta^v(\mu)$ is called the $[\mathcal{D}]_v$ -domain of convergence of μ_θ . In particular, Theorem 6 shows that the Taylor series converges strongly for $|\Delta| < R_\theta^v(\mu)$. Note, however, that in general this is not the maximal set for which the series converges for all $g \in [\mathcal{D}]_v$ since the inequality in (30) may be strict.

Example 14. Let us revisit Example 8 and consider the exponential distribution with rate θ denoted by μ_θ . We aim to determine the $[\mathcal{D}]_v$ -radius of convergence $R_\theta^v(\mu)$ of μ_θ , for $v(x) = 1 + x$, $\forall x \geq 0$, which will show that the exponential distribution μ_θ is weakly analytical, for $\theta > 0$.

Recall that an instance of the n^{th} -order derivative $\mu_\theta^{(n)}$ is given by

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

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:

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

In order to show analyticity we have to show that (28) holds true for $|\Delta| < \theta$. First, we note that the density $f(x, \theta)$ of μ_θ is analytical (in classical sense) in θ , i.e.,

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

Hence, (28) is equivalent to:

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

Fix $g \in [\mathcal{F}]_v$. In order 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 8, yields:

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

Since 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 $[\mathcal{F}]_v$ -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 classical analysis it is well known that the product of two analytical functions is again analytical. The following theorem establishes the counterpart of this fact for weak analyticity of measures.

Theorem 7. *Let $(\mathcal{D}(\mathbb{S}), v)$ and $(\mathcal{D}(\mathbb{T}), u)$ be Banach bases on \mathbb{S} and \mathbb{T} , respectively. Let μ_θ be $[\mathcal{D}(\mathbb{S})]_v$ -analytic with domain of convergence $D_\theta^v(\mu)$, and let ν_θ be $[\mathcal{D}(\mathbb{T})]_u$ -analytic with domain of convergence $D_\theta^u(\nu)$. Then, $\mu_\theta \times \nu_\theta$ is $[\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$ -analytic and for each $g \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$, and the domain of convergence $D_\theta(g, \mu \times \nu)$ satisfies:*

$$D_\theta^v(\mu) \cap D_\theta^u(\nu) \subset D_\theta(g, \mu \times \nu).$$

Proof. For $\theta \in \Theta$ let us denote by $(\mu \times \nu)_\theta$ the product measure $\mu_\theta \times \nu_\theta$. Recall that $D_\theta^v(\mu) = \Theta \cap (\theta - R_\theta^v(\mu), \theta + R_\theta^v(\mu))$ with $R_\theta^v(\mu)$ as defined in (29). Similarly, $D_\theta^u(\nu) = \Theta \cap (\theta - R_\theta^u(\nu), \theta + R_\theta^u(\nu))$. Let $\rho = \min\{R_\theta^v(\mu), R_\theta^u(\nu)\}$ and choose $g \in [\mathcal{D}(\mathbb{S}) \otimes \mathcal{D}(\mathbb{T})]_{v \otimes u}$, arbitrarily. We show that for $|\Delta| < \rho$, it holds that:

$$\int g(s, t) (\mu \times \nu)_{\theta + \Delta}(ds, dt) = \lim_{n \rightarrow \infty} \sum_{k=0}^n \frac{\Delta^k}{k!} \int g(s, t) (\mu \times \nu)_\theta^{(k)}(ds, dt). \quad (31)$$

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

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

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, $H(t) = \lim_{n \rightarrow \infty} H_n(t)$. 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-Schwarz inequality yields

$$|H_n(t)| = |T_n(g(\cdot, t))| \leq \|T_n\|_v \|g(\cdot, t)\|_v \leq \left(\sup_n \|T_n\|_v \right) \|g\|_{v \otimes u} u(t). \quad (32)$$

In order to show that $\sup_n \|T_n\|_v < \infty$, we note that weak analyticity of μ_θ implies that $\{T_n(f) : n \in \mathbb{N}\}$ is bounded for each $f \in [\mathcal{D}(\mathbb{S})]_v$ and apply the Banach-Steinhaus Theorem; see Remark 1.

