Electronic Journal of Statistics
Vol. 14 (2020) 1611–1647
ISSN: 1935-7524
https://doi.org/10.1214/20-EJS1697

Random distributions via Sequential Quantile Array

Annalisa Fabretti and Samantha Leorato

Department of Economics and Finance
University of Rome Tor Vergata
e-mail: annalisa.fabretti@uniroma2.it

Department of Economics, Management, and Quantitative Methods
University of Milan
e-mail: samantha.leorato@unimi.it

Abstract: We propose a method to generate random distributions with known quantile distribution, or, more generally, with known distribution for some form of generalized quantile. The method takes inspiration from the random Sequential Barycenter Array distributions (SBA) proposed by Hill and Monticino (1998) which generates a Random Probability Measure (RPM) with known expected value. We define the Sequential Quantile Array (SQA) and show how to generate a random SQA from which we can derive RPMs. The distribution of the generated SQA-RPM can have full support and the RPMs can be both discrete, continuous and differentiable. We face also the problem of the efficient implementation of the procedure that ensures that the approximation of the SQA-RPM by a finite number of steps stays close to the SQA-RPM obtained theoretically by the procedure. Finally, we compare SQA-RPMs with similar approaches as Polya Tree.

MSC 2010 subject classifications: Primary 60G57; secondary 62G07.
Keywords and phrases: Quantiles, random probability measures, M-quantiles.

Received November 2018.

Contents

1. Introduction

Random probability measures (RPM) find their applications in several different fields, such as statistics, mathematical finance, stochastic processes.

In nonparametric Bayesian statistics, for example, the construction of random probability distributions permits to draw a prior at random from the space of probability measures. In this context, the Dirichlet process (DP) model represents a rather common choice despite being only able to generate discrete distributions with probability one (see [12] and [13] for a review on the topic). Several generalizations of the Dirichlet processes have been proposed, among which we can mention the wider class of the Stick-breaking priors ([19]). Methods that permit to generate continuous prior distributions have also been studied: for example the Polya Tree models (including DP as particular cases) and Bernstein processes (see [26] for a more detailed review on RPM used in Bayesian data analysis). A method to produce RPM with given mean is proposed by [17], while [4] studied a method to construct RPM with given mean and variance.

RPM also occur in some stochastic processes describing mass in space ([25] and references therein) or in random dynamical systems ([9]). In particular, random dynamical systems are defined through a random measurable map $\varphi : T \times \Omega \times X \mapsto T \times \Omega \times X$ that, for all $(t, \omega) \in T \times \Omega$ maps Borel measures on X into a random Borel measure on X. RPM can be used to generate scenarios on which taking decisions under uncertainty or ambiguity. For example, [25] suggests that RPM may be used to solve a random optimal stopping problem.

Inspired by the Sequential Barycenter Array distributions (SBA) of [17], in this paper we propose an alternative to the SBA procedure, which we denote as the Sequential Quantile Array (SQA), that allows to construct random distribution functions with the τ-quantile following a given distribution (for some $\tau \in (0, 1)$). For example in finance, this method can be used to generate random probability measures whose associated Value at Risk (VaR), that is a risk measure largely used, follows a specified distribution. Moreover, by exploiting their link with the quantiles, it is possible to generate RPMs with specified distribution for alternative risk measures based on the general notion of M-quantiles defined by [5]. The expectiles belong to the family of M-quantiles, and might be seen as a generalization of the expectation, because in particular the expectile of level $\tau = 1/2$ coincides with the mean. In this sense, our method can be seen as a possible way to broaden the principle behind the work of [17]. We note that [6] simulate RPMs with a specified value of a certain risk measure (they focus in particular on the VaR and the expected shortfall (ES)). However, except for a common interest in controlling for quantiles of RPMs, our work is different from [6] in all respects. In fact, we construct a new procedure for the generation of RPMs, that allows for the chosen quantile to have a given distribution

λ_1, which needs not be concentrated on a constant as in [6]. Further, we study the properties of the random distributions generated through our procedure, by assessing conditions allowing for continuity, full support and differentiability of the RPMs. Conversely, their approach to the problem is purely algorithmic and is based on an *ad-hoc* transformation of a realized RPM obtained from the implementation of either [11], [15] or [17] procedures.

The paper is organized as follows: in Section 2, we present the notion of Sequential Quantile Array and we show that it shares most of the properties of the SBA. Section 2.2 focuses in particular on the case when the base measure generating the SQA sequence is discrete. In Section 3 we describe how to construct SQA-RPM and study their properties, namely we study conditions for continuity, differentiability and full support. Moreover, we show that the SQA-RPM behaves differently from the SBA-RPM, because it can produce families of distributions that include both discrete and continuous distributions with positive probability. In Section 3 we also present a stopping rule and a truncation algorithm for computational issues relative to the efficient implementation of the procedure. In this regard, we prove that the approximated probability measure obtained by truncation is in a γ-neighborhood relative to strong topologies of the *true* SQA-RPM with probability arbitrarily close to one. Section 4 discusses how the SQA-RPM can be used to obtain a distribution with prescribed generalized forms of quantiles and we describe how to generalize the procedure for two quantiles and to generate random distributions with unbounded support. Section 5 presents some simulations of RPM and we consider also the comparison of SQA-RPM to other methods proposed in the literature, such as Polya trees and quantile pyramids. The latter, proposed by [18], is based on the similar idea of generating dyadic quantiles. Finally Section 6 concludes with few possible applications.

2. Sequential Quantile Array (SQA)

Let $\tau \in (0,1)$ and G an arbitrary distribution function. We denote by $q^\tau(G)$ the τ-th level quantile of G, namely $q^\tau(G) = G^{-1}(\tau)$, where G^{-1} is replaced by the generalized inverse function, that is $q^\tau(G) = \inf\{y : G(y) \geq \tau\}$, if G is not invertible everywhere in its support.

Definition 1. *We call quantile of G in $(a,c]$ of level τ the following function:*

$$B_G^\tau((a,c]) = \begin{cases} q^\tau(G \mid X \in (a,c]) & \text{if } G(c) > G(a) \\ a & \text{if } G(c) = G(a) \end{cases} \tag{2.1}$$

where $q^\tau(G \mid X \in (a,c])$ is the generalized inverse of $G(\cdot \mid X \in (a,c])$ that is the distribution of $X \sim G$ conditional to $\{X \in (a,c]\}$.

In what follows, we shall write $G(\cdot \mid (a,c]) = G(\cdot \mid X \in (a,c])$ to shorten the notation.

2.1. SQA of continuous distributions

We first focus on the case of a continuous distribution G, with a density g, we will consider the definition of the SQA in the case of a discrete distribution later in this section.

Lemma 2.1. *Let $\tau \in (0,1)$ and G a continuous distribution function and $b = B_G^\tau(a,c]$, with $a \leq c$ then: (i) $G(c) > G(a)$ if and only if $b > a$; (ii) $G(b) = \tau(G(c) - G(a)) + G(a)$; (iii) $B_G^\tau(a,b] = b$ if and only if $B_G^\tau(b,c] = b$; (iv) for every $x \in (a,c]$, $b \geq B_G^\tau(a,x]$ for all τ.*

Proof. *(i)* If $G(c) > G(a)$ then the implication $b > a$ comes directly from the definition and the continuity of G. To prove the backward implication, let $b = a$, then $G(b) = G(a)$. If we assume by contradiction that $G(c) > G(a)$ we have, by the definition of b, that $\tau = G(b \mid (a,c])$, but if $G(b) = G(a)$, it comes that $G(b \mid (a,c]) = \frac{G(b)-G(a)}{G(c)-G(a)} = 0$.

(ii) Follows from $\tau = \frac{G(b)-G(a)}{G(c)-G(a)}$, if $G(c) > G(a)$, and $G(b) = G(a)$, if $G(c) = G(a)$.

(iii) Because of the continuity of G, $B_G^\tau(a,b] = b$ only if $a = b$ $(G(a) = G(b))$. From part *(i)*, this also implies that $G(a) = G(c)$, thus also $G(b) = G(c)$ and $B_G^\tau(b,c] = b$ too.

(iv) For every $x < c$, $G(c) - G(a) > G(x) - G(a)$, then, from the definition of $B_G^\tau(a,x]$ and $B_G^\tau(a,c]$, it comes

$$\tau = \frac{G(b) - G(a)}{G(c) - G(a)} = \frac{G(B_G^\tau(a,x]) - G(a)}{G(x) - G(a)} > \frac{G(B_G^\tau(a,x]) - G(a)}{G(c) - G(a)}.$$

Because of $G(B_G^\tau(a,x] \mid (a,c]) < G(b \mid (a,c])$, and by monotonicity and continuity of G, we must have that $b > B_G^\tau(a,x]$. □

The above lemma is analogous to Lemma 2.2. in Hill and Monticino (1998) [17] and it is useful to study the properties of the sequential quantile arrays defined below.

