Applied Mathematical Sciences, Vol. 3, 2009, no. 20, 979 - 1004

Moment-Based Approximations for the Law of Functionals of Dirichlet Processes

Ilenia Epifani

Dipartimento di Matematica, Politecnico di Milano
P.zza Leonardo da Vinci, 32, 20133 Milano, Italy
ilenia.epifani@polimi.it

Alessandra Guglielmi

Dipartimento di Matematica, Politecnico di Milano
P.zza Leonardo da Vinci, 32, 20133 Milano, Italy
alessandra.guglielmi@polimi.it

Eugenio Melilli

Dipartimento di Scienze delle Decisioni, Università L. Bocconi
viale Isonzo 25, 20135 Milano, Italy
eugenio.melilli@unibocconi.it

Abstract

The paper deals with the approximation of the law of a random functional of a Dirichlet process using a finite number of its moments. In particular, three classes of approximation procedures – expansions in series of orthonormal polynomials, the maximum entropy method and mixtures of known distributions – are discussed. A comparison of the different approximation procedures is performed by a few examples. Moreover, some new results on the support and the existence of the moment generating function of the Dirichlet functional variance are given.

Mathematics Subject Classification: 62F15, 60E10, 62E17

Keywords: Approximations by moments, Distributions of functionals of Dirichlet processes, Moments of a distribution

1 Introduction

In this paper we discuss the problem of approximating the laws of functionals of a Dirichlet process P_α with parameter α, using their moment sequences, here, *moment-based approximation problem.* In particular, we address our analysis to the approximation of the law of the variance $V_\alpha = \int x^2 P_\alpha(dx) - (\int x\, P_\alpha(dx))^2$ but, of course, it could be carried over to other functionals for which moments are known. By the conjugacy property of the Dirichlet process, any distributional results regarding its functionals will concern both prior and posterior analysis.

The (exact or approximated) distribution of the variance of a Dirichlet process is a useful tool in nonparametric Bayesian inference. For example, we make inference about the unknown variance of some characteristic in a quality control setting when the variability of the manufacturing process must be kept under control. If we think that a parametric model is not appropriate for the data, then the "unknown" distribution can be assumed to be a Dirichlet random probability P_α. In this case, we could choose "$P(V_\alpha \geq k|data)$ less than some desirably low probability" as a checking rule of the homogeneity of the manufacturing process.

According to the Bayesian nonparametric viewpoint, the choice of a Dirichlet process is not restrictive at all. First of all, any random probability measure can be approximated in distribution, with any precision, by a finite mixture P_m of Dirichlet processes; see [11] and [38]. Therefore, since the law of the variance of P_m is the mixture of the distributions of the variances of the components in P_m, results about the law of the variance of a Dirichlet process can be considered as being quite general. Moreover, we should consider the analysis of V_α as the first step towards further generalizations, since many new popular random probability measures extend Dirichlet processes. See [20], and [34] for recent reviews. In particular, our results can apply to mixtures of parametric families with a Dirichlet process as mixing measure, first introduced by [30]; see the last section of this paper.

However, results about distributions of functionals of a Dirichlet process P_α are often stated in terms of integral transforms (such as Laplace, Stieltjies or Zolotarev) or as solutions of stochastic equations involving the base measure α of the process. See [8], [36], [37] and [24] as far as linear functionals of P_α are concerned; on the other hand, very few results are known about the variance functional ([7] and [16]). In any case, explicit expressions for the distribution functions or the densities of these functionals can be obtained in very few situations. At the same time, integral transforms or solutions of stochastic equations easily yield the moments of the functionals as functions of the moments of the parameter α. Therefore, one can use the moments to obtain information on these functionals to make inferences in a Bayesian nonparametric perspective. How do we use a finite string of moments to approximate the underlying distribution or some of its features? This is a classical problem, widely discussed in the literature. Of course, preliminary to the approximation problem, there is an identification issue; *i.e.*, finding out conditions under which the moments of a distribution exist and uniquely identify it. This latter problem has found a complete answer in classical works; very good reviews are, for instance, [42], [1] and [12]. The approximation problem, on the contrary, suffers an intrinsic difficulty, since, even for distributions identified by moments, finite sequences of moments give relatively poor information on the distribution itself. In other words, the class of all distributions having the same first k moments, even when it reduces to a singleton as k goes to $+\infty$, is somewhat wide for each finite k; [45] argues about how informative is the reduced moment problem. The situation is quite different if we are interested only in tail probabilities. In fact classical results ([1] and [27]) show that approximations based on moments are precise enough when used to compute the high quantiles of the distribution of interest. A sharp enough bound is available in this case: if H and M are distribution functions sharing the first $2k$ moments, then $|H(t) - M(t)|$ is bounded by the reciprocal of a polynomial of order $2k$ in t, whose coefficients are functions of the moments. This kind of approximation can be very useful in Bayesian hypotheses testing, where upper tail probabilities are needed, as in the manufacturing process homogeneity problem mentioned above. Since the literature on the problem is extremely huge and "ancient" and there exists a large number of approximation methods, we will only focus on those we believe are tailor-made for the functionals of interest. On the other hand, any approximation method using moments is numerically unstable; see, for instance, [23], [19] and [44] for a description of numerical issues connected to moment approximation problems. Of course, in a different perspective, one can use simulation methods to approximate the distribution of functionals of P_α, when draws from α are available (see, for instance, [21]). But this is beyond the scope of this paper.

In approximating a distribution using some of its moments, one aims at using as many information on the underlying distribution as he can. For this reason, as far as the identification issue is concerned, we give a sufficient condition on α characterizing the existence of the moment generating function of V_α. Second, since knowledge of the support of the distribution is useful to drive towards the choice of the "best" approximation, then we give some new contributions on the support of V_α. Further information on the distribution of V_α, *i.e.* the boundedness or the smoothness of its density, can in principle be obtained from the same moments, but such conditions are generally difficult to verify, since an infinite number of inequalities must be checked.

The set-up of the paper is the following. In Section 2 some background on moments is reviewed. Section 3 recalls general results on the Dirichlet process and its functionals and yields new contributions on the support of V_α, on the existence of its moment generating function and on its asymptotic normality. Sections 4, 5 and 6 discuss three different classes of approximation procedures, by means of expansions in series of orthonormal polynomials (Section 4), maximum entropy (Section 5) and mixtures of known distributions (Section 6). Section 7 shows some illustrative examples, through which a comparison of the approximation methods is carried on. Some comments are given in Section 8.

2 Background on moments

In this section we briefly recall notation and classical results related to the moment problem, *i.e.* finding a probability whose moments are equal to a given sequence of real numbers $(\mu_k)_{k\geq 1}$. If $(\mu_k)_{k\geq 1}$ is a sequence of moments, there exists at least one solution of the moment problem; if this solution is unique, the moment problem is *determinate*. A probability P is said *uniquely determined by its moments* if all its moments exist and the moment problem is determinate. Sufficient (but not generally necessary) conditions for the uniqueness of P are the existence of the moment generating function of P on some interval containing the origin or the classical Carleman's criterion on $(\mu_k)_{k\geq 1}$. Moment problems are usually classified into 3 categories according to the support of their solutions; the Hausdorff moment problem concerns distributions with bounded support, whilst Stieltjies and Hamburger moment problems are those corresponding to bounded from below and unbounded support, respectively. An important result about the moment-based approximation problem is the following:

Proposition 2.1 Suppose that a probability measure P on $\mathbb{R}$ is determined by its moments $\mu_1, \mu_2, \ldots$. Let $(P_n)_{n=1}^{+\infty}$ be a sequence of probability measures on $\mathbb{R}$ having all moments $\nu_k^n = \int x^k P_n(dx)$ such that

$$\lim_{n\to+\infty} \nu_k^n = \mu_k, \quad k = 1, 2, \ldots .$$

Then $P_n \xrightarrow{w} P$.

For the proof see [3]. Hence, when approximating a probability measure P by means of a distribution P_n whose moments coincide or converge for $n \to +\infty$ to those of P, then the approximating P_n converges weakly to the target P. When the target distribution P is continuous (and this is the case for the law of the variance or of the mean of a Dirichlet process), the convergence of the associated distribution functions is uniform.

In principle, at least in the Markov problem, knowledge of all moments yields information on the associated density. Let Δ be the difference operator $\Delta(\mu_j) := \mu_{j+1} - \mu_j$, Δ^k its k-th power with $\Delta^0 \mu_j := \mu_j$ and

$$\mu_{n,j} := (-1)^{n-j} \binom{n}{j} \Delta^{n-j}(\mu_j), \quad n = 0, 1, \dots, j = 0, 1, \dots, n. \tag{1}$$

A necessary and sufficient condition to guarantee that the moment problem is determinate and its unique solution P is absolutely continuous with an almost everywhere bounded density f is

$$0 \leq \mu_{n,j} \leq \frac{d}{n+1} \quad \text{for all } n, j \text{ and some positive real number } d; \tag{2}$$

see Theorem 2 in [13], where an analogous condition ensuring boundedness of f in L^p-norm is also stated (Theorem 3 and 4).