Thus, $H_n \in [\mathcal{D}(\mathbb{T})]_u$ and since $u \in \mathcal{L}^1(\{\nu_\theta : \theta \in \Theta\})$ the Dominated Convergence Theorem applies to the sequence $\{H_n\}_n$. Hence, interchanging limit with integration on the right-hand side of (31) is justified, which

yields:

$$\int H(t)\nu_{\theta+\Delta}(dt) = \int \lim_{n \rightarrow \infty} H_n(t) \nu_{\theta+\Delta}(dt) = \lim_{n \rightarrow \infty} \int H_n(t)\nu_{\theta+\Delta}(dt). \quad (33)$$

Moreover, due to $[\mathcal{D}(\mathbb{T})]_u$ -analyticity of ν_θ , the right-hand side in (33) equals to:

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

Finally, inserting the expression of $H_n(t)$ in the above expression, we conclude that the left-hand side of (31) equals to:

$$\lim_{n \rightarrow \infty} \lim_{m \rightarrow \infty} \sum_{l=0}^m \sum_{j=0}^n \frac{\Delta^{j+l}}{j!l!} \int \int g(s, t)\mu_\theta^{(j)}(ds)\nu_\theta^{(l)}(dt). \quad (34)$$

According to Theorem 4, the right-hand side of (31) can be re-written as:

$$\lim_{k \rightarrow \infty} \sum_{0 \leq j+l \leq k} \frac{\Delta^{j+l}}{j!l!} \int \int g(s, t)\mu_\theta^{(j)}(ds)\nu_\theta^{(l)}(dt). \quad (35)$$

The power series in (34) is convergent for $|\Delta| < \rho$. Hence it is absolutely convergent, so its limit is not affected by re-shuffling terms; see [18]. It follows that the limits in (34) and (35) coincide and (31) holds true for $|\Delta| < \rho$. The fact that $D_\theta^v(\mu) \cap D_\theta^v(\nu) = \Theta \cap (\theta - \rho, \theta + \rho)$, 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 $\{Y_i : i \geq 1\}$ arrive according to a Poisson process with rate $\lambda > 0$. Let $\{X_i : i \geq 1\}$ denote inter-arrival 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 12 (ii); and 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 n^{th} claim, denoted by $V(n)$, follows the recurrence relation

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

Ruin occurs before time τ if at least one $n \leq N_\tau$ exists such that $V(n) < 0$. See Figure 3.

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 \mathfrak{R}_τ the event that *ruin occurs up to time τ* and note that, given the event $\{N_\tau = n\}$ it can be expressed as follows:

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

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

$$\begin{aligned} g_n(x_1, \dots, x_n, y_1, \dots, y_n) \\ = 1 - \prod_{j=1}^n \mathbb{I}_{\{r \cdot \sum_{i=1}^j x_i > \sum_{i=1}^j y_i\}}(x_1, \dots, x_n, y_1, \dots, y_n) \end{aligned} \quad (36)$$

we can write

$$\forall n \geq 1 : \mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau = n\}) = \mathbb{E}_\theta [\mathbb{I}_{\{N_\tau = n\}} g_n(X_1, \dots, X_n, Y_1, \dots, Y_n)] \quad (37)$$

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

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

Applying Theorem 5 with $v = 1$ yields that the product measure $\mu \times \pi_\theta$ is weakly $\mathcal{C}_B(\mathbb{S}^2)$ -differentiable with $(\mu \times \pi_\theta)' = \mu \times \pi'_\theta$ (for a proof use the fact that μ is independent of θ). Applying Theorem 5 with $v = 1$ again to the n -fold product of $\mu \times \pi_\theta$ yields that $(\mu \times \pi_\theta)^n$ is weakly $\mathcal{C}_B(\mathbb{S}^{2n})$ -differentiable. Hence, for any $g \in \mathcal{C}_B(\mathbb{S}^{2n})$, the derivative of the $\int g d(\mu \times \pi_\theta)^n$ can be obtained in closed form. See Example 13 for the derivative expression. Note, however, that the sample performance g_n introduced for modeling the ruin probability lies not in $\mathcal{C}_B(\mathbb{S}^{2n})$. Fortunately, since 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 $\mathcal{C}_B(\mathbb{S}^n) \cup \{g_n\}$ -differentiable; see Section II in the Appendix for details. Hence, we arrive at