Definition 2. *The Sequential Quantile Array (SQA) of level τ of the distribution function G is the triangular array $\mathcal{Q}(G,\tau) := \{q_{n,k}(G,\tau)\} = \{q_{n,k}\}_{n\geq 1, k\leq 2^n}$ defined by induction:*

$$q_{1,1} = B_G^\tau(-\infty,\infty) = q^\tau(G)$$
$$q_{n,2j} = q_{n-1,j}, \quad n \geq 1,\ j = 1,\dots,2^{n-1} - 1 \tag{2.2}$$

and

$$q_{n,2j-1} = B_G^\tau(q_{n-1,j-1}, q_{n-1,j}], \quad n \geq 1,\ j = 1,\dots,2^{n-1} \tag{2.3}$$

with $q_{n,0} = -\infty$ and $q_{n,2^n} = \infty$ for all n.

Let us define the intervals $I_{n,k} = (q_{n,k-1}, q_{n,k}]$.

Lemma 2.2. *Let $\{q_{n,k}(G,\tau)\}$ be the SQA of G of level τ. Then: (i) If $G(c) > G(a)$, there exist j and n such that $q_{n,j} \in [a,c]$; (ii) $\{q_{n,k}(G,\tau)\}$ is dense in the support of G; (iii) for every $n \geq 1$, $\{I_{n,k}\}_k$ defines a partition of $\mathbb{R}$. $\{I_{n+1,k}\}_k$ is a refinement of $\{I_{n,k}\}_k$; (iv) $\Pr(cl(I_{n,k})) > 0$ for all $n \geq 1$, $k = 1, \ldots, 2^n$, where $cl(I)$ is the closure of the interval I.*

Proof. Points *(i)* to *(iii)* are straightforward from the Definition 1.

Let's prove *(iv)* by induction. Let $n = 1$. Then, $\Pr([q_{1,0}, q_{1,1}]) = G(q_{1,1}) = \tau > 0$ and $\Pr([q_{1,1}, q_{1,2}]) = 1 - \tau > 0$. Now, assume that $\Pr([q_{n-1,k-1}, q_{n-1,k}]) > 0$ holds true for all $k \leq 2^{n-1}$. We show that this implies $\Pr([q_{n,k-1}, q_{n,k}]) > 0$ for all $k \leq 2^n$. In fact, let $k = 2j$, $j = 1, \ldots, 2^{n-1}$, then, $\Pr([q_{n,2j-1}, q_{n,2j}]) = \Pr([q_{n,2j-1}, q_{n-1,j}])$. Since $q_{n,2j-1} = B_G^\tau(q_{n-1,j-1}, q_{n-1,j}] > q_{n-1,j-1}$, because of the induction assumption and of Lemma 2.1-*(i)* and *(ii)*, we get

$$\Pr([q_{n,2j-1}, q_{n,2j}]) \geq \tau \Pr((q_{n-1,j-1}, q_{n-1,j}]) > 0.$$

Similarly, if $k = 2j - 1$, $j = 1, \ldots, 2^{n-1}$, we have $\Pr([q_{n,2(j-1)}, q_{n,2j-1}]) = \Pr([q_{n-1,j-1}, q_{n,2j-1}]) > 0$ because $q_{n,2j-1} = B_G^\tau(q_{n-1,j-1}, q_{n-1,j}] > q_{n-1,j-1}$ and from Lemma 2.1-*(i)*. □

The following theorem is a straightforward consequence of Lemma 2.1 and of Definition 2.

Theorem 2.1. *The distribution G is completely determined by $\{q_{n,k}(G)\}$. In particular, G is given inductively by $G(q_{n,0}) = 0 = 1 - G(q_{n,2^n})$ and*

$$G(q_{n,2k-1}) = \tau G(q_{n-1,k}) + (1-\tau) G(q_{n-1,k-1})$$

while

$$G(q_{n,2k}) = G(q_{n-1,k}).$$

Theorem 2.2. *A sequence $\mathcal{Q} = \{q_{n,k}\}$ is a SQA for some continuous distribution function G if and only if $q_{n,2k} = q_{n-1,k}$ and $q_{n,k-1} < q_{n,k}$, for all $n \geq 1$, $k = 1, \ldots, 2^n$.*

Proof. Let us assume that $\mathcal{Q}$ is a SQA sequence for a continuous d.f. G. Then, from the definition of SQA, we have that $q_{n,2k} = q_{n-1,k}$, for all n and k. Moreover, if we assume by contradiction that $q_{n,k} = q_{n,k-1}$ for some $1 < n < \infty$, then $G(q_{n+1,2k-1}) = G(q_{n,k}) = G(q_{n,k-1})$ and the distribution G could not be continuous because it would have a probability mass in $q_{n,k}$ greater than or equal to $\min(\tau^n, (1-\tau)^n)$.

Let us now assume the triangular array $\mathcal{Q}$ satisfies the two conditions of the theorem. Then, we can define a function G, by the recursive relation:

$$\begin{aligned} &G(q_{n,0}) = 0 \\ &G(q_{n,2^n}) = 1 \\ &G(q_{n,2k-1}) = \tau G(q_{n-1,k}) + (1-\tau) G(q_{n-1,k-1}). \end{aligned}$$

We immediately get that G is a distribution function satisfying $G(q_{n,2k-2}) = G(q_{n-1,k-1}) < G(q_{n,2k-1}) < G(q_{n-1,k}) = G(q_{n,2k})$, for all $n \geq 1$ and $k \leq 2^{n-1}$ and is, therefore, continuous. □

2.2. SQA of discrete distributions

In this section, we consider the case of a discrete distribution G, with jumps $\{\alpha_j\}$ at points $\{x_j\}$, $j \in J$. We prove that, although the SQA is able to reconstruct the support of G, it is not necessarily able to find the masses α_j in general, and thus there is not a one-to-one correspondence between G and its SQA of level τ.

The definition of $B_G^\tau((a,c])$ in the case of a discrete G is the same as (2.1), but here $q^\tau(G \mid (a,c]) = \inf\{y \in (a,c] : G(y) \geq \tau(G(c) - G(a)) + G(a)\}$. Then, for given τ, a, c, we have that $G(B_G^\tau((a,c])) \geq \tau(G(c) - G(a)) + G(a)$, where the identity always holds if $G(a) = G(c)$, but not in general.

Lemma 2.3. *Let G be a discrete distribution with support $\mathcal{X} = \{x_j\}$ and mass probabilities $\{\alpha_j\}$, $j \in J$. Let moreover $b = B_G^\tau(a,c]$, with $a \leq c$. Let $\mathcal{Q} = \{q_{n,k}(G,\tau)\}$ be a SQA of G of level τ and let $I_{n,k} = (q_{n,k-1}, q_{n,k}]$, $n \geq 1$, $k \leq 2^n$.*

(i) For all n, k, there exists a $j \in J$ such that $q_{n,k} = x_j$.
(ii) $G(c) = G(a)$ if and only if $b = a$.
(iii) $G(b) = \tau(G(c) - G(a)) + G(a)$ if and only if either $G(a) = G(c)$ or there exists $J_2 \subset J_1 \subset J$ such that $\tau = \frac{\sum_{i\in J_2}\alpha_i}{\sum_{i\in J_1}\alpha_i}$, where $\sum_{i\in J_1}\alpha_i = G(c) - G(a)$ and $\sum_{i\in J_2}\alpha_i = G(b) - G(a)$.
(iv) For every $x \in (a,c]$, $b \geq B_G^\tau(a,x]$ for all τ.
(v) $\mathcal{Q} = \{q_{n,k}(G,\tau)\}_{n=1,k}^\infty = \mathcal{X}$ if and only if $\tau \leq \min_j \alpha_j/(\alpha_j + \alpha_{j+1})$.
(vi) For every $n \geq 1$, $\{I_{n,k}\}_k$ defines a partition of $\mathbb{R}$. $\{I_{n+1,k}\}_k$ is a refinement of $\{I_{n,k}\}_k$
(vii) $\Pr(cl(I_{n,k})) > 0$ for all $n \geq 1$, $k = 1, \ldots, 2^n$, where $cl(I)$ is the closure of the interval I.

Proof. Points *(iii)* and *(vi)* are straightforward, while *(vii)* is a direct consequence of *(i)*.