3 Some new results on V_α

Here we first review notation and results about Dirichlet process and its functionals. Let $\mathbb{P}$ be the space of all probability measures on the Borel σ-field $\mathcal{B}(\mathbb{R})$ on $\mathbb{R}$, endowed with the Prohorov metric; in this way, $\mathbb{P}$ is a Polish space, whose Borel σ-field we denote by $\mathcal{P}$. Given a probability space $(\Omega, \mathcal{F}, \mathsf{P})$, a *random probability measure* P on $\mathbb{R}$ is any measurable function from $(\Omega, \mathcal{F})$ into $(\mathbb{P}, \mathcal{P})$; let $P(A; \omega)$ denote the value of the probability measure $P(\omega)$ at A in $\mathcal{B}(\mathbb{R})$. If α is a finite measure on $\mathcal{B}(\mathbb{R})$, with total mass $\alpha(\mathbb{R}) = c > 0$, then a *Dirichlet process* with parameter α is a random probability measure P_α such that, for any finite measurable partition $(A_1, \dots, A_n)$ of $\mathbb{R}$ with $n \geq 2$ and $\alpha(A_j) > 0$, $j = 1, \dots, n$, the distribution of $(P_\alpha(A_1), \dots, P_\alpha(A_n))$ is Dirichlet with parameter $(\alpha(A_1), \dots, \alpha(A_n))$. If $\alpha(A_j) = 0$ for some j, the j-th coordinate of the random vector $(P_\alpha(A_1), \dots, P_\alpha(A_n))$ is P-a.s. equal to 0. Denote by α_0 the probability measure defined by $\alpha_0(A) = \alpha(A)/c$, A in $\mathcal{B}(\mathbb{R})$. Given a Dirichlet process P_α, we can consider some of its functionals, such as the mean $\Gamma_\alpha(\omega) := \int x P_\alpha(dx; \omega)$ and the variance $V_\alpha(\omega) := \int x^2 P_\alpha(dx; \omega) - \left(\int x P_\alpha(dx; \omega) \right)^2$. First of all, Γ_α and V_α are P-a.s. finite r.v. on $(\Omega, \mathcal{F})$ if and only if the base measure α is such that

$$\int_{\mathbb{R}} \log(1 + |x|) \alpha(dx) < +\infty.$$

Moreover, if α is non-degenerate, Γ_α and V_α are absolutely continuous r.v.; in what follows we shall denote by f_{V_α} and F_{V_α} a density and the distribution function of V_α, respectively.

The laws of Γ_α and V_α are strictly related themselves by means of the distributional equation provided in [16] that connects V_α to the squared difference of two independent copies of the Dirichlet mean Γ_α. Indeed, if $c \le 1/2$, then $V_\alpha \overset{d}{=} B_1 U_\alpha$, where B_1, U_α are independent r.v's with $B_1 \sim beta(1, 1/2-c)$ and , $U_\alpha = (X-Y)^2/4$ with X, Y independent copies of Γ_α. Moreover, if $c > 1/2$, then $B_2 V_\alpha \overset{d}{=} U_\alpha$, with B_2, V_α independent r.v's and $B_2 \sim beta(1, c-1/2)$. As far as the moments of Γ_α and V_α are concerned, if the $2k$-th moment of α_0 is finite, then Γ_α and V_α have finite $2k$-th and k-th moment, respectively; in particular, Γ_α and V_α admit all moments if α_0 does. The moments of Γ_α and V_α are

$$\mathsf{E}(\Gamma_\alpha^k) = \frac{c}{\Gamma(c+k)} \sum_{i=0}^{k-1} \frac{(k-1)!}{i!} \Gamma(c+i) \xi_{k-i} \mathsf{E}(\Gamma_\alpha^i), \quad k = 1, 2, \dots, \tag{3}$$

where $\xi_k := \int_{\mathbb{R}} x^k \alpha_0(dx)$ (see, for example, [14], and

$$\begin{aligned} \mathsf{E}(V_\alpha^n) &= \frac{\Gamma(a+n)\Gamma(\frac{1}{2})}{4^n \Gamma(n+\frac{1}{2})\Gamma(a)} \sum_{j=0}^{2n} \binom{2n}{j} (-1)^j \mathsf{E}(\Gamma_\alpha^j) \mathsf{E}(\Gamma_\alpha^{2n-j}) \\ &= \frac{\Gamma(a+n)\Gamma(\frac{1}{2})}{\Gamma(n+\frac{1}{2})\Gamma(a)} \mathsf{E}(U_\alpha^n) \end{aligned} \tag{4}$$

(see (11) in [7]).

We can verify that V_α is uniquely determined by its moments if $\alpha \circ (x^2)^{-1}$ is. In fact, we prove the following:

Theorem 3.1 *Let Z be a r.v. with α_0 distribution. The moment generating function M_{V_α} of V_α exists if, and only if, the moment generating function M_{Z^2} of Z^2 exists.*

Proof. We use the above-mentioned distributional relationship between V_α and U_α to prove the statement. First, by the distributional equation, $\limsup_{k\to\infty} \left[|\mathsf{E}(V_\alpha^k)|/k!\right]^{\frac{1}{k}}$ is finite if, and only if, $\limsup_{k\to\infty} \left[|\mathsf{E}(U_\alpha^k)|/k!\right]^{\frac{1}{k}}$ is finite. Hence M_{V_α} exists if, and only if, M_{U_α} exists. Second, let us prove that U_α admits moment generating function if, and only if, Γ_α^2 does. Since Γ_α and $\Gamma_\alpha + b$ admit exactly the same numbers of moments, there is no loss of generality in assuming that Γ_α has median 0. Thus it is easy to verify that $P(U_\alpha > t) \ge P(\Gamma_\alpha^2 > 4t)$ for all $t > 0$ and

$$\mathsf{E}(U_\alpha^k) = k \int_0^\infty t^{k-1} P(U_\alpha > t)\, dt \ge \frac{k}{2} \int_0^\infty t^{k-1} P(\Gamma_\alpha^2 > 4t)\, dt = \frac{1}{2} \mathsf{E}\left(\left(\frac{\Gamma_\alpha}{2} \right)^{2k} \right).$$

Hence

$$\limsup_{k\to\infty}\left[\frac{|\mathsf{E}(U_\alpha^k)|}{k!}\right]^{\frac{1}{k}} \geq \limsup_{k\to\infty}\left[\frac{|\mathsf{E}\big((\frac{\Gamma_\alpha}{2})^{2k}\big)|}{k!}\right]^{\frac{1}{k}},$$

so that the existence of M_{U_α} implies the existence of $M_{\Gamma_\alpha^2}$. Conversely, it is sufficient to observe that $\mathsf{E}(\mathrm{e}^{tU_\alpha}) \leq \mathsf{E}(\mathrm{e}^{\frac{t}{2}(X^2+Y^2)}) = \mathsf{E}^2(\mathrm{e}^{\frac{t}{2}\Gamma_\alpha^2})$ to derive the existence of M_{U_α} from that of $M_{\Gamma_\alpha^2}$. As a third step, notice that $\int x^2 P_\alpha(dx) = V_\alpha + \Gamma_\alpha^2$, and $\int x^2 P_\alpha(dx)$ has the same distribution of a Dirichlet mean Γ_β with parameter $\beta = \alpha \circ (x^2)^{-1}$ and M_{Γ_β} exists if, and only if, $M_\beta = M_{Z^2}$ exists as proved in Lemma 2 in [16]. Hence, the existence of M_{Z^2} imply the existence of both M_{V_α} and $M_{\Gamma_\alpha^2}$ and the "if part" of Theorem 3.1 is proved. Conversely, the existence of M_{V_α} implies the existence of $M_{\Gamma_\alpha^2}$ and hence that of $\int x^2 P_\alpha(dx)$, which is equal to the sum $V_\alpha + \Gamma_\alpha^2$, and therefore that of M_{Z^2}. □

We now determine the support of V_α. We shall denote the support of a random element X with law q equivalently by means of $S(X)$ or $S(q)$, and the closure of the convex hull of the support of q by $\overline{co(S(q))}$.

Lemma 3.2 *If (x, y) belongs to the support $S(X, Y)$ of a random vector (X, Y), then $(x^2, y) \in S(X^2, Y)$.*

Proof. Let $(x, y) \in S(X, Y)$, $\epsilon_2 > 0$, and $0 < \epsilon_1 < x^2$ s.t. $\eta := \min\{x - \sqrt{x^2-\epsilon_1}, \sqrt{x^2+\epsilon_1} - x\} > 0$ (if $x = 0$, the proof is trivial). Thus

$$\begin{aligned}
&\{|X^2 - x^2| < \epsilon_1, |Y - y| < \epsilon_2\} = \\
&\quad \{-\sqrt{x^2+\epsilon_1} < X < -\sqrt{x^2-\epsilon_1}, |Y-y| < \epsilon_2\} \cup \\
&\quad \cup \{\sqrt{x^2-\epsilon_1} < X < \sqrt{x^2+\epsilon_1}\}, |Y-y| < \epsilon_2\} \\
&\quad \supseteq \{\sqrt{x^2-\epsilon_1} < X < \sqrt{x^2+\epsilon_1}, |Y-y| < \epsilon_2\} \supseteq \{|X-x| < \eta, |Y-y| < \epsilon_2\},
\end{aligned}$$

so that $\mathsf{P}(|X^2 - x^2| < \epsilon_1, |Y - y| < \epsilon_2) \geq \mathsf{P}(|X - x| < \eta, |Y - y| < \epsilon_2) > 0$, since $(x, y) \in S(X, Y)$. □

Lemma 3.3 *Let B, W be independent r.v.'s and suppose $S(B) = [0, 1]$. Then $S(W) \subseteq S(BW)$.*

Proof. Let $w \in S(W)$. Then, for all $\epsilon > 0$ and $0 < \eta < \epsilon$ we have

$$\{w - \epsilon < BW < w + \epsilon\} \supseteq \{w - \eta < W < w + \eta, \frac{w-\epsilon}{w-\eta} < B < 1\}$$

and then

$$\begin{aligned}
\mathsf{P}(w - \epsilon < BW < w + \epsilon) &\geq \mathsf{P}(w - \eta < W < w + \eta, \frac{w-\epsilon}{w-\eta} < B \leq 1) = \\
&= \mathsf{P}(w - \eta < W < w + \eta)\mathsf{P}(\frac{w-\epsilon}{w-\eta} < B \leq 1) > 0. \ \square
\end{aligned}$$

Lemma 3.4 *Suppose the parameter α of a Dirichlet process P_α has bounded support $S(\alpha)$. If u, v belong to $S(\alpha)$, and $0 < \theta < 1$, then $\theta(1-\theta)(v-u)^2$ belongs to $S(V_\alpha)$.*