$$\begin{aligned} \frac{d}{d\theta} \mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau = n\}) &= \frac{d}{d\theta} \mathbb{E}_\theta [\mathbb{I}_{\{N_\tau = n\}} g_n(X_1, \dots, X_n, Y_1, \dots, Y_n)] \\ &= \frac{\beta}{\theta} \sum_{i=1}^n \mathbb{E}_\theta \left[\mathbb{I}_{\{N_\tau = n\}} g_n(X_1, \dots, X_n, Y_1, \dots, Y_n) \right. \\ &\quad \left. - \mathbb{I}_{\{N_\tau = n\}} g_n(X_1, \dots, X_n, Y_1, \dots, Y_{j-1}, \theta, Y_{j+1}, \dots, Y_n) \right] \\ &= \frac{\beta}{\theta} \sum_{i=1}^n (\mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau = n\}) - \mathbb{P}_\theta(\mathfrak{R}_\tau^i \cap \{N_\tau = n\})), \end{aligned}$$

where \mathfrak{R}_τ^i denotes the event that there is ruin up to time τ , when the value of the i^{th} claim is replaced by the constant θ , i.e.,

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

Provided that interchanging limit with differentiation is allowed we obtain

$$\frac{d}{d\theta} \mathbb{P}_\theta(\mathfrak{R}_\tau) = \frac{d}{d\theta} \sum_{n=1}^{\infty} \mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau = n\}) \quad (38)$$

$$\begin{aligned} &= \sum_{n=1}^{\infty} \frac{d}{d\theta} \mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau = n\}) \\ &= \sum_{n=1}^{\infty} \frac{\beta}{\theta} \sum_{i=1}^n (\mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau = n\}) - \mathbb{P}_\theta(\mathfrak{R}_\tau^i \cap \{N_\tau = n\})) \\ &= \frac{\beta}{\theta} (\mathbb{P}_\theta(\mathfrak{R}_\tau \cap \{N_\tau \geq n\}) - \mathbb{P}_\theta(\mathfrak{R}_\tau^n \cap \{N_\tau \geq n\})). \end{aligned} \quad (39)$$

Note that the n^{th} remainder term of the series in (39) is bounded by:

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

Since the bound is independent of θ and converges to 0 as $n \rightarrow \infty$, it means that we deal with an uniformly convergent series of functions of θ , so that interchanging limit with differentiation in (38) is justified.

Taking into account that $Y_n > \theta$ a.s. a sample path analysis together with a monotonicity argument yield $\mathfrak{R}_\tau^n \subset \mathfrak{R}_\tau$. Moreover, the difference $\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^n$ represents the event that ruin occurs up to time τ but it does not occur anymore if one reduces the value of the n^{th} claim by $Y_n - \theta$; a graphical representation of these facts can be found in Figure 3. One can easily note that this event is incompatible with $\{N_\tau < n\}$, i.e., if the "reduced claim" comes after time τ . Hence, it holds that $\mathbb{P}_\theta((\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^n) \cap \{N_\tau < n\}) = 0$, so that (39) becomes:

$$\frac{d}{d\theta} \mathbb{P}_\theta(\mathfrak{R}_\tau) = \frac{\beta}{\theta} \sum_{n=1}^{\infty} \mathbb{P}_\theta(\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^n). \quad (40)$$

Remark 8. *Following the line of argument that lead from (38) to (39) we obtain*

$$\begin{aligned} \frac{d}{d\theta} \mathbb{P}_\theta(\mathfrak{R}_\tau) &= \sum_{n=1}^{\infty} \frac{\beta}{\theta} \sum_{i=1}^n \mathbb{E}_\theta \left[\mathbb{I}_{\{N_\tau = n\}} g_n(X_1, \dots, X_n, Y_1, \dots, Y_n) \right. \\ &\quad \left. - \mathbb{I}_{\{N_\tau = n\}} g_n(X_1, \dots, X_n, Y_1, \dots, Y_{j-1}, \theta, Y_{j+1}, \dots, Y_n) \right] \\ &= \frac{\beta}{\theta} \mathbb{E}_\theta \left[N_\tau g_{N_\tau}(X_1, \dots, X_{N_\tau}, Y_1, \dots, Y_{N_\tau}) \right. \\ &\quad \left. - \sum_{j=1}^{N_\tau} g_{N_\tau}(X_1, \dots, X_{N_\tau}, Y_1, \dots, Y_{j-1}, \theta, Y_{j+1}, \dots, Y_{N_\tau}) \right], \end{aligned}$$

c.f. Example 13. The expression on the righthand side provides an unbiased estimator for the derivative of the ruin probability. For details on the relation between weak derivatives and unbiased estimators, we refer to [9].