We prove *(i)* by induction. Given $\tau \in (0,1)$, $q_{1,1}$ satisfies $G(q_{1,1}) \geq \tau$ and $G(q') < \tau$ for all $q' < q_{1,1}$. Then $q_{1,1}$ must coincide with a discontinuity point of G. For $n > 1$, given that all $q_{n-1,k}$ are discontinuity points of G, we have that $q_{n,2k} = q_{n-1,k}$ is a discontinuity point of G and $q_{n,2k-1}$ satisfies $G(q_{n,2k-1}) \geq \tau(G(q_{n-1,k}) - G(q_{n-1,k-1})) + G(q_{n-1,k-1})$, but $G(q') < \tau(G(q_{n-1,k}) - G(q_{n-1,k-1})) + G(q_{n-1,k-1})$, for all $q' < q_{n,2k-1}$, which implies that in $q_{n,2k-1}$ there must be a jump of the distribution.

(ii) If $G(c) = G(a)$, then $b = a$ by definition. If $b = a$, then $G(b) = G(a)$ can satisfy the inequality $G(b) = G(a) \geq \tau(G(c) - G(a)) + G(a)$ if and only if $G(c) = G(a)$.

(iv) If $b = a$, then (point *(ii)*) $G(c) = G(a)$ and consequently also $G(x) = G(a)$ and the equality holds. Let then $b > a$. From the definition of b and from monotonicity of G, $G(b) \geq \tau(G(c)-G(a))+G(a) \geq \tau(G(x)-G(a))+G(a)$. Since $B^\tau((a,x]) = \inf\{y \in (a,x] : G(y) \geq \tau(G(x) - G(a)) + G(a)\}$, if $B^\tau((a,x]) > b$, then necessarily $a < b < x$ and thus $B^\tau((a,x])$ would not be the smallest point in $(a,x]$ to satisfy $G(B^\tau((a,x])) \geq \tau(G(x) - G(a)) + G(a)$, thus contradicting the definition of $B^\tau((a,x])$.

(v) We prove that for every $(a, c] \cap \mathcal{X} \neq \emptyset$, there exists a n, k such that $q_{n,k} \in [a, c] \cap \mathcal{X}$, if and only if $\tau \leq \min_j \alpha_j/(\alpha_j + \alpha_{j+1})$. Using then *(i)* yields immediately *(v)*. If either a or c belong to $\mathcal{X}$ the claim is proved, then let us assume that neither of them is in $\mathcal{X}$. Since we assume that $(a, c] \cap \mathcal{X} = \{x_j, j \in J_1\} \neq \emptyset$, we also have that $G(c) > G(a)$. Note that, $\min_j \alpha_j/(\alpha_j + \alpha_{j+1}) \geq \tau$ implies $\sum_{j \leq h} \alpha_j / \sum_{j \leq h+1} \alpha_j \geq \tau$. Then, by writing $J_1 = \{j_a, j_a + 1, \ldots, j_c\}$, we have

$$\frac{G(x_{j_c-1}) - G(a)}{G(c) - G(a)} = \frac{G(x_{j_c-1}) - G(x_{j_a-1})}{G(x_{j_c}) - G(x_{j_a-1})} = \frac{\sum_{j=j_a}^{j_c-1} \alpha_j}{\sum_{j=j_a}^{j_c} \alpha_j} \geq \tau$$

and $b = B^\tau((a, c])$ is equal to x_{j_c-1} if further $\frac{G(x_{j_c-2}) - G(x_{j_a}-1)}{G(x_{j_c}) - G(x_{j_a}-a)} < \tau$. In general, we have that

$$b = \min \left\{ x_j, j \in J_1 : \frac{G(x_j) - G(x_{j_a} - 1)}{G(x_{j_c}) - G(x_{j_a} - a)} \geq \tau \right\}.$$

To prove the *only if* part, it is enough to show that if $\tau > \alpha_{j^*}/(\alpha_{j^*} + \alpha_{j^*+1})$ for some $j^* \in J$, then there is a $x_j \notin \mathcal{Q}$. Let in fact $q_{n,k} = x_{j^*+1}$ and $q_{n,k-1} = x_{j^*-1}$ (if no such n, k exist, then either $x_{j^*+1} \notin \mathcal{Q}$ or $x_{j^*-1} \notin \mathcal{Q}$). Then, by definition, we have that $q_{n+1,2k-1}$ satisfies $\tau \leq \frac{G(q_{n+1,2k-1}) - G(q_{n,k-1})}{G(q_{n,k}) - G(q_{n,k-1})}$. Since however $\frac{G(x_{j^*}) - G(q_{n,k-1})}{G(q_{n,k}) - G(q_{n,k-1})} = \alpha_{j^*}/(\alpha_{j^*} + \alpha_{j^*+1}) < \tau$, $q_{n+1,2k-1}$ must be equal to $q_{n,k} = x_{j^*+1}$. Therefore, there is no way x_{j^*} can be in $\mathcal{Q}$. □

Part *(iii)* of Lemma 2.3 prevents the SQA sequence to completely define an arbitrary discrete distribution G.

Example 1. *As a simple counterexample, let us consider the distribution G, with finite support $\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}$ and probability masses $\mathbf{p} = \{0.1, 0.2, 0.4, 0.2, 0.1\}$. Let $\tau = 0.1$. Then, we easily find that the SQA is $\mathcal{Q} = \{x_1, x_2, x_3, x_4, x_5\}$. However, knowing $\tau = 0.1$ and $\mathcal{Q}$ does not allow to define univocally the probabilities $\mathbf{p}$. In fact, any distribution satisfying the condition in Lemma 2.3-(v) would produce the same SQA (try for example $\{0.1, 0.3, 0.2, 0.3, 0.1\}$). Note also that in this example, if $\tau > 0.34$, then $\mathcal{Q} = \{x_2, x_3, x_4, x_5\} \subset \mathcal{X}$.*

Example 2. *Let G be a uniform discrete distribution over $\mathcal{X} = \{x_1, \ldots, x_n\}$. Then $\mathcal{Q} = \mathcal{X}$ holds if and only if $\tau \leq 0.5$.*

The following theorem is a generalization of Theorem 2.2 and includes the case of discrete distributions G.

Theorem 2.3. *A sequence $\mathcal{Q} = \{q_{n,k}\}$ is a SQA for some distribution function G if and only if: (i) $q_{n,2k} = q_{n-1,k}$; (ii) $q_{n,k-1} \leq q_{n,k}$, for all $n \geq 1$, $k = 1, \ldots, 2^n$; (iii) $q_{n,2k-1} = q_{n,2k}$ if and only if $q_{n-1,k-1} = q_{n-1,k}$.*

Proof. The necessity of *(i)–(iii)* follows from the definition of a SQA of a distribution function G and from Lemma 2.2. For the sufficiency, let $\{q_{n,k}\}$ satisfy *(i)–(iii)*. Let us define a sequence of discrete random variables: X_n, such that X_n has support $\{q_{n,1}, \ldots, q_{n,2^n}\}$, and probabilities defined recursively. For $n = 1$,

$X_1 = q_{1,1}$ with probability τ and $X_1 = q_{1,2}$ with probability $1-\tau$. For $n > 1$ and $k \leq 2^{n-1}$, if $q_{n,2k-1} \neq q_{n,2k}$:

$$\begin{aligned} &\Pr\left(X_n \in (q_{n,2k-2}, q_{n,2k-1}]\right) = \tau \Pr\left(X_{n-1} \in (q_{n-1,k-1}, q_{n-1,k}]\right), \\ &\Pr\left(X_n \in (q_{n,2k-1}, q_{n,2k}]\right) = (1-\tau) \Pr\left(X_{n-1} \in (q_{n-1,k-1}, q_{n-1,k}]\right). \end{aligned} \tag{2.4}$$

If $q_{n,2k-1} = q_{n,2k}$:

$$\Pr\left(X_n \in (q_{n,2k-2}, q_{n,2k-1}]\right) = \Pr\left(X_{n-1} \in (q_{n-1,k-1}, q_{n-1,k}]\right). \tag{2.5}$$

Note that, by construction, if G_n is the cumulative distribution function (cdf) of X_n, we have that, for all $N \leq n$, $k = 1, \ldots, 2^{N-1}$

$$B^{\tau}_{G_n}\left((q_{N,k}, q_{N,k+1}]\right) = q_{N+1,2k+1}.$$

It remains to prove that X_n converges to a r.v. X that has the correct quantiles. This follows because, by construction, for all $N \geq 1$ and for all $k = 1, \ldots 2^N$, we clearly have that

$$G_N(q_{N,k}) = \Pr\left\{X_N \leq q_{N,k}\right\} = G_{N+1}(q_{N,k}) = \ldots = G_n(q_{N,k})$$

for every $n \geq N$. Thus, the sequence of distribution functions $\{G_n\}_{n \geq 1}$, admits a limit G that, for every $q \in \mathcal{Q}$, assigns a probability $G(q) = \lim_n G_n(q)$ and X is the associated random variable. □

The simple counterexample in Example 1 is sufficient to show that the definition of G is not necessarily unique for a given discrete set $\mathcal{Q}$. And in fact, if $\mathcal{Q}$ is generated by an arbitrary discrete distribution G_0, we might have that $G \neq G_0$. The construction of G given by (2.4) and (2.5) is however the unique G with support in $\mathcal{Q}$ that guarantees the following recursive property: for every n, k such that $q_{n,2k-1} \neq q_{n-1,k}$: $G(q_{n,2k-1}) = \tau G(q_{n-1,k}) + (1-\tau) G(q_{n-1,k-1})$.