Proof. Let u, v be in $S(\alpha)$, $0 < \theta < 1$ and μ the discrete measure defined by $\mu(dz) = \theta\delta_u(dz) + (1-\theta)\delta_v(dz)$. Thus μ belongs to the support of P_α, $S(P_\alpha)$, *i.e.*

$$\mathcal{D}_\alpha \left\{ p : \left| \int f_i(x)\, p(dx) - \int f_i(x)\, \mu(dx) \right| < \epsilon, i = 1, \ldots, k \right\} > 0, \tag{5}$$

where ϵ is positive, $k = 1, 2, \ldots$ and f_i's are bounded, continuous functions on $\overline{co(S(\alpha))}$. In particular, by applying Equation (5) to $f_1(x) = x$ and $f_2(x) = x^2$ for any $\epsilon > 0$, one has

$$\mathcal{D}_\alpha \left\{ p : \left| \int x\, p(dx) - (u\theta + v(1-\theta)) \right| < \epsilon, \right.$$
$$\left. \left| \int x^2\, p(dx) - (u^2\theta + v^2(1-\theta)) \right| < \epsilon, \right\} > 0. \tag{6}$$

It follows from (6) and Lemma 3.2 that

$$\mathcal{D}_\alpha \left\{ p : \left| (\int x\, p(dx))^2 - (u\theta + v(1-\theta))^2 \right| < \epsilon, \right.$$
$$\left. \left| \int x^2\, p(dx) - (u^2\theta + v^2(1-\theta)) \right| < \epsilon \right\} > 0 \tag{7}$$

$\forall \epsilon > 0$. Finally, Equation (7) and the definition of V_α as $V_\alpha = \int x^2\, P_\alpha(dx) - (\int x\, P_\alpha(dx))^2$ yield $u^2\theta + v^2(1-\theta) - (u\theta + v(1-\theta))^2 = \theta(1-\theta)(v-u)^2 \in S(V_\alpha)$. $\square$

Theorem 3.5 *If $\overline{co(S(\alpha))}$ is a bounded interval $[\sigma, \tau]$, then*

$$S(V_\alpha) = \left[0, \frac{(\tau - \sigma)^2}{4} \right]. \tag{8}$$

Otherwise, $S(V_\alpha) = [0, \infty)$.

Proof. First, we suppose α has a bounded support. Secondly we consider the case when $\overline{co(S(\alpha))} = [0, \infty)$. This latter is equivalent to assume more generally that the support of α is unbounded on one side only, since the law of V_α is α-shift and α-reflection invariant. Indeed, if the measure α^- is defined via $\alpha^-(B) := \alpha(-B)$ and $\alpha_c(B) := \alpha(B - c)$, for any real constant c and $B \in \mathcal{B}(\mathbb{R})$, then $V_{\alpha^-}, V_{\alpha_c}$ and V_α have the same distribution; Sethuraman's representation of P_α ([41]) can be useful in proving it. Finally we tackle the case $\overline{co(S(\alpha))} = \mathbb{R}$.

Step 1. If $\overline{co(S(\alpha))}$ is $[\sigma, \tau]$, then by Lemma 3.4 we obtain $\left[0, \frac{(\tau-\sigma)^2}{4}\right] \subseteq S(V_\alpha)$. Conversely, to prove that $\left[0, (\tau-\sigma)^2/4\right] \supseteq S(V_\alpha)$, we use the distributional equation that connects U_α to V_α. First of all, the internality propriety of the mean Γ_α ($\mathbb{P}(\sigma \leq \Gamma_\alpha \leq \tau) = 1$) and (6) yield that $S(\Gamma_\alpha) = [\sigma, \tau]$ and therefore $S(U_\alpha) = [0, (\tau-\sigma)^2/4]$. Moreover, if $a \leq 1/2$, then

$$\mathsf{P}\left(V_\alpha > \frac{(\tau-\sigma)^2}{4}\right) = \mathsf{P}\left(B_1 U_\alpha > \frac{(\tau-\sigma)^2}{4}\right) \leq \mathsf{P}\left(U_\alpha > \frac{(\tau-\sigma)^2}{4}\right) = 0,$$

i.e. $S(V_\alpha) = [0, (\tau-\sigma)^2/4]$. On the other hand, if $a > 1/2$, then $S(B_2 V_\alpha) = S(U_\alpha)$. So we can apply Lemma 3.3 above to obtain $S(V_\alpha) \subseteq S(B_2 V_\alpha)$. Last inclusion completes the proof of (8).

Step 2. Suppose now $\overline{co(S(\alpha))} = [0, \infty)$. Here we need a "truncation procedure" first appeared in [8]. Let $A(x) := \alpha((\infty, x])$ be the distribution function corresponding to α, let A_m be the distribution function defined as follows

$$A_m(x) = \begin{cases} A(x) & x < m \\ c & x \geq m \end{cases} \tag{9}$$

and denote by α_m the finite measure corresponding to A_m. Of course $\overline{co(S(\alpha_m))} = [0, m]$ and, as observed in [16], the sequence of the random variances $(V_m)_{m\geq 1}$, with $V_m := V_{\alpha_m}$, converges weakly to variance V_α. Moreover, V_α is an absolutely continuous r.v. and $S(V_m) = [0, m^2/4]$ by (8). We will prove by contradiction that there does not exist any finite $K > 0$ such that $S(V_\alpha) = [0, K]$. Suppose that such a K exists, *i.e.* $\mathsf{P}(V_\alpha > K) = 0$. By the weak convergence of $(V_m)_m$ to the absolutely continuous r.v. V_α , it follows that

$$\forall \epsilon > 0 \ \exists M_1 \text{ such that } \ 0 < 1 - \mathsf{P}(V_m \leq K) < \epsilon, \ \ \forall m > M_1.$$

On the other hand, since $S(V_m) = [0, m^2/4]$, then there exists M_2 such that K is an interior point of the support of V_m for all $m > M_2$ and hence

$$\mathsf{P}(V_m \leq K) < 1 - 2\epsilon, \qquad \forall m > M_2 \text{ and for some } \epsilon > 0.$$

So, we have for some $\epsilon > 0$ and for all $m > \max\{M_1, M_2\}$

$$1 - \epsilon < \mathsf{P}(V_m \leq K) < 1 - 2\epsilon,$$

which is clearly impossible.

Step 3. Finally, let $\overline{co(S(\alpha))}$ be $\mathbb{R}$ and consider the measure α_m defined via the truncation procedure (9) and the corresponding Dirichlet variance V_m. In this case, $\overline{co(S(\alpha_m))} = (-\infty, m]$, so that $S(V_m) = [0, \infty)$ (see Step 2) and, as before, V_m weakly converges to V_α. Thus, arguing as in the Step 2, we obtain the result. □

We conclude this section with a proposition stating the asymptotic normality of V_α. A similar result concerning linear functionals of P_α has been given by [22].

Proposition 3.6 Let α_0 be a probability measure on $(\mathbb{R}, \mathcal{B}(\mathbb{R}))$ such that $\xi_4 < +\infty$. Then, for $\alpha = c\alpha_0$,

$$\sqrt{c}(V_\alpha - V_0) \xrightarrow{d} \mathcal{N}(0, \sigma_0^2), \quad c \to +\infty,$$

where $\sigma_0^2 = \xi_4 - 4\xi_3\xi_1 + 4\xi_2\xi_1^2 - \xi_2^2 - 4\xi_1^4 > 0$.

Proof. Let $\Lambda_\alpha := \int x^2 P_\alpha(dx)$ be the random second moment of P_α and G_c a gamma$(c, 1)$ variable independent of Γ_α and Λ_α. Thus the characteristic function of $(G_c\Gamma_\alpha, G_c\Lambda_\alpha, G_c)$ is

$$\mathsf{E}\Big(\exp(i(uG_c\Gamma_\alpha + vG_c\Lambda_\alpha + zG_c))\Big) = \exp\Big(-ic\int \log(1 + ut + vt^2 + z)\alpha_0(dt)\Big),$$

from which it follows that $(G_c\Gamma_\alpha, G_c\Lambda_\alpha, G_c)$ is infinitely divisible. Then, an application of the central limit theorem and the delta method gives

$$\sqrt{c}\,(\Gamma_\alpha - \xi_1, \Lambda_\alpha - \xi_2) \xrightarrow{d} \mathcal{N}_2(0, \Sigma), \quad c \to +\infty, \tag{10}$$

where $\mathcal{N}_2(0, \Sigma)$ denotes the bivariate normal distribution with 0-mean vector and covariance matrix $\Sigma = \begin{pmatrix} \xi_2 - \xi_1^2 & \xi_3 - \xi_1\xi_2 \\ \xi_3 - \xi_1\xi_2 & \xi_4 - \xi_2^2 \end{pmatrix}$. On the other hand, if $h(x, y) = y - x^2$, then its gradient, evaluated in (ξ_1, ξ_2), is $\nabla^T = (-2\xi_1, \xi_2)$, so that (10) and an application of the delta method gives

$$\sqrt{c}\,(h(\Gamma_\alpha, \Lambda_\alpha) - h(\xi_1, \xi_2)) \xrightarrow{d} h(W), \quad c \to +\infty,$$

where W is a $\mathcal{N}_2(0, \Sigma)$ random vector and $h(W)$ is distributed according to $\mathcal{N}(0, \nabla^T\Sigma\nabla)$. □

4 Moment-based approximations of $\mathcal{L}(V_\alpha)$: expansion in orthonormal polynomials

As proved in Section 3, the support of V_α can only be or a bounded interval, say $[0, \delta]$, or the interval $[0, +\infty)$. We consider the case of bounded support and assume, without loss of generality, $\delta = 1$. Of course, in this case all the moments of V_α exist and uniquely identify its distribution. For $a, b > -1$, let us consider the function $w_{a,b}(x) = x^b(1-x)^a$ and denote by $\nu_{a,b}$ the (finite) measure with density $w_{a,b}(x)$, w.r.t. Lebesgue measure on $(0,1)$. The distribution of V_α is absolutely continuous w.r.t. $\nu_{a,b}$ with density $g_{a,b}(x) = f_{V_\alpha}(x)/w_{a,b}(x)$.