8.2 Stochastic Activity Networks

Stochastic Activity Networks (SAN) such as those arising in Project Evaluation Review Technique (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, which 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 1 below. The network has 5 nodes, labeled from 1 (source) to 5 (sink) and the edges denote the activities under consideration. The weights X_i , $1 \leq i \leq 7$, denote the durations of the corresponding activities. For instance, activity 6 can only begin when both activities 2 and 3 have finished.

Place Figure 1 here.

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:

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

For more details on SAN, we refer to [17]. For instance, in the above example, the set of paths from source node 1 to sink node 5, 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 [5]. In the following 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 presented in this paper.

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

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

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

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

where we denote by μ_i the distribution of X_i , for $1 \leq i \leq 7$. In accordance with Theorem 5 it holds that if μ_i is weakly differentiable with respect to some parameter θ , for all $1 \leq i \leq 7$, then the distribution of T is

weakly differentiable w.r.t. θ , as well. Roughly speaking, that means that “the distribution of T is differentiable w.r.t. each μ_i ”¹⁰.

Assume for instance that r.v. X_i , $1 \leq i \leq 7$ are independent and exponentially distributed, with rates λ_i . We let $\lambda_1 = \lambda_3 = \theta$ variable, the other rates being fixed, i.e., independent of θ . By Example 14, the exponential distribution is weakly $[\mathcal{F}]_v$ -analytical, for $v(x) = 1 + x$ and the domain of convergence is given by $|\Delta| < \theta$. Since the distributions which are independent of θ are trivially weakly analytical, from Theorem 7 we conclude that the joint distribution of the vector (X_1, \dots, X_7) is weakly $[\mathcal{F}(S^7)]_{v \otimes \dots \otimes v}$ -analytical. Moreover, the radius of convergence of the Taylor series is equal to θ . Finally, we note that

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

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

Next we proceed to computation of derivatives, in accordance with Corollary 1. Since only the derivatives of $\mu_{1,\theta}$ and $\mu_{3,\theta}$ are significant, inspired by Example 8, for $j, k \geq 0$ 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 whereas 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 $jk = 0$. With this notation Theorem 5 yields

$$\forall n \geq 0 : \frac{d^n}{d\theta^n} \mathbb{E}_\theta[T] = (-1)^n \frac{n!}{\theta^n} \sum_{i+j=n} \mathbb{E}_\theta[T_{i+1,j+1} - T_{i+1,j} - T_{i,j+1} + T_{i,j}] \quad (41)$$

and for each $n \geq 1$ we call

$$\mathbf{T}_n(\theta, \Delta) \stackrel{\text{def}}{=} \sum_{k=0}^n (-1)^k \left(\frac{\Delta}{\theta}\right)^k \sum_{i+j=k} \mathbb{E}_\theta[T_{i+1,j+1} - T_{i+1,j} - T_{i,j+1} + T_{i,j}] \quad (42)$$

the n^{th} order Taylor polynomial for $\mathbb{E}_{\theta+\Delta}[T]$, at θ , where \mathbb{E}_θ denotes the expectation operator w.r.t. the product measure $\mu_{1,\theta} \times \mu_2 \times \mu_{3,\theta} \times \mu_4 \times \mu_5 \times \mu_6 \times \mu_7$. Using a monotonicity argument, one can easily check that

$$\forall i, j \geq 0 : |\mathbb{E}_\theta[T_{i+1,j+1} - T_{i+1,j} - T_{i,j+1} + T_{i,j}]| \leq \mathbb{E}_\theta[X_1] + \mathbb{E}_\theta[X_3] = \frac{2}{\theta}. \quad (43)$$

Hence, a bound for the error of the n^{th} order Taylor polynomial is given by

$$\begin{aligned} \forall |\Delta| < \theta : |\mathbb{E}_{\theta+\Delta}[T] - \mathbf{T}_n(\theta, \Delta)| &\leq \frac{2}{\theta} \sum_{k=n+1}^{\infty} (k+1) \left(\frac{|\Delta|}{\theta}\right)^k \\ &= \frac{2(n+2) - (n+1)\frac{|\Delta|}{\theta}}{\left(1 - \frac{|\Delta|}{\theta}\right)^2} \left(\frac{|\Delta|}{\theta}\right)^{n+1} \\ &\leq \frac{2(n+1)}{(\theta - |\Delta|)} \left(\frac{|\Delta|}{\theta}\right)^{n+1}. \end{aligned} \quad (44)$$

¹⁰Note that for a deterministic system, i.e., all the weights are deterministic, the completion time is, in general, not everywhere differentiable w.r.t. the weights. That is because the Dirac distribution δ_θ is not weakly differentiable w.r.t. θ .