Example 3. *Let us consider the SQA generated in Example 1 by the distribution, with support* $\mathcal{X} = \{x_1, x_2, x_3, x_4, x_5\}$ *and respective probability masses* $\mathbf{p} = \{0.1, 0.2, 0.4, 0.2, 0.1\}$*, and with* $\tau = 0.1$*. Using Theorem 2.3 we have*

X_1 *has support* $\{x_1, x_5\}$ *with probability mass* $\mathbf{p}_1 = \{\tau, 1-\tau\}$
$X_2 \in \{x_1, x_2, x_5\}$ *with probability mass* $\mathbf{p}_2 = \{\tau, \tau(1-\tau), (1-\tau)^2\}$
$X_3 \in \{x_1, x_2, x_3, x_5\}$ *with probability mass* $\mathbf{p}_3 = \{\tau, \tau(1-\tau), \tau(1-\tau)^2, (1-\tau)^3\}$
$X_4 \in \{x_1, x_2, x_3, x_4, x_5\}$ *with probability mass* $\mathbf{p}_4 = \{\tau, \tau(1-\tau), \tau(1-\tau)^2, \tau(1-\tau)^3, (1-\tau)^4\}$
X_n *is equal to* X_4 *for* $n \geq 5$

Indeed for $n \geq 5$ *it holds always* $q_{n,2k-1} = q_{n,2k}$ *hence, from (2.5), we have*

$$\Pr\left(X_n \in (q_{n,2k-2}, q_{n,2k-1}]\right) = \Pr\left(X_{n-1} \in (q_{n-1,k-1}, q_{n-1,k}]\right).$$

Thus, X_n *has a limiting distribution* X*,*

$$\mathbf{p}_\infty = \mathbf{p}_4 = \{\tau, \tau(1-\tau), \tau(1-\tau)^2, \tau(1-\tau)^3, (1-\tau)^4\},$$

that is different from $\mathbf{p} = \{0.1, 0.2, 0.4, 0.2, 0.1\}$*, used to generate the SQA.*

3. SQA random probability measures

In this section, we define the random SQA distributions and derive some of their properties. First we introduce a random SQA, to be used to generate an RPM. To construct a random SQA distribution, we select two distributions λ_1 and λ. Under this definition we show in Section 3.1 some properties including full support under the weak topology. Then in Section 3.2 we generalize random SQA with λ changing with n, denoted $\lambda_n, n \geq 1$, and under specific hypotheses we show that, in this case, SQA-RPM has full support under the Kullback-Lieber topology and it is differentiable.

The procedure follows closely the ideas in [17] and consists of the following steps: (i) choose distributions λ_1 and λ; (ii) generate the first quantile of level τ according to λ_1 and proceed with the other terms of the SQA sequence by extracting from λ; (iii) use Theorem 2.1 to obtain the distribution G.

Without loss of generality, we consider the generation of random distributions with support in $[0, 1]$. The extension to distributions with unbounded support will be considered in Section 4.1. Let λ_1 and λ be two probability measures with support $[0, 1]$ and $[0, 1)$ respectively.

Let $\{X_{n,2k-1}\}_{n,k}$, $n \geq 1$, $k = 1, \dots, 2^{n-1}$ be a triangular array of independent random variables, with $X_{1,1} \sim \lambda_1$ and $X_{n,2k-1} \sim \lambda$, for all $n \geq 2$. We build a random SQA(τ), denoted by $\mathcal{Q} = \{q_{n,k}\}$, by the following recursive procedure:

$$\begin{aligned} q_{1,1} &= X_{1,1} \\ q_{n,2k} &= q_{n-1,k}, \quad n \geq 1,\ k = 1, \dots, 2^{n-1} \\ q_{n,2k-1} &= \begin{cases} q_{n-1,k-1} + X_{n,2k-1}(q_{n-1,k} - q_{n-1,k-1}) & \text{if} \begin{cases} \text{k odd and (3.2) holds} \\ \text{k even and (3.3) holds} \end{cases} \\ q_{n-1,k-1} & \text{otherwise} \end{cases} \\ q_{n,0} &= 0 \quad \text{for all } n \geq 1 \\ q_{n,2^n} &= 1 \quad \text{for all } n \geq 1 \end{aligned} \tag{3.1}$$

where

$$q_{n-1,k-1} < q_{n-1,k} < q_{n-1,k+1} \text{ and } \min(X_{n,2k-1}, X_{n,2k+1}) > 0 \tag{3.2}$$

$$q_{n-1,k-2} < q_{n-1,k-1} < q_{n-1,k} \text{ and } \min(X_{n,2k-3}, X_{n,2k-1}) > 0. \tag{3.3}$$

Let us introduce some notation. We endow the space of triangular arrays $\mathfrak{A} = [0, 1] \times [0, 1]^3 \times \cdots \times [0, 1]^{2^n - 1} \times \cdots$ with the product topology and let $\mathcal{A}$ be the subset of $\mathfrak{A}$ satisfying the conditions *(i)–(iii)* in Theorem 2.3. Let $P_{\lambda_1,\lambda}$ be the probability distribution of $\mathcal{Q}$ on $\mathcal{A}$ and let T be the mapping described in Theorem 2.3, that transforms $\mathcal{Q}$ into a probability distribution on $[0, 1]$. Then, T is Borel-measurable, given the weak topology, and $T : (\mathcal{A}, P_{\lambda_1,\lambda}) \mapsto (\mathcal{P}([0, 1]), B_{\lambda_1,\lambda})$, where we define $B_{\lambda_1,\lambda} := P_{\lambda_1,\lambda} \circ T^{-1}$.

Theorem 3.1. *The distribution of the τ-quantile of the random probability measure generated from the SQA sequence of level τ is λ_1:*

$$B_{\lambda_1,\lambda} \left\{G : G^{-1}(\tau) \leq z\right\} = \lambda_1([0, z]).$$

Proof. The proof follows from the fact that, by construction, conditional on a given $\mathcal{Q} = \mathbf{q}$, $G = T(\mathbf{q})$ (where G is the limit of the sequence of probabilities G_n described in Theorem 2.3) satisfies $G(q_{1,1}) = \tau$, and thus $G^{-1}(\tau) = q_{1,1}$. Since $q_{1,1} \sim \lambda_1$, we immediately have

$$B_{\lambda_1,\lambda}\left\{G : G^{-1}(\tau) \leq z\right\} = P_{\lambda_1,\lambda}\left\{\mathcal{Q} \in \mathcal{A} : q_{1,1} \leq z\right\} = \Pr\left(q_{1,1} \leq z\right) = \lambda_1([0, z]).$$

□

Example 4. *It is straightforward to see that, if $\lambda_1 = \lambda$ and assigns probability 1 to $\{\tau\}$, then the SQA, SBA and Dubins and Freedman's RPMs ([11]) coincide, since they all assign probability 1 to the uniform distribution $U(0, 1)$.*

3.1. Properties of SQA random distributions: λ fixed

In this section we study some properties of the SQA random probability measures described in the previous subsection.

Theorem 3.2. *$B_{\lambda_1,\lambda}$-almost all SQA-RPMs distributions are continuous if and only if $\lambda_1(\{0, 1\}) = \lambda(\{0\}) = 0$.*

Proof. Suppose that $\lambda_1(\{0, 1\}) = \eta > 0$. Then, in view of Theorem 3.1,

$$\begin{aligned}
\Pr\{G \text{ has at least one jump}\} &\geq B_{\lambda_1,\lambda}\left\{G : (G(\{0\}) = \tau) \cup (G(\{1\}) = 1 - \tau)\right\} \\
&= \Pr\left\{\omega : (q_{1,1}(\omega) = 0) \cup (q_{1,1}(\omega) = q_{1,2}(\omega) = 1)\right\} \\
&= P_{\lambda_1,\lambda}\left\{(q_{1,1} = 0) \cup (q_{1,1} = q_{1,2} = 1)\right\} \\
&= \Pr\left(X_{1,1} = 0\right) + \Pr\left(X_{1,1} = 1\right) = \lambda_1(\{0, 1\}) = \eta.
\end{aligned}$$

Now suppose that $\lambda(\{0\}) = \eta > 0$. Then, each $X_{n,k} = 0$ $(n > 1)$ with probability η. Let $\omega \in \Omega$ be such that $X_{2,2k-1}(\omega) = 0$ for some $k = 1, 2$ (without loss of generality, let us set $k = 1$), while $X_{1,1} > 0$. Then, from (3.1), we have that $q_{2,1}(\omega) = q_{1,1}(\omega) = q_{2,2}(\omega)$ and consequently, also $q_{3,2} = q_{3,3} = q_{3,4} = q_{1,1}$. In general, $q_{n,2^{n-2}} = \cdots = q_{n,2^{n-1}} = q_{1,1}$ and

$$G_n\left((q_{n,2^{n-2}-1}, q_{n,2^{n-1}}]\right) = G_1(q_{1,1}) - G_n(q_{n,2^{n-2}-1}) \geq \tau - \tau^2 + \tau^2(1 - \tau)^{n-2} \overset{n\to\infty}{\longrightarrow} \tau(1 - \tau).$$

Thus, the distribution G has at least one point with mass probability greater than or equal to $\tau(1 - \tau)$ at $q_{1,1}$[1].