We assume that there exist $a, b > -1$ for which the function $g_{a,b}$ belongs to $L^2(\nu_{a,b})$, *i.e.*

$$\int_{(0,1)} g_{a,b}(x)^2 \nu_{a,b}(dx) = \int_0^1 f_{V_\alpha}(x)^2 \frac{1}{w_{a,b}(x)} dx < +\infty. \tag{11}$$

Let $(J_n(x; a, b))_{n\geq 0}$ be the sequence of (normalized) Jacobi polynomial on $(0,1)$ with parameters a, b (see, for instance, [43] for their definition). As it is well known, $(J_n(x; a, b))_{n\geq 0}$ is an orthonormal complete system of polynomials w.r.t. the weight function $w_{a,b}$ on $(0,1)$. For instance, the first few elements of the sequence are:

$$J_0(x; a, b) = \sqrt{\frac{1}{B(b+1, a+1)}}, \; J_1(x; a, b) = \sqrt{\frac{a+b+2}{B(b+2, a+2)}}\Big(x - \frac{b+1}{a+b+2}\Big),$$

$$J_2(x; a, b) = \sqrt{\frac{(a+b+3)(a+b+4)}{2B(a+3, b+3)}}\Big(x^2 - \frac{2(b+2)}{a+b+4}x + \frac{(b+1)(b+2)}{(a+b+3)(a+b+4)}\Big).$$

Well known results on Hilbert spaces give the following formal representation of $g_{a,b}$:

$$g_{a,b}(x) \sim \sum_{k=0}^{+\infty} a_k J_k(x; a, b) \tag{12}$$

that means

$$\lim_{n\to+\infty} \int_{[0,1]} \left\{ g_{a,b}(x) - \sum_{k=0}^{n} a_k J_k(x; a, b) \right\}^2 \nu_{a,b}(dx) =$$

$$= \lim_{n\to+\infty} \int_0^1 \left\{ f_{V_\alpha}(x) - w_{a,b}(x) \sum_{k=0}^{n} a_k J_k(x; a, b) \right\}^2 \frac{1}{w_{a,b}(x)} dx = 0 \; ,$$

where, for each non-negative integer k,

$$a_k = \int_{(0,1)} g_{a,b}(x) J_k(x; a, b) \nu_{a,b}(dx) = \int_0^1 f_{V_\alpha}(x) J_k(x; a, b) dx = \mathsf{E}(J_k(V_\alpha; a, b)) \; .$$

Moreover, observe that if $\nu^*_{a,b}$ is the measure with density $w^{-1}_{a,b}$ w.r.t. the Lebesgue measure on $(0,1)$ and $a, b < 1$ so that $\int_0^1 w^{-1}_{a,b}(x) dx < +\infty$ (*i.e.* $\nu^*_{a,b}$

is a finite measure), then $f_{V_\alpha}(x)$ itself can be approximated, in $L^2(\nu^*_{a,b})$, by the finite sum $w_{a,b}(x)\sum_{k=0}^{n} a_k J_k(x;a,b)$. Of course, this does not imply in general the uniform or even pointwise convergence of the series $w_{a,b}(x)\sum_{k=0}^{+\infty} a_k J_k(x;a,b)$ to $f_{V_\alpha}(x)$; some additional conditions are needed for this. For instance, a sufficient (but not necessary) condition for uniform convergence of the series is given by the following proposition, whose proof can be found in [17].

Proposition 4.1 If f_{V_α} is continuous and has a piecewise continuous derivative in $(0,1)$, then the series $w_{a,b}(x)\sum_{k=0}^{+\infty} a_k J_k(x;a,b)$, $a,b>-1$, converges uniformly to $f_{V_\alpha}(x)$ in $[\varepsilon, 1-\varepsilon]$, for each positive ε.

Other sufficient conditions require the function f_{V_α} belonging to a Lipschitz class of some order; see [35].

Remark **4.2** It follows from (11) that, for any $\nu_{a,b}$-integrable function h,

$$\begin{aligned}\int_0^1 h(x) f_{V_\alpha}(x)dx &= \int_{(0,1)} h(x) g_{a,b}(x)\nu_{a,b}(dx)\\ &= \sum_{k=0}^{+\infty} a_k \int_{(0,1)} J_k(x;a,b)h(x)\nu_{a,b}(dx) = \sum_{k=0}^{+\infty} a_k \int_0^1 J_k(x;a,b)h(x)w_{a,b}(x)dx\ .\end{aligned}$$

In particular, for $h(x) = I_{(0,t)}(x)$, $t\in(0,1)$, if $J_k(x;a,b) = \sum_{j=0}^{k} b_{k,j}x^j$, we have

$$\begin{aligned}F_{V_\alpha}(t) &= \int_0^t f_{V_\alpha}(x)dx = \sum_{k=0}^{+\infty} a_k \int_0^t J_k(x;a,b)x^b(1-x)^a dx\\ &= \sum_{k=0}^{+\infty} a_k \int_0^t \sum_{j=0}^{k} b_{k,j}x^j x^b (1-x)^a dx = \sum_{k=0}^{+\infty} a_k \sum_{j=0}^{k} b_{k,j} B_t(b+j+1, a+1)\ ,\end{aligned}$$

$B_t(p,q)$ denoting the incomplete beta function with parameters p and q. Moreover, the convergence of the series is uniform in $(0,1)$, as shown in [40], Section 7, Theorem 32. Hence, the distribution function of V_α can be uniformly approximated by the finite sum

$$\sum_{k=0}^{n} a_k \sum_{j=0}^{k} b_{k,j} B_t(b+j+1, a+1)\ ,$$

whose coefficients depends only upon the moments of V_α.

Remark **4.3** We emphasize that approximations based on expansions in orthonormal polynomials are not necessarily densities or distribution functions themselves (*i.e.*, approximations of densities may assume negative values and approximations of distribution functions can be decreasing in some intervals). On the other hand, when the convergence of the expansion is uniform, such approximations are generally very good.

REMARK **4.4** Similar approximations of f_{V_α} and F_{V_α} can be built when the support of V_α is $[0, +\infty)$, using orthonormal Laguerre polynomials $L_n^a(x)$ (*e.g.*, see [40]) instead of the Jacobi ones, under the assumption that, for some $a > -1$, $\int_0^{+\infty} f_{V_\alpha}^2(x) \mathrm{e}^x x^{-a} dx < +\infty$.

5 Moment-based approximations of $\mathcal{L}(V_\alpha)$: maximum entropy method

As in the previous section, we introduce notation and definitions when the support of V_α is $[0, 1]$. The *Boltzmann-Shannon entropy* of f_{V_α} is defined as

$$H(f_{V_\alpha}) := -\int_0^1 f_{V_\alpha}(x) \log f_{V_\alpha}(x) dx = -\int_{(0,1)\cap\{x: f_{V_\alpha}(x)>0\}} f_{V_\alpha}(x) \log f_{V_\alpha}(x) dx,$$

and, according to notation in [9], $-H(f_{V_\alpha})$ is called *I-divergence* of the law of V_α (w.r.t. the Lebesgue measure on $[0, 1]$). Observe that $H(f) \leq 0$, for any density f, since $x \log x \geq x - 1$ for any $x \geq 0$. A function g_n is the *maximum entropy estimate* of f_{V_α}, based on its first n moments, if

$$\max_{q \in \mathcal{Q}_n} H(q) = H(g_n), \tag{13}$$

$$\text{where } \mathcal{Q}_n := \{q : [0, 1] \to \mathbb{R}^+ : \int_0^1 x^j q(x) dx = \mu_j, j = 0, 1, \ldots, n\}.$$

Since the support of V_α is bounded, Theorem 3.3 in [9] ensures that, if there exists a density $q(x)$ with first n moments equal to $\mu_1, \ldots, \mu_n$, such that the I-divergence $-H(q)$ is finite, then there exists a unique solution g_n of (13). Here we cannot check directly that $H(f_{V_\alpha}) > -\infty$ since the analytic expression of f_{V_α} is unknown in general. However, if $\exp(\sum_{i=0}^n \lambda_i x^i)$ belongs to $\mathcal{Q}_n$, then it is a density and its entropy is finite. Since $(\mu_j)_{j \geq 0}$ is an infinite sequence of moments of a density on $[0, 1]$, then the system of equations

$$\int_0^1 x^j \exp(\sum_{i=0}^n \lambda_i x^i) dx = \mu_j, \quad j = 0, 1, \ldots, n \tag{14}$$

has a unique solution $(\lambda_0^*, \lambda_1^*, \ldots, \lambda_n^*)$ in $\mathbb{R}^{n+1}$ (see [31], Theorem 1, or [4], Lemma 1). Summing up, the maximum entropy estimate of f_{V_α} exists, is unique and its expression is

$$g_n(x) = \exp\Big(\sum_{i=0}^n \lambda_i^* x^i\Big), \ x \in [0, 1]. \tag{15}$$

It can be proved that $H(g_n) \downarrow H(f_{V_\alpha})$ (finite or infinite) as $n \to +\infty$, *i.e.* g_n converges in entropy to f_{V_α} as n grows to infinity. On the other hand, if $H(f_{V_\alpha}) > -\infty$, then $\int_0^1 |g_n(x) - f_{V_\alpha}(x)| dx \to 0$ as $n \to +\infty$, *i.e.* g_n converges in L_1 to f_{V_α}; this is equivalent to convergence in total variation distance, yielding $\sup_x |G_n(x) - F_{V_\alpha}(x)| \to 0$ for $n \to +\infty$, where G_n represents the distribution function of the density g_n; see [9] and [5]. Convergence of g_n to the exact density in L_∞ or L_p, $p \geq 2$, as well as reasonable error bounds, holds under more "smoothness" assumptions on f_{V_α}, but they will not be considered here, since they cannot be granted in general.