Example 15. *In order to perform a numerical experiment, we consider the following rates:*

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

Computing the coefficients of the Taylor polynomial is quite demanding and it is worth noting that the coefficients can alternatively be evaluated by simulation. Figure 2 shows the Taylor polynomial $\mathbf{T}_3(1, \Delta)$ of order 3 compared to the true value of $\mathbb{E}[T_{1+\Delta}]$, for $|\Delta| \leq 0.6$.

Place Figure 2 here.

As the figure shows, the Taylor polynomial approximates the true function quite well, for $|\Delta| \leq 0.4$. Indeed, the relative error, according to (44) is below 3.4%.

Acknowledgement

The authors are grateful to Arie Hordijk for pointing out to them the result in Theorem 2.

References

- [1] Billingsley, B. *Weak Convergence of Probability Measures*. J.Wiley and Sons, New York, 1966.
- [2] Cohn, D. *Measure Theory*. Birkäuser, Stuttgart, 1980.
- [3] Dekker, R. and Hordijk, A. Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. Report no.83-36, Institute of Applied Mathematics and Computing Science, 1983.
- [4] Dekker, R. and Hordijk, A. Average, sensitive and Blackwell optimal policies in denumerable Markov decision chains with unbounded rewards. *Mathematics of Operations Research*, 13, pp. 395–421, 1988.
- [5] Devroye, L.P. Inequalities for completion times of stochastic pert networks. *Mathematics of Operation Research*, 4, pp.441–447, 1979.
- [6] B. Heidergott, F. Vázquez-Abad and W. Volk-Makarewicz. Sensitivity estimation for gaussian systems. *European Journal of Operational Research*, (to appear).
- [7] Heidergott, B. and Hordijk, A. Taylor series expansions for stationary Markov chains. *Advances in Applied Probability*, 23, pp.1046–1070, 2003.
- [8] Heidergott, B. and Vázquez-Abad, F. Gradient estimation for a class of systems with bulk services: a problem in public transportation. (*submitted*), 2007.

- [9] Heidergott, B. and Vázquez-Abad, F. Measure valued differentiation for Markov chains. *Journal of Optimization and Applications*, to appear.
- [10] Hordijk, A. and Yushkevich, A.A. Blackwell optimality in the class of all policies in Markov decision chains with a Borel state space and unbounded rewards. *Mathematics of Operations Research*, 50, pp. 421–448, 1999.
- [11] Kartashov, N. *Strong Stable Markov Chains*. VSP, Utrecht, 1996.
- [12] Kushner, H. and Vázquez-Abad, F. Estimation of the derivative of a stationary measure with respect to a control parameter. *Journal of Applied Probability*, 29, pp. 343–352, 1992.
- [13] Lipman, S. On dynamic programming with unbounded rewards. *Management Science*, 21, pp. 1225–1233, 1974.
- [14] Meyn, S.P. and Tweedie, R.L. *Markov Chains and Stochastic Stability*. Springer, London, 1993.
- [15] Pflug, G. Derivatives of probability measures - concepts and applications to the optimization of stochastic systems. In *Lecture Notes in Control and Information Science 103*, pages 252–274. Springer, Berlin, 1988.
- [16] Pflug, G. *Optimization of Stochastic Models*. Kluwer Academic, Boston, 1996.
- [17] Pich, M., Loch, C., and de Meyer, A. On uncertainty, ambiguity and complexity in project management. *Management Science*, 75(2), pp. 137–176, 1996.
- [18] Rudin, W. *Principles of Mathematical Analysis (Third Edition)*. McGraw-Hill, Tokyo, 1976.
- [19] Semadeni, Z. *Banach Spaces of Continuous Functions*. Polish Scientific Publishers, Warszawa, 1971.
- [20] Vázquez-Abad, F. and Kushner, H. Stochastic approximations for systems of interest over an infinite time interval. *SIAM Journal on Control and Optimization*, 29, pp. 712–756, 1996.