Following the same argument for arbitrary n, k, with $n < \infty$, it comes that, if $X_{n,2k-1}(\omega) = 0$, then the sequence $\mathcal{Q}$ generated by (3.1) has a tie at $q_{n,2k-1}$

[1]In fact, $G_n(q_{n,2^{n-2}-1}) = \tau^2 - \tau^2(1 - \tau)^{n-2}$ if all $X_{m,k} > 0$ for $m \leq n$ and $k \leq 2^{n-2} - 1$. This is an immediate consequence of the fact that, when there are no ties, $G_n(q_{n,2^{n-2}-1}) = G_n(q_{n,2^{n-2}}) - G_n((q_{n,2^{n-2}-1}, q_{n,2^{n-2}}]) = G(q_{2,1}) - G_n((q_{n,2^{n-2}-1}, q_{n,2^{n-2}}]) = \tau^2 - \tau^2(1 - \tau)^{n-2}$. This last identity can be found in Section 3.3 for arbitrary values of n and k.

and the distribution $G = \lim G_n$ has (at least) a jump in that point. Since $\Pr(\omega : X_{n,2k-1}(\omega) = 0) = \lambda(\{0\}) = \eta$, this implies that $\Pr\{G \text{ is discrete}\} \geq \eta$.

To prove the inverse implication, we show that

$$\int \left\{ \int_{\{x,y:x=y\}} d(G \times G)(x,y) \right\} dB_{\lambda_1,\lambda}(G) = 0.$$

This immediately implies the continuity of G because the probability that two independent r.v. with the same distribution G coincide is zero only if G is continuous (see [24]). We prove the above identity by considering the integral

$$E_n = \int \left(\sum_{k=1}^{2^n} (G_n(q_{n,k}) - G_n(q_{n,k-1}))^2 \right) dB_{\lambda_1,\lambda}(G_n),$$

where $\{G_n\}$ is the sequence of probability distribution functions (pdf's) of the sequence of random variables $\{X_n\}$ defined in the proof of Theorem 2.3 and showing that $E_n \to 0$.

For all ω such that $q_{n,2k-1}(\omega) \neq q_{n,k-1}(\omega)$ for all n, k, we have (see Theorem 2.3):

$$\begin{aligned}
&\sum_k (G_n(q_{n,k}) - G_n(q_{n,k-1}))^2 \\
&\quad = \sum_{k=1}^{2^{n-1}} (G_n(q_{n,2k}) - G_n(q_{n,2k-1}))^2 + \sum_{k=1}^{2^{n-1}} (G_n(q_{n,2k-1}) - G_n(q_{n,2k-2}))^2 \\
&\quad = \sum_{k=1}^{2^{n-1}} \big[(1-\tau)^2 (G_n(q_{n-1,k}) - G_n(q_{n-1,k-1}))^2 \\
&\qquad + \tau^2 (G_n(q_{n-1,k}) - G_n(q_{n-1,k-1}))^2\big] \\
&\quad = \sum_k [\tau^2 + (1-\tau)^2](G_n(q_{n-1,k}) - G_n(q_{n-1,k-1}))^2 \\
&\quad \cdots = \sum_{k=1}^{2} [\tau^2 + (1-\tau)^2]^{n-1} (G_n(q_{1,k}) - G_n(q_{1,k-1}))^2 \\
&\quad = [\tau^2 + (1-\tau)^2]^n
\end{aligned}$$

where we omitted the ω for convenience. Moreover, since $\lambda(\{0\}) = \lambda_1(\{0,1\}) = 0$ for all n, k, we have $X_{n,2k-1} \in (0,1)$, where $X_{n,2k-1}$ is from (3.1) and thus, for all $n \geq 1$,

$$\begin{aligned}
&P_{\lambda_1,\lambda}(\mathcal{Q} : q_{n,2k-1} = q_{n,k-1}, \text{ for some } k \leq 2^n) \\
&\qquad = \Pr\{\omega : q_{n,2k-1}(\omega) = q_{n,k-1}(\omega), \text{ for some } k \leq 2^n\} \\
&\qquad = \Pr\{X_{n,2k-1} = 0 \text{ for some } k \leq 2^n\} = 0.
\end{aligned}$$

Thus,

$$\begin{aligned} E_n &= \int_{\{\mathcal{Q}: q_{n,2k-1}\neq q_{n,k-1},\ k\leq 2^n\}} \left(\sum_{k=1}^{2^n} (G_n(q_{n,k}) - G_n(q_{n,k-1}))^2 \right) dP_{\lambda_1,\lambda}(G_n) \\ &\quad + \int_{\left\{ \substack{\mathcal{Q}\,:\, q_{n,2k-1} = q_{n,k-1}, \\ \text{for some } k\leq 2^n} \right\}} \left(\sum_{k=1}^{2^n} (G_n(q_{n,k}) - G_n(q_{n,k-1}))^2 \right) dP_{\lambda_1,\lambda}(G_n) \\ &\leq [\tau^2 + (1-\tau)^2]^n \Pr\{\omega : q_{n,2k-1}(\omega) \neq q_{n,k-1}(\omega),\ \forall\, k\leq 2^n\} \\ &\qquad + P_{\lambda_1,\lambda}(\mathcal{Q} : q_{n,2k-1} = q_{n,k-1},\ \text{for some } k \leq 2^n) \\ &= [\tau^2 + (1-\tau)^2]^n \to_n 0. \end{aligned}$$

□

Theorem 3.3. *$B_{\lambda_1,\lambda}$-almost all SQA-RPM are discrete if and only if $\lambda(\{0\}) = 1$ or $\lambda_1(\{0,1\}) = 1$.*

Proof. Let us define the mean sum of jumps generated by the SQA-RPMs:

$$J = \int_{\mathcal{P}(0,1)} \Delta(G) dB_{(\lambda_1,\lambda)}(G) = \int_0^1 J(m) d\lambda_1(m)$$

where

$$J(m) = \int_{\mathcal{P}(0,1)} \Delta(G) dB_{(\delta_m,\lambda)}(G)$$

and $\Delta(G)$ denotes the sum of the jumps of a distribution function G. It is easy to see that, if $m = 0$ or $m = 1$, then $J(m) = 1$. We prove that, unless $\lambda(\{0\}) = 1$, $J(m) < 1$ for all $m \in (0,1)$.

For every n, k, let us define the random variable $Z_{n,k} = \min\{X_{n,4k-3}, X_{n,4k-1}\}$. Let $p = \lambda(\{0\})$ and let $R = p + p(1-p) = Pr\{Z_{n,k} = 0\}$. Then, for all $n \geq 2$, we define the sequences

$$\begin{aligned} E_{n,j} &= \{\text{exactly } j \text{ new ties occur at stage } n \text{ of the extraction of the quantiles}\} \\ &= \{\exists\ k_1, .., k_j < 2^{n-2} \text{ s.t. } Z_{n,k_i} = 0\} \end{aligned}$$

and

$$L_{n,j} = \{\text{total length of the } j \text{ jumps at step } n\}.$$

In particular

$$E_{n,0} = \{\text{no ties at stage } n\} = \left\{ \min_{k\leq 2^{n-2}} Z_{n,k} > 0 \right\} = \left\{ \min_{k\leq 2^{n-1}} X_{n,2k-1} > 0 \right\},$$

has probability $P(E_{n,0}) = (1-R)^{2^{n-2}} = (1-p)^{2^{n-1}}$.