When the support of V_α is $S(V_\alpha) = [0, +\infty)$, the existence of the maximum entropy estimate of V_α is not guaranteed, depending on the geometry of the set $\Lambda := \{(\lambda_1, \dots, \lambda_n) \in \mathbb{R}^n : \int_{S(V_\alpha)} \exp(\sum_{i=1}^n \lambda_i x^i) dx < +\infty\}$. In general, if Λ is open in $\mathbb{R}^n$, then the maximum entropy estimate of V_α exists and its expression is as in (15); see Corollaire 1 in [10]. However it is straightforward to verify that Λ is not open when $S(V_\alpha) = [0, +\infty)$, and other weaker conditions, involving bounds on moments, must be checked; see [25], Section 9.

6 Moment-based approximations of $\mathcal{L}(V_\alpha)$: mixtures of distributions

In this section we consider a few methods for approximating $\mathcal{L}(V_\alpha)$ by means of a (finite) mixture of known distributions. Whilst methods discussed in the previous sections aim at approximating the density of V_α, here a direct approximation of its distribution function is looked for. We begin with two classical approximation procedures valid when V_α has bounded support which, as usual, without loss of generality, we assume equal to $[0, 1]$. The approximating distributions will turn out to be mixtures of Dirac measures, *i.e.* discrete probability distributions.

The first one is based on properties of Bernstein polynomials and aims at approximate $\mathcal{L}(V_\alpha)$ through the distribution

$$P_n = \sum_{j=0}^{n} \binom{n}{j} (-1)^{n-j} \Delta^{n-j}(\mu_j) \delta_{j/n} = \sum_{j=0}^{n} \mu_{n,j} \delta_{j/n}, \tag{16}$$

where $\mu_{n,j}$ is defined in (1). Observe that P_n has fixed support points $\{j/n\}$ (*i.e.*, not depending on the moments of V_α) and associated masses $\{\mu_{n,j}\}$ determined through $\mu_1, \dots, \mu_n$. It is simple to prove that, for each $k \geq 1$, $\int_0^1 x^k P_n(dx) \to \mu_k$ as $n \to +\infty$ (see, for instance, [18]); by Proposition 2.1, the weak convergence of P_n to $\mathcal{L}(V_\alpha)$ follows and the uniform convergence of the associated distribution functions holds true too. The main drawback of this approximation scheme is its very low rate of convergence; [2] gives more details on the convergence of (P_n).

The second discrete approximation we briefly mention is based on well-known Chebyshev polynomials and, as before, is valid for bounded support distributions only. The procedure is very simple: for each $n \geq 1$, determine n points $\theta_1, \dots, \theta_n$ in $[0, 1]$ and n nonnegative weights $\pi_1, \dots, \pi_n$ with $\sum_j \pi_j = 1$ such that probability measure

$$P_n = \sum_{j=1}^{n} \pi_j \delta_{\theta_j} \tag{17}$$

has $\mu_1, \mu_2, \dots, \mu_{2n-1}$ as first $2n - 1$ moments. Since $\mu_1, \mu_2, \dots, \mu_{2n-1}$ is part of an infinite moment sequence, then a solution exists; moreover it is unique. The support points $\theta_1, \dots, \theta_n$ are exactly the roots of the polynomial (in t)

$$det \begin{pmatrix} 1 & \mu_1 & \mu_2 & \cdots & \mu_n & 1 \\ \mu_1 & \mu_2 & \mu_3 & \cdots & \mu_{n+1} & t \\ \mu_2 & \mu_3 & \mu_4 & \cdots & \mu_{n+2} & t^2 \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ \mu_n & \mu_{n+1} & \mu_{n+2} & \cdots & \mu_{2n} & t^n \end{pmatrix} = 0 \ , \tag{18}$$

while the weights $\pi_1, \dots, \pi_n$ are obtained as the solution of the simple linear system

$$\begin{pmatrix} 1 & 1 & \dots & 1 \\ \theta_1 & \theta_2 & \dots & \theta_n \\ \vdots & \vdots & \vdots & \vdots \\ \theta_1^{n-1} & \theta_2^{n-1} & \dots & \theta_n^{n-1} \end{pmatrix} \begin{pmatrix} \pi_1 \\ \pi_2 \\ \vdots \\ \pi_p \end{pmatrix} = \begin{pmatrix} 1 \\ \mu_1 \\ \vdots \\ \mu_{n-1} \end{pmatrix} , \tag{19}$$

whose coefficients matrix is a Vandermonde (hence nonsingular) matrix. A detailed discussion of this method can be found in [6]. Unfortunately, the relation between the moments $\mu_1, \mu_2, \dots, \mu_{2n-1}$ and the solution $\theta_1, \dots, \theta_n, \pi_1, \dots, \pi_n$ is extremely unstable, so the method is severely ill-conditioned.

According to the next technique, essentially due to Lindsay ([26], [28] and [29]), the mixture components are parametric distributions having specific features, making the approximation procedure particularly easy to implement. Let us consider a parametric family of distributions $\{F_\theta, \theta \in \Theta\}$ on $(0, +\infty)$ and, for each positive integer p, denote by Q_p a probability measure on Θ with p-point support, *i.e.* $Q_p = \sum_{j=1}^{p} \pi_j \delta_{\theta_j}$. The aim is to choose the weights $\pi_1, \dots, \pi_p$ and the support points $\theta_1, \dots, \theta_p$ so that the (finite) mixture

$$F_{Q_p} = \int_\Theta F_\theta Q_p(d\theta) = \sum_{j=1}^{p} \pi_j F_{\theta_j} \tag{20}$$

has the same first $2p-1$ moments as V_α. This problem, requiring the solution of a system of $2p-1$ equations on the $2p-1$ (free) unknown $\pi_1, \dots, \pi_{p-1}, \theta_1, \dots, \theta_p$, is generally very complex, both theoretically since existence and uniqueness of the solution are by no means guaranteed, and computationally since the equations may be highly non-linear). Lindsay and coauthors suggest considering particular parametric families $\{F_\theta^\lambda, \theta \in \Theta, \lambda \in \Lambda\}$, having an additional non-negative parameter λ, in order to make simpler the solution of the system. The proposed classes of distributions arise as generalizations of the quadratic variance exponential families, introduced by [32], [33], and are characterized by the existence, for each $k = 1, 2, \dots$, of a polynomial $\xi_k^\lambda(x) = \sum_{i=0}^k a_{k,i}^\lambda x^i$ of degree k, possibly depending on λ, for which

$$\int_0^{+\infty} \xi_k^\lambda(x) dF_\theta^\lambda(x) = \theta^k, \qquad \theta \in \Theta. \tag{21}$$

By integrating both sides of (21) with respect to Q_p one obtains

$$\int_0^{+\infty} \xi_k^\lambda(x) dF_{Q_p}^\lambda(x) = \int_\Theta \theta^k Q_p(d\theta),$$

which yields, by imposing equality of the first $2p$ moments of F_{Q_p} and F_{V_α} (now the free parameters are $2p$, *i.e.* $\theta_1, \dots, \theta_p, \pi_1, \dots, \pi_{p-1}, \lambda$):

$$\sum_{i=0}^k a_{k,i}^\lambda \mu_i = \sum_{j=1}^p \pi_j \theta_j^k, \qquad k = 1, 2, \dots, 2p. \tag{22}$$

Observe that systems (19) and (22) formally differ only by the constant terms (the moment vector in the former and a linear function of it in the latter). Hence, the original (and generally complex) approximation problem has been transformed into the following: finding a distribution Q_p on Θ whose support has exactly p points and whose first $2p$ moments are given. In other words, the moment constraints on the mixture F_{Q_p} have been carried over to the mixing measure Q_p, greatly simplifying the problem. Indeed the existence of a solution Q_p can be proved under mild assumptions and the solution itself can be (numerically) determined. As a first step, one needs to choose λ such that a solution of (22) does exist; that is, a value of λ has to be determined for which the $2p$ quantities in the left hand-sides of Equations (22) are the first $2p$ moments of a distribution on Θ having p points of support. As proved in [29] (Theorem 3.1 and Proposition 3.2) using moment matrices theory, such a λ exists and it is unique, under weak conditions on the family $\{F_\theta^\lambda, \theta \in \Theta, \lambda \in \Lambda\}$. Once the value of λ guaranteeing the existence of the solution has been determined, (22) can be numerically solved. In particular, choosing *gamma*$(1/\lambda, 1/(\lambda\theta))$ as F_θ^λ (so that the coefficient of variation is $\sqrt{\lambda}$), the approximating F_{Q_p} turns out to be a mixture of gamma distributions, well tailored to approximate a positive support distribution as the variance law F_{V_α}. If we assume

$$\xi_k^\lambda(x) = \sum_{i=0}^{k} a_{k,i}^\lambda x^i = \frac{x^k}{(1+\lambda)(1+2\lambda)\dots(1+[k-1]\lambda)} \ ,$$

(22) yields

$$\nu_k^\lambda := \frac{\mu_k}{(1+\lambda)(1+2\lambda)\cdots(1+(k-1)\lambda)} = \sum_{j=1}^{p} \pi_j \theta_j^k, \quad k = 1, 2, \dots, 2p. \quad (23)$$

As mentioned above, the aim is to determine a $\lambda > 0$ so that (23) has solution, that is a $\lambda > 0$ such that there exists a probability measure on $\Theta = (0, +\infty)$, with a p-point support, having $\nu_1^\lambda, \nu_2^\lambda, \cdots \nu_{2p}^\lambda$ as first $2p$ moments. Classical moment results (see, for instance, [26], Theorem 2A, p. 726) establish that $\{\nu_k^\lambda, k = 1, 2, \dots, 2p\}$ are the first $2p$ moments of a distribution on $(0, +\infty)$ with a p-point support if, and only if, the matrix

$$M_k^\lambda = \begin{pmatrix} 1 & \nu_1^\lambda & \nu_2^\lambda & \cdots & \nu_k^\lambda \\ \nu_1^\lambda & \nu_2^\lambda & \nu_3^\lambda & \cdots & \nu_{k+1}^\lambda \\ \nu_2^\lambda & \nu_3^\lambda & \nu_4^\lambda & \cdots & \nu_{k+2}^\lambda \\ \vdots & \vdots & \vdots & \vdots & \vdots \\ \nu_k^\lambda & \nu_{k+1}^\lambda & \nu_{k+2}^\lambda & \cdots & \nu_{2k}^\lambda \end{pmatrix} \quad (24)$$

is positive definite for $k = 1, 2, \dots, p-1$ and singular for $k = p$. In [29] the authors prove that if λ is the smallest positive root λ^* of equation $det(M_p^\lambda) = 0$, these conditions are satisfied, and (23) is solvable. Next step consists in the explicit resolution of system (23), *i.e.* finding the p support points $\theta_1, \theta_2, \dots, \theta_p$ and the $p-1$ weights $\pi_1, \pi_2, \dots, \pi_{p-1}$. Of course, since Lindsay's F_{Q_p} shares its first $2p$ moments with F_{V_α}, it converges uniformly to the latter for $p \to +\infty$ by Proposition 2.1.