Appendix

I. Uniqueness of the Weak \mathcal{D}_v -Limit

The \mathcal{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 $s \in \mathbb{S}$, and denote by δ_0 the Dirac measure, i.e., δ_0 assigns mass 1 to point 0. Assume that μ is a \mathcal{D}_v -limit of the sequence $\{\mu_n\}_n \subset \mathcal{M}$. Since for $g \in \mathcal{D}_v$ we have $g(0) = 0$, $\mu + \alpha \cdot \delta_0 \in \mathcal{M}$ is also a \mathcal{D}_v -limit of the sequence $\{\mu_n\}_n$, for each $\alpha \in \mathbb{R}$ and the \mathcal{D}_v -limit fails to be unique. In words, (4) still holds true if one assigns a different mass on the 'zero set' of v . Our next result will elucidate this issue.

In particular it shows that the set \mathcal{D}_v , likewise \mathcal{C}_B , is appropriate for introducing weak convergence.

Lemma 3. Let $v \in \mathcal{C}^+(\mathbb{S})$ and let $\mathbb{S}_v = \{s \in \mathbb{S} : v(s) > 0\}$. If $\mu, \nu \in \mathcal{M}(\mathbb{S})$ be such that $v \in \mathcal{L}^1(\{\mu, \nu\})$ and

$$\forall g \in \mathcal{D}_v : \int g(s)\mu(ds) = \int g(s)\nu(ds), \quad (45)$$

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

$$\forall A \in \mathcal{S} : \mu(A \cap \mathbb{S}_v) = \nu(A \cap \mathbb{S}_v). \quad (46)$$

Proof. Since \mathcal{S} is the Borel field of \mathbb{S} we may assume without losing generality that $A \in \mathcal{S}$ is an arbitrary non-empty open set. For $\epsilon > 0$ consider the set:

$$A_\epsilon \stackrel{\text{def}}{=} \{s \in A : \rho(s, \mathbb{C}A) \geq \epsilon^{-1}\} \subset A,$$

where, for $E \subset \mathbb{S}$ we denote $\mathbb{C}E = \mathbb{S} \setminus E$ and $\rho(s, E) = \inf\{\rho(s, t) : t \in E\}$. Note that, for sufficiently large $\epsilon > 0$, A_ϵ is a non-empty closed set satisfying $A_\epsilon \cap \mathbb{C}A = \emptyset$. Since A is an open set, i.e., $\mathbb{C}A$ is closed, according to Urysohn's Lemma there exists a continuous function $f_\epsilon : \mathbb{S} \rightarrow [0, 1]$ such that $f_\epsilon(x) = 1$ for $x \in A_\epsilon$ and $f_\epsilon(x) = 0$, for $x \in \mathbb{C}A$. On the other hand the family $\{A_\epsilon\}_{\epsilon > 0} \subset \mathcal{F}$ is ascendent and $\cup_{\epsilon > 0} A_\epsilon = A$. Hence, f_ϵ converges point-wise to \mathbb{I}_A , as $\epsilon \rightarrow \infty$.

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

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

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

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

Applying now the Dominated Convergence Theorem yields:

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

which concludes the proof of (46). □

Remark 9. If we denote by \sim the equivalence relation on \mathcal{M} given by $\mu \sim \nu$ if (46) holds true then Lemma 3 shows that if (45) holds true then $\mu \sim \nu$. Going back to (4), we conclude that the \mathcal{D}_v -limit is uniquely determined up to this equivalence relation and the precise definition of the \mathcal{D}_v -limit would be in terms of the equivalence class of μ , denoted by $[\mu]$. Note that since $g \in \mathcal{D}_v$ 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 \mathcal{D}_v^* of \mathcal{D}_v , i.e., the set of all linear functionals on \mathcal{D}_v , is \mathcal{M}/\sim , i.e., the quotient space of \mathcal{M} w.r.t. equivalence relation \sim .

II. Continuity Sets

According to Definition 3, weak \mathcal{C}_B -convergence can only handle continuous performance measures. In fact, the class of mappings g which satisfy equation (18) is much larger and includes, for instance, the indicator functions

of the so-called *continuity sets*. This fact is well known for classical weak convergence of probability measures; see, e.g., the *portmanteau theorem* in [1]. In the following, we show that a similar result holds true for weak differentiation of measures. In order to be able to use the results known from classical weak convergence theory we introduce the concept of *regular weak differentiability*.