Given these definitions, we can clearly decompose the mean jumps, for all m, to the following

$$J(m) = \sum_{h=2}^{\infty} \sum_{j=1}^{2^{h-2}} \Pr\left(E_{h,j} \cap \{E_{l,0}, l < h\}\right) L_{h,j}.$$

Note that $L_{n,j}$ is a random variable that depends wholly on which of the 2^{n-1} possible variables $Z_{n,k}$ is equal to 0. Since all $\{X_{n,2k-1}\}_{k\leq 2^{k-1}}$ are i.i.d., also the $Z_{n,k}$ are i.i.d. and therefore exchangeable. This also implies that in computing $J(m)$ we can replace $L_{n,j}$ to its average value.

We know (Lemma 3.1) that the mean value of $L_{n,j}$, $n \geq 2$, is equal to $j/2^{n-2}$. Then, we can write

$$\begin{aligned} J(m) &= \sum_{h=2}^{\infty} \sum_{j=1}^{2^{h-2}} \Pr(E_{h,j}) \Pr\{E_{l,0}, l < h\} \frac{j}{2^{h-2}} \\ &= \sum_{h=2}^{\infty} \sum_{j=1}^{2^{h-2}} \Pr(E_{h,j}) \prod_{l=2}^{h-1} (1-p)^{2^{l-1}} \frac{j}{2^{h-2}} \\ &= \sum_{h=2}^{\infty} (1-p)^{2^{h-1}-2} \sum_{j=1}^{2^{h-2}} \Pr(E_{h,j}) \frac{j}{2^{h-2}} \end{aligned}$$

For the term $\Pr(E_{h,j})$, we note that

$$\Pr(E_{h,j}) = \binom{2^{h-2}}{j} P(Z=0)^j [1 - P(Z=0)]^{2^{h-2}-j}$$

where $Z \sim Z_{h,k}$, $k \leq 2^{h-2}$. Thus,

$$\begin{aligned} \sum_{j=1}^{2^{h-2}} \Pr(E_{h,j}) \frac{j}{2^{h-2}} &= \sum_{j=1}^{2^{h-2}} \frac{j}{2^{h-2}} \binom{2^{h-2}}{j} R^j (1-R)^{2^{h-2}-j} \\ &= R \sum_{j=1}^{2^{h-2}} \binom{2^{h-2}-1}{j-1} R^{j-1} (1-R)^{2^{h-2}-j} = R. \end{aligned}$$

Noting that $1 - R = (1-p)^2$, we can write, for any $m \in (0,1)$,

$$J(m) = \sum_{n=2}^{\infty} R(1-R)^{2^{n-2}-1} = R + R(1-R) + R(1-R)^3 + R(1-R)^7 + \cdots.$$

The above equation implies that $J(m) \leq \sum_{n=0}^{\infty} R(1-R)^n = 1$ and the identity occurs only if $R = 1$, that is, when $p = 1$.

Thus, given $\lambda_1(\{0,1\}) < 1$, the SQA-RPM are $B_{\lambda_1,\lambda}$-a.s. discrete if and only if $\lambda(\{0\}) = 1$. If instead $\lambda(\{0\}) < 1$ there is a set of SQA-RPM with strictly positive $B_{\lambda_1,\lambda}$-measure that is not discrete. □

Note that Theorem 3.3 differs from the analogous result in [17] (Theorem 3.6), because in our case by choosing a distribution λ s.t. $\lambda(\{0\}) > 0$, we can eventually obtain both continuous and discrete random probability measures, with positive $B_{\lambda_1,\lambda}$-probability.

Lemma 3.1. *For all $n \geq 2$,*

$$\mathbb{E}L_{n,k} = \frac{\sum L_{n,k}}{\binom{2^{n-2}}{k}} = \frac{k}{2^{n-2}},$$

where the sum in the above equation is over all possible (not distinct) $\binom{2^{n-2}}{k}$ *values that can occur for* $L_{n,k}$.

Proof. Let $n = 2$, $L_{2,1} = 1$, because in this case there can be only one tie at $q_{2,1} = q_{2,3} = m$, and the total length of the jumps is $\tau + (1-\tau) = 1$. If $n = 3$, and there is only one tie, it can be in $q_{3,1} = q_{3,2} = q_{3,3}$ (if $X_{3,1} = 0$ or $X_{3,3} = 0$) or in $q_{3,5} = q_{3,6} = q_{3,7}$ (if $X_{3,5} = 0$ or $X_{3,7} = 0$): the sum of the jumps before and after the tie will be $\tau^2 + \tau(1-\tau) = \tau$, in the first case, and $\tau(1-\tau) + (1-\tau)^2 = 1-\tau$ in the second case. Because of all $X_{n,j}$ are identically distributed, these two values have the same weight, and the average value of $L_{3,1}$ is therefore $\mathbb{E}L_{3,1} = 1/2$. If at stage 3 there are two new ties, then all $X_{3,1} = X_{3,3} = X_{3,5} = X_{3,7} = 0$. In this case, the total jumps $L_{3,2} = 1$. In general, it is straightforward that $L_{n,2^{n-1}} = 1$ for all n, and, using a binary tree representation of $L_{n,1}$ (see Figure 1), it is easy to see that there are 2^{n-2} possible *elementary* values for $L_{n,1}$, among which we can distinguish the $n-1$ distinct values $\tau^i(1-\tau)^{n-2-i}$, occurring $\binom{n-2}{i}$ times each. Therefore, $\mathbb{E}L_{n,1} = \sum_i \binom{n-2}{i}\tau^i(1-\tau)^{n-2-i}/2^{n-2} = 1/2^{n-2}$.

The lengths of $L_{n,2}$ are obtained by summing the jumps corresponding to the two ties that have occurred. Therefore, since the weighted sum of possible values of $L_{n,1}$ is one, we have that, $\mathbb{E}L_{n,2}$ is obtained by considering the sum of all possible values $L_{n,1} + L'_{n,1}$, where, given $L_{n,1}$, $L'_{n,1}$ is constrained to be one of the possible $2^{n-2} - 1$ values left (some of which might coincide numerically with $L_{n,1}$). Since the sum of all possible values of $L'_{n,1}$ satisfying the constraint is $1 - L_{n,1}$, we get

$$
\begin{aligned}
2\mathbb{E}L_{n,2} &= \sum_{L_{n,1}} \frac{1}{\#\{\text{choices for values of } L_{n,2}\}} (L_{n,1}(2^{n-2} - 1) + 1 - L_{n,1}) \\
&= \frac{2(2^{n-2} - 1)}{\binom{2^{n-2}}{2}} = \frac{4}{2^{n-2}}
\end{aligned}
$$

and thus $\mathbb{E}L_{n,2} = \frac{2}{2^{n-2}}$.[2]

Now we assume that $\mathbb{E}L_{n,j} = \sum_{L_{n,j}} \frac{L_{n,j}}{\binom{2^{n-2}}{j}} = \frac{j}{2^{n-2}}$ and prove by induction that $L_{n,j+1} = \frac{j+1}{2^{n-2}}$. By generalizing the argument used for $L_{n,2}$, we can write $(j+1)\sum_{L_{n,j+1}} L_{n,j+1} = \sum_{L_{n,j}} \sum_{L'_{n,1} \neq L_{n,j}} (L_{n,j} + L'_{n,1})$, where the second sum is over all possible values of $L_{n,1}$ that are not included in $L_{n,j}$. Thus, we clearly have that:

$$
\begin{aligned}
(j+1)\mathbb{E}L_{n,j+1} &= \frac{\sum_{L_{n,j}} L_{n,j}(2^{n-2} - j) + 1 - L_{n,j}}{\binom{2^{n-2}}{j+1}} \\
&= (\text{after some computation}) = \frac{(j+1)^2}{2^{n-2}}. \qquad \square
\end{aligned}
$$

[2]Note that the above formula counts couples $(L_{n,1}, L'_{n,1})$ twice, therefore we need to correct to obtain the mean by dividing by 2. A similar correction is used for the computation of $\mathbb{E}L_{n,j+1}$

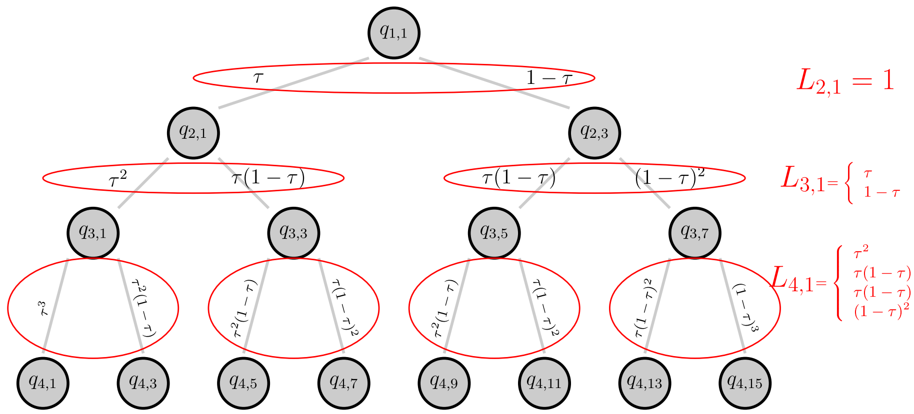

FIG 1. *Set of possible total heights of the jumps $L_{n,k}$, when there is a single ($k=1$) new tie occurring at the nth step of the construction of the sequence of the quantiles, $n \geq 2$. The values of $L_{n,1}$ are given by the sum of the probabilities in each ellipse.*