7 Numerical examples and comparisons

In this section, three different moment-based approximation procedures (expansions in orthonormal polynomials, maximum entropy estimates and Lindsay's gamma-mixtures) are compared by some examples.

Example 7.1 We assume $c = 1$ and $\alpha_0 = beta(1/2, 1/2)$, so that the exact density of V_α (see [7], Example 3.5(α)) is

$$f_{V_\alpha}(v) = \frac{32}{\pi}\int_{4v}^{1} x^{-1/2}(1-x)^{1/2}dx = \\ = \frac{32}{\pi}\left(\arctan(\sqrt{\frac{1-4v}{4v}}) - \sqrt{4v(1-4v)}\right), \quad v \in [0, 1/4].$$

The density f_{V_α} is bounded and continuous on $[0, 0.25]$ (this implies $f_{V_\alpha} \in L^2([0,1], dx)$) and its derivative f'_{V_α} is continuous on $(0, 0.25)$. Moreover, all the moments of V_α exist and

$$\mu_n = \mathsf{E}(V_\alpha^n) = \frac{1}{4^n(n+1)}\frac{B(n+\frac{3}{2},\frac{3}{2})}{B(\frac{3}{2},\frac{3}{2})} = \frac{1}{4^n(n+1)}\frac{4\Gamma(n+\frac{3}{2})}{\sqrt{\pi}(n+2)!}, \quad n = 0, 1, \ldots .$$

We decided to use the first 12 moments $\mu_1, \ldots, \mu_{12}$ to compare the procedures illustrated in this paper. First, we approximate the law of V_α via expansion in Jacobi polynomials $\{J_k(\cdot; 0, 0)\}$. In this case, by Proposition 4.1, the convergence of $\sum_0^{+\infty} a_k J_k(\cdot; 0, 0)$ is also uniform in $[\epsilon, 1/4 - \epsilon]$ for any $\epsilon > 0$. Actually we approximate the density of $Y := 4V_\alpha$, in order to consider the interval $[0, 1]$ as support, and finally we transform back the approximating density $\sum_0^{12} a_k J_k(\cdot; 0, 0)$, $a_k = \mathsf{E}(J_k(Y; 0, 0))$ (we use the same notation for the approximating density on $[0, 0.25]$ or on $[0, 1]$). In this case, the function $\sum_0^{12} a_k J_k(\cdot; 0, 0)$ is a density with the same first 12 moments as Y and the two functions intersect at 13 points. Moreover, the L_2-error is $\|f_Y - \sum_{k=0}^{12} a_k J_k(; 0, 0)\|_2 = 0.5340 \times 10^{-2}$. If we represent $J_k(x; 0, 0)$ as $J_k(x; 0, 0) = \sum_{j=0}^{k} b_{k,j} x^j$, then the distribution function F_Y will be uniformly approximated by

$$\sum_0^{12} a_k \sum_{j=0}^{k} b_{k,j} B_t(j+1, 1) = \sum_0^{12} a_k \sum_{j=0}^{k} b_{k,j} \frac{t^j}{j+1};$$

the error in the uniform metric between the two distribution functions (both transformed back to have $[0, 0.25]$ support) is $\sup_t |F_{V_\alpha}(t) - \sum_0^{12} a_k \sum_{j=0}^{k} b_{k,j} B_t(j+1, 1)| = 0.1802 \times 10^{-3}$. The maximum entropy estimate $g_{12}(v)$ of f_{V_α} (which has finite entropy $H(f_{V_\alpha}) = -1.8073$), using the first 12 moments as before, yields $\sup_t |F_{V_\alpha}(t) - G_{12}(t)| = 0.6022 \times 10^{-3}$. The plots of f_{V_α}, $\sum_0^{12} a_k J_k(\cdot; 0, 0)$ and g_{12} are displayed in Figure 1(a); they look indistinguishable, even if, as you can see if you zoom in on (see Figure 1(b)), they present some differences at least in the monotonicity. In any case, Figure 1, as well as the plots of corresponding distribution functions, shows a very good approximation via Jacobi polynomials and maximum entropy methods. Finally, following Lindsay, we computed a mixture of 6 gamma distributions that matches $\mu_1, \dots, \mu_{12}$, with distribution function $F_{Q_6}(v) = \sum_{j=1}^{6} \pi_j F_{\nu_j,\lambda}(v)$. The 6-component mixture has well-separated modes, as you can see from Figure 2(a), where F_{Q_6} climbs and wriggle on F_V. The error in the uniform metric is $\sup_v |F_{V_\alpha}(v) - F_{Q_6}(v)| = 0.9885 \times 10^{-1}$, but it greatly decreases in the right tail of F_{V_α} (see Figure 2(b)).

Example 7.2 Let $c = 1$, $\alpha_0 = \frac{1}{2}\delta_0 + \frac{1}{2}beta(1/2, 1/2)$, so that, by (3)-(4),

$$\mu_n = \mathsf{E}((4V_\alpha)^n) = \Big(\frac{\Gamma(n+1/2)}{\Gamma(1/2)}\Big) / \Big(\frac{\Gamma(n+2)}{\Gamma(2)}\Big), \quad n = 0, 1, 2, \dots ;$$

since the moments uniquely identify the law of $Y = 4V_\alpha$, it is easy to see that $Y \overset{d}{=} Z_1/Z_2$, where Z_1 and Z_2 are independent, $Z_1 \sim gamma(1/2, 1)$, $Z_2 \sim gamma(2, 1)$, yielding $Y \sim beta(1/2, 3/2)$, *i.e.*,

$$f_{V_\alpha}(v) = \frac{4}{\pi}\sqrt{\frac{1-4v}{v}}, \quad v \in (0, \frac{1}{4}).$$

Of course, this density has an asymptote in 0, and it is continuous, together with its first derivative f'_{V_α}, in $(0, 0.25)$. We used the first 6 moments $\mu_1, \dots, \mu_6$ to compare the methods. Jacobi polynomials $\{J_k(\cdot; 0, -\frac{1}{2})\}$ were considered, since $\int_0^1 (f_Y(x))^2 \nu^*_{0,-1/2}(dx) = \int_0^1 (f_Y(x))^2 \sqrt{x} dx < +\infty$, so that the convergence of $x^{-1/2}\sum_0^{+\infty} a_k J_k(x; 0, -1/2)$ is uniform in $[\epsilon, 1-\epsilon]$ for any $\epsilon > 0$. The $L_2(\nu^*_{0,-1/2})$-error is $\|f_Y(x) - \sum_0^6 a_k J_k(x; 0, -1/2)\|_{L_2(\nu^*_{0,-1/2})} = 0.246674 \times 10^{-2}$, while the error in the uniform metric is $\sup_v |F_{V_\alpha}(v) - \sum_0^6 a_k \sum_{j=0}^{k} b_{k,j} B_t(j + \frac{1}{2}, 1)| = 0.1596 \times 10^{-3}$. As in Example 7.1, plots of the Jacobi polynomial density (not included here) show a very good approximation of the exact distribution. On the other hand, the maximum entropy estimate $g_6(v)$ of f_{V_α}, using the first 6 moments as before, do not show the same very good behaviour. In fact, $\sup_t |F_{V_\alpha}(t) - G_6(t)| = 0.4634 \times 10^{-1}$. Figure 3($a$) displays the plot of the error between F_{V_α} and G_6. The maximum entropy approximation slightly improves considering 12 moments, instead of 6, as $\sup_t |F_{V_\alpha}(t) - G_{12}(t)| = 0.2839 \times 10^{-1}$ shows. In this case, since $H(g_6) - H(f_{V_\alpha}) = 0.5387 \times 10^{-1}$, while $H(g_{12}) - H(f_{V_\alpha}) = 0.3335 \times 10^{-1}$, the convergence of g_n to f_{V_α} in entropy seems very slow, at least comparing to Example 7.1. The 3-gamma mixture $F_{Q_3}(v) = \sum_{j=1}^{3} \pi_j F_{\nu_j,\lambda}(v)$, matching the first 6 moments of V_α, is a poorer approximation, in this case. The error in the uniform metric is $\sup_v |F_{V_\alpha}(v) - F_{Q_3}(v)| = 0.1784$; see Figure 3($b$) for a plot of F_V and F_{Q_3}.