Definition 5. We say that $\mu : \Theta \rightarrow \mathcal{M}$ is regular weakly $[\mathcal{D}]_v$ -differentiable at θ if for each h there exist a decomposition

$$\frac{\mu_{\theta+h} - \mu_\theta}{h} = \left[\frac{\mu_{\theta+h} - \mu_\theta}{h} \right]^+ - \left[\frac{\mu_{\theta+h} - \mu_\theta}{h} \right]^-$$

such that for $h \rightarrow 0$ it holds that

$$\left[\frac{\mu_{\theta+h} - \mu_\theta}{h} \right]^+ \xrightarrow{[\mathcal{D}]_v} \mu_\theta^p \quad \text{and} \quad \left[\frac{\mu_{\theta+h} - \mu_\theta}{h} \right]^- \xrightarrow{[\mathcal{D}]_v} \mu_\theta^m.$$

Note that, if μ_θ is regular weakly differentiable, then it is weakly differentiable cf. Definition 3, and its weak derivative satisfies

$$\mu'_\theta = \mu_\theta^p - \mu_\theta^m.$$

Moreover, it can be shown without difficulty that, if μ_θ is regular weakly differentiable then the n -fold product $\mu_\theta \times \dots \times \mu_\theta$ is regular weakly differentiable.

Example 16. *The Pareto distribution*

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

is regular weakly differentiable. Indeed, for $h > 0$, we can write:

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

Hence, it holds that

$$\left[\frac{\pi_{\theta+h} - \pi_\theta}{h} \right]^+(dx) = \frac{\beta((\theta+h)^\beta - \theta^\beta)}{hx^{\beta+1}} \mathbb{I}_{(\theta+h, \infty)}(x) dx \xrightarrow{\mathcal{C}_B} \frac{\beta}{\theta} \pi_\theta(dx)$$

and

$$\left[\frac{\pi_{\theta+h} - \pi_\theta}{h} \right]^-(dx) = -\frac{\beta\theta^\beta}{hx^{\beta+1}} \mathbb{I}_{(\theta, \theta+h]}(x) dx \xrightarrow{\mathcal{C}_B} \frac{\beta}{\theta} \delta_\theta(dx).$$

One can proceed similarly for $h < 0$.

We say that A is a *continuity set* for $\mu \in \mathcal{M}^+$ if $\mu(\partial A) = 0$, where ∂A denotes the boundary of A . The next statement provides sufficient conditions for "set wise" differentiation.

Lemma 4. If μ_θ is regular-weakly \mathcal{C}_B -differentiable and if $A \in \mathcal{F}$ is a continuity set for both μ_θ^p and μ_θ^m , then it holds that:

$$\frac{d}{d\theta} \mu_\theta(A) = \mu'_\theta(A).$$

Proof. The conclusion follows from the portmanteau theorem¹¹; see [1]. □

¹¹The cited result is formulated in terms of probability measures, only. However, extension to positive measures is straightforward.

Example 17. Recall the situation in Section 8.1. In order to be able to differentiate $\mathbb{P}_\theta(\mathfrak{A}_\tau)$ it suffices to look at the following. Let μ denote the exponential distribution and let X be distributed according to μ and let Y_i be distributed according to $\pi_{i,\theta}$, for $1 \leq i \leq j$. Then, the set

$$\left\{ r \cdot \sum_{i=1}^j x_i = \sum_{i=1}^j y_i \right\}$$

is a continuity set of the product measure

$$\Pi_\theta(dx_1, \dots, dx_j, dy_1, \dots, dy_j) \stackrel{\text{def}}{=} \mu(dx_1) \times \dots \times \mu(dx_j) \times \pi_{1,\theta}(dy_1) \times \dots \times \pi_{j,\theta}(dy_j).$$

Indeed, since μ is a continuous distribution, we have

$$\forall j \geq 1; y_1, \dots, y_j \in \mathbb{R}: \mu^j \left(\left\{ r \sum_{i=1}^j x_i = \sum_{i=1}^j y_i \right\} \right) = 0.$$

Hence, applying Fubini's Theorem yields

$$\forall j \geq 1: \Pi_\theta \left(\left\{ r \cdot \sum_{i=1}^j x_i = \sum_{i=1}^j y_i \right\} \right) = \int \dots \int \mu^j \left\{ r \sum_{i=1}^j x_i = \sum_{i=1}^j y_i \right\} \pi_{1,\theta}(dy_1) \dots \pi_{j,\theta}(dy_j) = 0.$$