Theorem 3.4. *If λ_1 and λ have full support on $[0,1)$ then $B_{\lambda_1,\lambda}$ has full support on $\mathcal{P}([0,1])$ with respect to the weak topology in $\mathcal{P}([0,1])$.*

Proof. $B_{\lambda_1,\lambda}$ has full support if any non empty set of $\mathcal{P}([0,1])$ has a positive $B_{\lambda_1,\lambda}$-measure. This holds if each set of the base of the weak topology of $\mathcal{P}([0,1])$ has positive $B_{\lambda_1,\lambda}$-measure. According to Billingsley (1968) [3] a basis consists of sets of the form

$$\{\sigma \in \mathcal{P}([0,1]) : \sigma(O_i) > \sigma_0(O_i) - \varepsilon_i \quad i = 1,..,m\},$$

where O_i is an open set of $[0,1]$, $\varepsilon_i > 0$ and σ_0 is a given measure in $\mathcal{P}([0,1])$. Since each open set in $[0,1]$ can be written as a disjoint union of r open sets $O_{i,j}$, we can write

$$E = \{\sigma \in \mathcal{P}([0,1]) : \sigma(\cup_j^r O_{i,j}) > \sigma_0(\cup_j^r O_{i,j}) - \varepsilon_i \quad i = 1,..,m\}.$$

We define the following subset of E:

$$C = \cap_j \{\sigma \in \mathcal{P}([0,1]) : \sigma(O_{i,j}) > \sigma_0(O_{i,j}) - \frac{\varepsilon}{r} \quad i = 1,..,m\},$$

with $\varepsilon = \min_i \varepsilon_i$. If we prove that $B_{\lambda_1,\lambda}(C) > 0$, since $C \subset E$ it follows that $B_{\lambda_1,\lambda}(E) > 0$.

Let G_σ and $q^\sigma_{n,k}$ denote the distribution function of σ and the SQA of G_σ, respectively. Let G_{σ_0} and $q^{\sigma_0}_{n,k}$ the analogous for σ_0. Since $q^{\sigma_0}_{n,k}$ is dense in the support of σ_0, let $q^{\sigma_0}_{n_{ij},k_{ij}}$ and $q^{\sigma_0}_{n_{ij},l_{ij}}$ be such that $(q^{\sigma_0}_{n_{ij},k_{ij}}, q^{\sigma_0}_{n_{ij},l_{ij}}] \subset O_{ij}$ and

$$\sigma_0\left((q^{\sigma_0}_{n_{ij},k_{ij}}, q^{\sigma_0}_{n_{ij},l_{ij}}]\right) > \sigma_0(O_{ij}) - \frac{2\varepsilon}{r}.$$

Note that, from the definition of $q^{\sigma_0}_{n,k}$ and $q^\sigma_{n,k}$, we have $G_{\sigma_0}(q^{\sigma_0}_{n,k}) = G_\sigma(q^\sigma_{n,k})$, for all n,k.

Let $N = \max_{ij} n_{ij}$ and consider the set

$$D_\delta = \{\sigma \in \mathcal{P}([0,1]) : |q^{\sigma}_{N,k} - q^{\sigma_0}_{N,k}| < \delta, \quad \text{for all } k = 1, ...2^N - 1\},$$

if $\sigma \in D_\delta$ then, for all i, j we have that

$$\left| q^{\sigma}_{n_{ij},l_{ij}} - q^{\sigma_0}_{n_{ij},l_{ij}} \right| < \delta \qquad \text{and} \qquad \left| q^{\sigma}_{n_{ij},k_{ij}} - q^{\sigma_0}_{n_{ij},k_{ij}} \right| < \delta.$$

This implies

$$q^{\sigma}_{n_{ij},l_{ij}} < \delta + q^{\sigma_0}_{n_{ij},l_{ij}} \qquad \text{and} \qquad q^{\sigma}_{n_{ij},k_{ij}} > q^{\sigma_0}_{n_{ij},k_{ij}} - \delta \quad \text{for all } i, j;$$

thus, we have, for all i, j,

$$G_\sigma(q^{\sigma}_{n_{ij},l_{ij}}) < G_\sigma(q^{\sigma_0}_{n_{ij},l_{ij}} + \delta) \qquad \text{and} \quad G_\sigma(q^{\sigma}_{n_{ij},k_{ij}}) > G_\sigma(q^{\sigma_0}_{n_{ij},k_{ij}} - \delta),$$

that is, for all i, j,

$$\begin{aligned} \sigma\left((q^{\sigma_0}_{n_{ij},k_{ij}} - \delta, q^{\sigma_0}_{n_{ij},l_{ij}} + \delta] \right) &> \sigma\left((q^{\sigma}_{n_{ij},k_{ij}}, q^{\sigma}_{n_{ij},l_{ij}}] \right) \\ &= \sigma_0\left((q^{\sigma_0}_{n_{ij},k_{ij}}, q^{\sigma_0}_{n_{ij},l_{ij}}] \right) > \sigma(O_{ij}) - \frac{2\varepsilon}{r}. \end{aligned} \tag{3.4}$$

The set $(q^{\sigma_0}_{n_{ij},k_{ij}} - \delta, q^{\sigma_0}_{n_{ij},l_{ij}} + \delta]$ is larger than $(q^{\sigma_0}_{n_{ij},k_{ij}}, q^{\sigma_0}_{n_{ij},l_{ij}}]$. We can find a δ small enough such that the left hand side of (3.4) is bounded by:

$$\sigma\left((q^{\sigma_0}_{n_{ij},k_{ij}} - \delta, q^{\sigma_0}_{n_{ij},l_{ij}} + \delta] \right) < \sigma(O_{ij}) - \frac{\varepsilon}{r}.$$

For this δ, we have that, if $\sigma \in D_\delta$, then also

$$\sigma(O_{ij}) - \frac{\varepsilon}{r} > \sigma_0(O_{ij}) - \frac{2\varepsilon}{r}, \quad \forall i, j,$$

that is equivalent to $\sigma \in C$.

Given that λ_1 and λ have full support on $[0, 1)$, $B_{\lambda_1,\lambda}(D_\delta) > 0$ and since we showed that $D_\delta \subset C$ it follows $B_{\lambda_1,\lambda}(C) > 0$, this ends the proof. □

3.2. Properties of SQA-RPM when λ_n changes with n

In this section we consider the more general definition of SQA-RPM obtained when each $X_{n,2k-1}$ in equation (3.1) is generated independently from a distribution λ_n that is allowed to change with n. We point out that all the results of the previous subsection are still valid, with little adjustment, if the SQA-RPM is generated by the sequence $\lambda_n, n \geq 1$. For example, if $\lambda_1(\{0, 1\}) = 0$ and all measures λ_n assign mass 0 to $\{0\}$, then according to Theorem 3.2, almost all SQA-RPMs are continuous.

We further show that when the measures λ_n are allowed to change with n the SQA-RPM can have stronger properties, such as differentiability and full support in stronger topologies.

We assume that $\lambda_1(\{0,1\}) = 0$ and that the SQA sequence is defined through a sequence of generating measures $\{\lambda_n,\ n \geq 1\}$, such that $X_{n,2k-1} \sim \lambda_n$ for all k and n with $\lambda_n(\{0\}) = 0$. The probability measure in the space of SQA-RPM is now denoted by $B_{\lambda_n, n\geq 1}$, and, repeating the arguments of Theorem 3.2, it can be seen that it contains only continuous distribution functions and, because of Lemma 2.2*(ii)*, the set $\{q_{n,k}\}$ is dense in $(0,1)$. Under this condition, the main equations in (3.1) simplify to

$$q_{n,2k-1} = q_{n-1,k-1} + X_{n,2k-1}(q_{n-1,k} - q_{n-1,k-1}).$$

Let us denote by $\mathbf{1}\{A\}$ the indicator function, $\mathbf{1}\{A\} = 1$ if A is true and is $\mathbf{1}\{A\} = 0$ otherwise. For all $u \in [0,1]$ and $n > 1$, we have the following representation of G_n.

$$G_n(u) = \sum_{k=1}^{2^n}[G_n(q_{n,k}) - G_n(q_{n,k-1})]\mathbf{1}\{u \geq q_{n,k}\} = \sum_{k=1}^{2^n} \tau^{z_{n,k}}(1-\tau)^{n-z_{n,k}}\mathbf{1}\{u \geq q_{n,k}\},$$

where $z_{n,k}$ is the number of zeros in the n-cyphers binary representation of $k-1$. Hence we have $G(u) = \lim_{n\to\infty} G_n(u)$.