Example 7.3 Given $c > 0$, we choose α_0 on $\mathbb{R}$ such that $\Gamma_\alpha \sim \mathcal{N}(0, 1)$; the existence of such an α_0 is a special case of the *inversion of the Markov transform* (see [39] for details) and can be guaranteed for any $c > 0$ as showed in [15]. In this case, $V_\alpha \sim gamma(c, 1)$. Of course, Laguerre polynomial expansion and Lindsay's method are expected to perform very well in this case, since approximating distributions are gamma mixtures. In fact, the expansion in Laguerre polynomials $L_n^a(x)$, with $a = c - 1$, is exact: $a_0 = \mathsf{E}(L_0^{c-1}(V_\alpha)) > 0$, $a_k = \mathsf{E}(L_k^{c-1}(V_\alpha)) = 0$ for any $k \geq 1$. If we fix a different parameter a (for instance, $a = -0.5$ with $c = 3$), and assume the first 10 moments, the approximation is still good: $\sup_v |F_{V_\alpha}(v) - \int \sum_0^{10} a_k L_k^{-0.5}(\cdot)| = 0.2293 \times 10^{-3}$. A very good performance is achieved by Lindsay's 5-gamma mixture: $\sup_v |F_{V_\alpha}(v) - L_{10}(v)| = 0.5204 \times 10^{-4}$, while the maximum entropy distribution function G_{10} does not behave so well, since $\sup_v |F_{V_\alpha}(v) - G_{10}(v)| = 0.1164 \times 10^{-1}$.

Example 7.4 Let us consider the dataset of $n = 101$ stress-rupture lifetimes of Kevlar strands, $x_1, \ldots, x_n$, extracted from the book by Andrews and Herzberg, Table 29.1, also available via StatLib at CMU. We approximate the posterior distribution of the unknown V_α when a priori the random probability measure is a Dirichlet process with a negative exponential distribution of parameter 1 ($\mathcal{E}(1)$) as α_0 and prior total mass equal to 1. By the conjugacy property of the Dirichlet process, the posterior distribution of the variance, given $x_1, \ldots, x_n$, coincides with the law of V_α with $\alpha = \mathcal{E}(1) + \sum_1^{101} \delta_{x_i}$. Hence $c = 102$ and the support of V_α is $[0, +\infty)$. We approximate $\mathcal{L}(V_\alpha)$ using 8 moments.

We do not have an explicit expression of the exact distribution of V_α, but we are able to simulate from a Markov chain with $\mathcal{L}(V_\alpha)$ as invariant distribution; therefore we compare the different approximations of the previous sections with the empirical distribution function F_{emp} from the simulation procedure. First of all, since the total mass c is relatively large, the normal limit law discussed in Section 3 holds. In particular, $\mathcal{L}(V_\alpha) \simeq \mathcal{N}(1.238, 0.376)$. Secondly, we compute a Laguerre polynomial approximation of the density, for different values of the parameter a, but all of them turn out to be negative in some subintervals (close to $\mathsf{E}(V_\alpha)$). Third, the maximum entropy estimate g_8, which is unimodal and not too asymmetric, yields a distribution function which is not a good approximation of the empirical distribution function F_{emp} at all. On the other hand, the 4-gamma mixture $F_{Q_4}(v)$ behaves well, with an error in the uniform metric (w.r.t. F_{emp}) equal to 0.1571626×10^{-1}, which reduces to 0.7084339×10^{-2} for $v > \mathsf{E}(V_\alpha)$. Figure 4 displays a plot of F_{emp} and F_{Q_4}, showing a very good behaviour of the latter.

8 Concluding remarks

Summing up, as already pointed out, knowledge of α yields information on the support of V_α, as well as on its moment sequence $(\mu_n)_n$, but generally we have no clue on the smoothness of f_{V_α}. However the examples show that further information on $\mathcal{L}(V_\alpha)$ can help selecting the "best" approximation procedure. For instance, expansion in orthonormal polynomials works well if $f_{V_\alpha}/w_{a,b}$ satisfies the $L^2([0,1],\nu_{a,b})$-boundedness condition (11); the same when the support of V_α is $[0,+\infty)$. This condition can in principle be checked via the knowledge of $(\mu_n)_n$, but an infinite number of inequalities should be verified. However, if f_{V_α} is $L^2(\nu^*_{a,b})$-bounded for some $a, b < 1$, approximating functions based on finite expansions in orthonormal polynomials can behave very well, provided they are densities, *i.e.* if they are nonnegative. On the other hand, maximum entropy estimates might not exist when the support of V_α is $[0,+\infty)$; moreover, the solution of (14) is numerically unstable, so that a very accurate coding of the algorithm must be implemented. Finally, Lindsay's gamma-mixtures suggest an "automatic" approximation method for F_{V_α} that really works very well when the support of V_α is unbounded and no other information is available, except for the moment sequence. Besides, Lindsay's approximating distributions require the computation of a sum of a number of terms smaller than the other two methods considered here.

Note that our approximation results apply also to the variance of a Dirichlet mixture of location-invariant densities, $f(t) = \int_{\mathbb{R}} g(t-x)P_\alpha(dx)$, where g is a density with finite second moment. Popular choices for g are the normal density with x-mean and known variance and the exponential distribution shifted by x. The model can be described as follows:

- the conditional density of an observation T, given X, is $g(\,\cdot\, - X)$;
- the conditional distribution of X, given P_α, is P_α;
- P_α is a Dirichlet process on $\mathbb{R}$.

The random variance V_f of f can be represented as $V_f = V_\alpha + \sigma_g^2$, being σ_g^2 the variance of g. Hence, the prior distribution of V_f is that of V_α shifted by σ_g^2. Having observed a sample $T_1, \ldots, T_n$ from f, the posterior distribution of V_f is a mixture of laws of Dirichlet random variances $V_{\alpha+\sum_1^n \delta_{x_i}}$, with mixing measure equal to the posterior distribution of $X_1, \ldots, X_n$, given $T_1 = t_1, \ldots, T_n = t_n$; we refer to [30] for more details.

ACKNOWLEDGEMENTS. We are grateful to thank Aldo Tagliani, who kindly provided us a computer code.

References

[1] N.I. Akhiezer, *Classical Moment Problem and Some Related Questions in*

Analysis (translated by N.Kemmer), Oliver & Boyd, London, 1965.

[2] F. Altomare and M. Campiti, *Korovkin-type approximation theory and its applications,* De Guyter, Berlin, 1994.

[3] P. Billingsley, *Probability and Measure,* Wiley, New York, 1995.

[4] J.M. Borwein and W.Z. Huang, A fast heuristic method for polynomial moment problems with Boltzmann-Shannon entropy, *SIAM J. Optimization,* **5** (1995), 68 - 99.

[5] J.M. Borwein and A.S. Lewis, Convergence of best entropy estimates, *SIAM J. Optimization,* **1** (1991), 191 - 205.

[6] G. Castelnuovo, Sul problema dei momenti, *Giornale dell'Istituto Italiano degli Attuari,* **1** (1930), 137 - 169.

[7] D.M. Cifarelli and E. Melilli, Some new results for Dirichlet priors, *Ann. Statist.,* **28** (2000), 1390 - 1413.

[8] D.M. Cifarelli and E. Regazzini, Distribution functions of means of a Dirichlet process, *Ann. Statist.,* **18** (1990), 429 - 442. (Correction in *Ann. Statist.,* **22** (1994), 1633 - 1634).

[9] I. Csiszar, *I*-divergence geometry of probability distributions and minimization problems, *Ann. Probab.,* **3** (1975), 146 - 158.

[10] D. Dacunha-Castelle and F. Gamboa, Maximum d'entropie et probleme des moments, *Ann. Inst. H. Poincaré Probab. Statist.,* **26** (1990), 567 - 596.

[11] S.R. Dalal and G.J. Hall, On approximating parametric Bayes models by nonparametric Bayes models, *Ann. Statist.,* **8** (1980), 664 - 672.

[12] H. Dette and W.J. Studden, *The theory of canonical moments with applications in statistics, probability and analysis,* Wiley, New York, 1997.

[13] P. Diaconis and D. Freedman, The Markov moment problem and de Finetti's theorem: Part I, *Math. Z.,* **247** (2004), 183 - 212.

[14] I. Epifani, Some results on the moments of vector-valued functionals of the Ferguson-Dirichlet process, *Technical Report Dipartimento di Matematica, Politecnico di Milano,* **No 487/P** (2001).

[15] I. Epifani, A. Guglielmi and E. Melilli, Some new results on random Dirichlet variances, *Technical Report IMATI,* **2004 - 15-MI** (2004).

[16] I. Epifani, A. Guglielmi and E. Melilli, A stochastic equation for the law of the random Dirichlet variance, *Statist. Probab. Lett.,* **76** (2006) 495 - 502.

[17] A. Erdely, *Asymptotic Expansions,* Dover, New York, 1956.

[18] W. Feller, *An introduction to probability theory and its applications, Vol. II,* Wiley, New York, 1971.

[19] M. Frontini and A. Tagliani, Entropy-convergence in Stieltjies and Hamburger moment problems, *Appl. Math. Comput.,* **88** (1997) 39 - 51.

[20] J.K. Ghosh and R.V. Ramamoorthi, *Bayesian Nonparametrics,* Springer, New York, 2003.

[21] A. Guglielmi and R.L. Tweedie, Markov chain Monte Carlo estimation of the law of the mean of a Dirichlet process, *Bernoulli,* **7** (2001), 573 - 592.

[22] N.L. Hjort and A. Ongaro, Exact inference for random Dirichlet means, *Stat. Inference Stoch. Process.,* **8** (2005), 227 - 254.

[23] G. Inglese, A note about the discretization of finite moment problems, *Inverse Problems,* **10** (1994) , 401 - 414.

[24] L.F. James, New Dirichlet mean identities, *The arXiv, arXiv:0708.0614v2 [math.PR],* (2007), Available at http://arxiv.org/abs/0708.0614v2.

[25] M. Junk, Maximum Entropy for Reduced Moment Problems, *Math. Models Methods Appl. Sci.,* **10** (2000), 1001 - 1025.