Note that, this is equivalent to

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

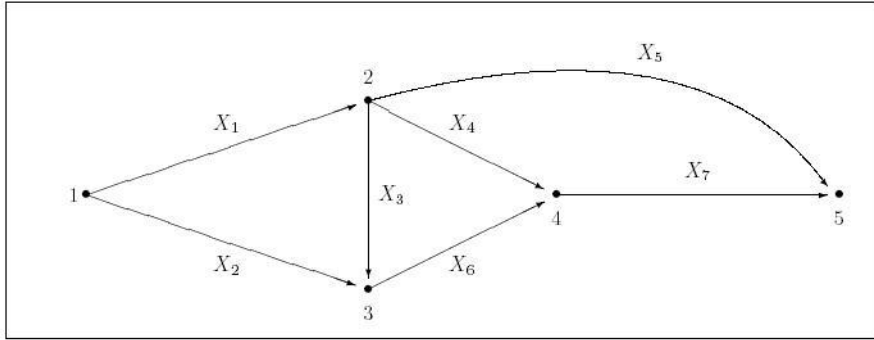


Figure 1: SAN example with source node 1 and sink node 5.

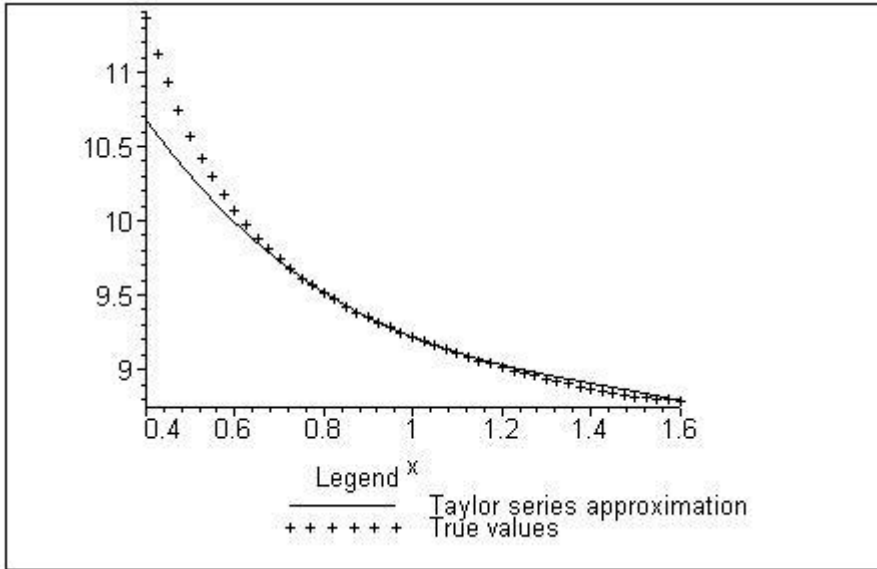


Figure 2: The Taylor polynomial of order 3 compared to the true value.

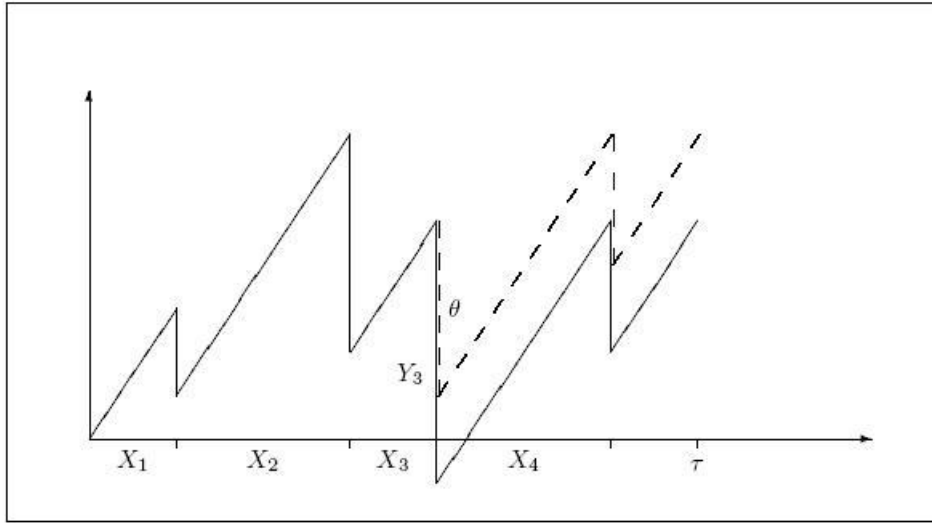


Figure 3: An occurrence of the event $\mathfrak{R}_\tau \setminus \mathfrak{R}_\tau^3$ and $N_\tau = 4$. The dashed line represents a version of the process where the value of the 3^{rd} claim is reduced.