[26] B.G. Lindsay, Moment matrices: applications in mixtures, *Ann. Statist.,* **17** (1989), 722 - 740.

[27] B.G. Lindsay and P. Basak, Moments determine the tail of a distribution (but not much else), *Amer. Statist.,* **54** (2000) , 248 - 251.

[28] B.G. Lindsay and K. Roeder, Moment-based oscillation properties of mixture models, *Ann. Statist.,* **25** (1997), 378 - 386.

[29] B.G. Lindsay, R.S. Pilla and P. Basak, Moment-based approximations of distributions using mixtures: Theory and applications, *Ann. Inst. Statist. Math.,* **52** (2000), 215 - 230.

[30] A.Y. Lo, On a class of Bayesian nonparametric estimates: I. Density estimates, *Ann. Statist.,* **12** (1984), 351 - 357.

[31] L.R. Mead and P. Papanicolaou, Maximum Entropy in the Problem of Moments, *J. Math. Phys.,* **25** (1984) , 2404 - 2417.

[32] C. N. Morris, Natural exponential families with quadratic variance functions, *Ann. Statist.,* **10** (1982) , 65 - 80.

[33] C. N. Morris, Natural exponential families with quadratic variance functions: Statistical theory, *Ann. Statist.,* **11** (1983), 515 - 529.

[34] P. Müller and F.A. Quintana, Nonparametric Bayesian Data Analysis, *Statist. Sci.,* **19** (2004), 95 - 110.

[35] J. Prasad and H. Hayashi, On the Uniform Convergence of Fourier-Jacobi Series, *SIAM J. Numer. Anal.,* **10** (1973), 23 - 27.

[36] E. Regazzini, A. Guglielmi and G. Di Nunno, Theory and numerical analysis for exact distributions of functionals of a Dirichlet process, *Ann. Statist.,* **30** (2002) , 1376 - 1411.

[37] E. Regazzini, A. Lijoi and I. Prünster, Distributional results for means of normalized random measures with independent increments, *Ann. Statist.,* **31** (2003), 560 - 585.

[38] E. Regazzini and V.V. Sazonov, Approximation of laws of random probabilities by mixtures of Dirichlet distributions with applications to nonparametric Bayesian inference. *Theory Probab. Appl.,* **45** (2000), 93 - 110.

[39] D. Romik, Roots of the derivative of a polynomial, *Amer. Math. Monthly,* **112** (2005) , 66 - 68.

[40] G. Sansone, *Orthogonal functions,* Interscience, London, 1959.

[41] J. Sethuraman, A constructive definition of Dirichlet priors, *Statist. Sinica,* **4** (1994), 639 - 650.

[42] J.A. Shohat and J.D. Tamarkin, *The problem of Moments,* AMS, Providence, Rhode Island, 1963.

[43] G. Szego, *Orthogonal polynomials, Third Edition,* AMS, Providence, Rhode Island, 1967.

[44] A. Tagliani, Numerical aspects of finite Hausdorff moment problem by maximum entropy approach, *Appl. Math. Comput.,* **118** (2001), 133 - 149.

[45] L. Tardella, A note on estimating the diameter of a truncated moment class. *Statist. Probab. Lett.,* 54 (2001), 115 - 124.

Received: October, 2008

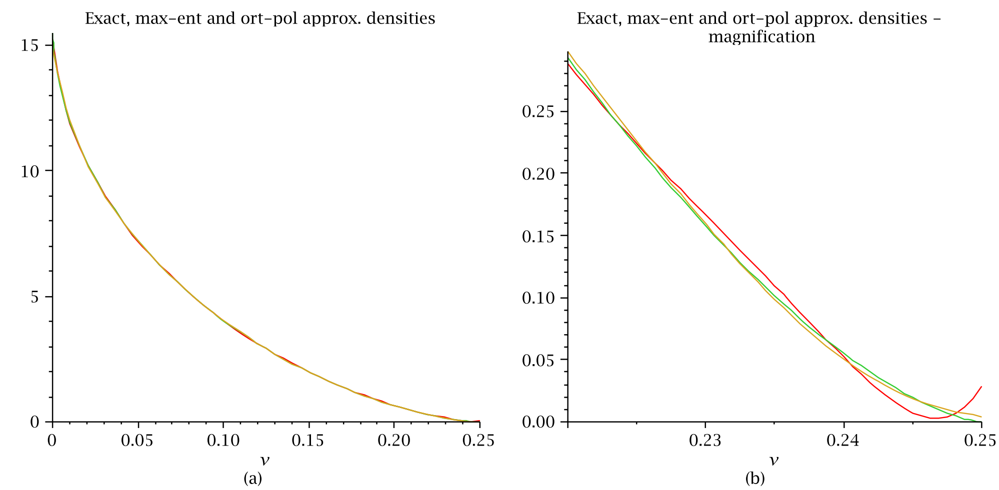

Figure 1. Comparison between f_{V_α} (green) and two approximating densities using the first 12 moments in Example 7.1: the Jacobi polynomial $\sum_0^{12} a_k J_k(\cdot; 0, 0)$ (red) and the maximum entropy g_{12} (yellow).

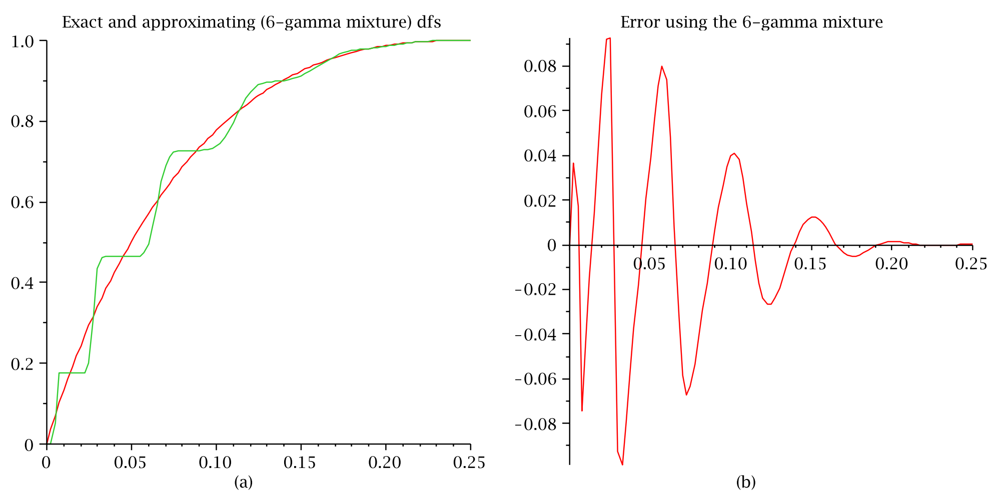

Figure 2. Comparison between F_{V_α} (red) and F_{Q_6} (green) on panel (a), plot of the error $F_{V_\alpha} - F_{Q_6}$ on panel (b), for Example 7.1.

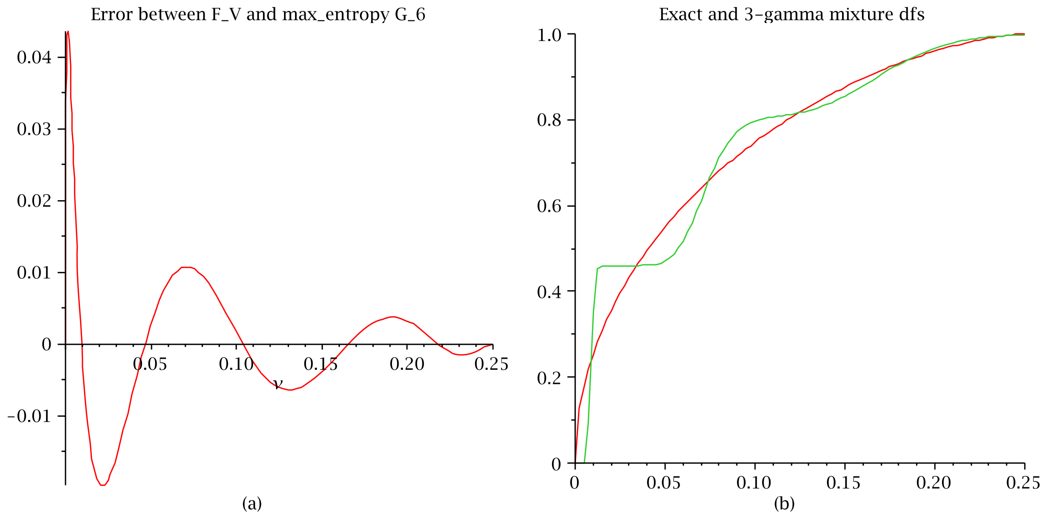

Figure 3. Comparison between approximations in Example 7.2: maximum entropy error error $F_{V_\alpha}(v) - F_{Q_3}(v)$ on the left (a), $F_{V_\alpha}(v)$ (red) and 3-gamma mixture $F_{Q_3}(v)$ (green) on the right (b).

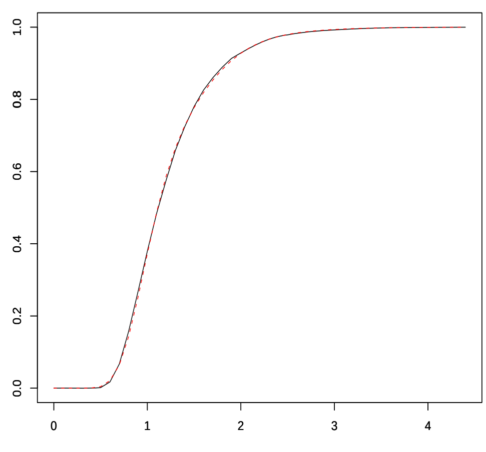

Figure 4. Comparison between $F_{emp}(v)$ and $F_{Q_4}(v)$ (red dashed) in Example 7.4.