Every G_n depends in a complex way on the distributions λ_n, through each $q_{n,k}$. Indeed, it is possible to see that all $q_{n,k}$ can be written as a sum of products of $X_{m,2j-1}$:

$$q_{n,k} = \sum_{h=1}^{k}\prod_{m=1}^{n} X_{m,2j_{m,h}-1}^{1-d_{m,h}}(1 - X_{m,2j_{m,h}-1})^{d_{m,h}} \tag{3.5}$$

where $d_{m,h}$ and $j_{m,h}$ both depend on m and h.

Specifically, $d_{m,h} = 1$ if the m-th cypher (from the right) of the binary representation of the number $h-1$ ($h \leq 2^m$) is a 1 (and $d_{m,h} = 0$ otherwise), while $j_{m,h}$ is defined recursively by:

$$j_{m,h} = 2j_{m-1,h} - 1 + 2d_{m-1,h},\ m \geq 1.$$

Thus, each summand in $q_{n,k}$ depends on products of independent variables $X_{m,2j-1}$, but summands are not independent because they can have terms in common.

Theorem 3.5. *Assume the SQA follows (3.1) and all the variables $X_{n,2k-1} \sim \lambda_n$ are independent for all $k \leq 2^n$ and $n \geq 1$, with mean τ_n and variance σ_n^2. Let $\inf_n \tau_n > t > 0$, and for some $\delta > 0$, $\tau_n = \tau + O(n^{-1-\delta})$. If moreover $\sum_n \sigma_n^2 < \infty$, then $B_{\lambda_1,\lambda}$-almost all SQA-RPMs are differentiable in $u \in (0,1)$.*

Proof. Let us define

$$g_n(u) = \sum_k \mathbf{1}\{u \in (q_{n,k-1}, q_{n,k})\}\frac{G_n(q_{n,k}) - G_n(q_{n,k-1})}{q_{n,k} - q_{n,k-1}}.$$

Using (3.5), we have:

$$g_n(u) = g_{n-1}(u)\sum_k \mathbf{1}\{u \in (q_{n,k-1}, q_{n,k})\}\frac{\tau^{1-d_{n,k}}(1-\tau)^{d_{n,k}}}{X_{n,2j_{n,k}-1}^{1-d_{n,k}}(1 - X_{n,2j_{n,k}-1})^{d_{n,k}}}.$$

We have that the RPM G is differentiable if the limit $\lim_n g_n(u)$ exists for all u. Now, let us define $\tilde{g}_n(u)$ as

$$\tilde{g}_n(u) = \tilde{g}_{n-1}(u) \sum_k \mathbf{1}\{u \in (q_{n,k-1}, q_{n,k})\} \frac{1}{Z_n^{1-d_{n,k}} W_n^{d_{n,k}}},$$

where $Z_n = X_{n,2j_{n,k}-1}/\tau_n$ and $W_n = (1 - X_{n,2j_{n,k}-1})/(1-\tau_n)$. Clearly, Z_n and W_n are independent on $\tilde{g}_{n-1}$ and both have mean 1.

Further, if all $\tau_n = \tau$, then $\tilde{g}_n = g_n$ for all n, while

$$g_n(u) = \tilde{g}_n(u) \times \prod_{m \le n} \frac{\tau^{1-d_{m,k}}(1-\tau)^{d_{m,k}}}{\tau_n^{1-d_{m,k}}(1-\tau_n)^{d_{m,k}}},$$

otherwise.

We define the process $\tilde{s}_n(u) = 1/\tilde{g}_n(u)$ and show that it is a martingale; in fact,

$$\mathbb{E}\left(\tilde{s}_n(u) \mid \tilde{s}_{n-1}\right) = \tilde{s}_{n-1}(u)\mathbb{E}(Z_n^{1-d_{n,k}} W_n^{d_{n,k}}) = \tilde{s}_{n-1}(u).$$

We can apply the martingale convergence theorem if $\mathbb{E}\tilde{s}_1(u) < \infty$. This is an immediate consequence of $\tilde{s}_1(u) = Z_1 \mathbf{1}\{u \in (0, X_{11}]\} + W_1 \mathbf{1}\{u \in (X_{11}, 1]\}$, which implies $\mathbb{E}(\tilde{s}_1(u)) \le 2$ for all n and u. Thus, the limit, $\tilde{s}(u) = \lim_n \tilde{s}_n(u)$ exists for all u.

In order to be able to define $\tilde{g}(u) = \lim_n 1/\tilde{s}_n(u)$, we need $s(u) > 0$ for all $u \in (0,1)$. Then, by writing $\bar{Z}_n = Z_n - 1$ and $\bar{W}_n = W_n - 1$, we have $\lim_n \tilde{s}_n(u) > 0$ if and only if

$$\infty > |\lim_n \log(\tilde{s}_n(u))| = |\lim_n \sum_{m \ge n} (1 - d_{n,m}) \log(1 + \bar{Z}_m) + d_{n,m} \log(1 + \bar{W}_m)|.$$

It is then enough to show that $\sum_{m \le \infty} \log(1 + |\bar{Z}_m|) < \infty$ and $\sum_{m \le \infty} \log(1 + |\bar{W}_m|) < \infty$, and we can limit ourselves to do it for the term in Z, because for W it is the same.

We invoke the two–series theorem (see [33], IV.2.2): the series $\sum_{m \le \infty} \log(1 + |\bar{Z}_m|) < \sum_m |\bar{Z}_m|$ converges if $\sum_n \mathbb{E}|\bar{Z}_n| < \infty$ and $\sum_n Var|\bar{Z}_n| < \infty$. Both conditions are satisfied if $\sum_{n=1}^{\infty} \sigma_n^2/\tau_n^2 < t^{-2} \sum_{n=1}^{\infty} \sigma_n^2 < \infty$.

Finally, to prove that

$$g(u) = \lim_n \tilde{g}_n(u) \times \prod_m \frac{\tau^{1-d_{m,k}}(1-\tau)^{d_{m,k}}}{\tau_n^{1-d_{m,k}}(1-\tau_n)^{d_{m,k}}}$$

exists, we now need to show the convergence of the series

$$\lim_n \log(g_n(u)) = \lim_n \log(\tilde{g}_n(u)) + \sum_{m \le n} \log\left(\frac{\tau^{1-d_{m,k}}(1-\tau)^{d_{m,k}}}{\tau_n^{1-d_{m,k}}(1-\tau_n)^{d_{m,k}}}\right).$$

Provided that each τ_n is bounded below by a positive constant $t > 0$, we have that the series in the above equation converges if $\sum_n |1 - \frac{\tau}{\tau_n}| < \infty$. Since

$\tau_n > t > 0$ for all t, we must have $|\tau_n - \tau| = O(1/n^{1+\delta})$, for $\delta > 0$. Finally, the summability of the sequence of variances also permits to apply Corollary in [21], implying that $G(x) = \int_0^x g(u)du$. □

From the above theorem, the only differentiable distribution generated by a SQA sequence, when $\lambda_n = \lambda$ is constant with n, is the uniform distribution, obtained for $\lambda = \delta_\tau$.

For arbitrary measures $\{\lambda_n, n \geq 1\}$, the characterization of G and of its density can't be easily used to derive a closed form expression for G and, through this, of the expected random measure $\bar{G}$. However, as [18] pointed out, *with the recent advent of simulation based inference the need for clear-cut conjugacy and analytically tractable posteriors is no longer critical*, since it deflates the importance of analytical tractability.

Theorem 3.4 proves the large support of the SQA-RPM in the weak topology; while, in applications, it is often desirable that the random probability measures have full support in a stronger sense.

As in [22] and [18], we consider the extension of Theorem 3.4 to Kullback-Leibler topology, i.e. the topology induced by relative entropy neighborhoods. We recall the definition of relative entropy (also called KL-divergence): given two probability distributions G_0, G with densities g_0, g, we have $KL(G_0, G) = \int g_0(x) \log \frac{g_0(x)}{g(x)} dx$. Note that $KL(G_0, G) = \infty$ if G_0 is not absolutely continuous w.r.t. (the probability measure associated to) G.

In this topology, neighborhoods of an arbitrary probability distribution function G_0 have the form $\{G : KL(G_0, G) < \epsilon\}$.

Theorem 3.6. *Let all the assumptions of Theorem 3.5 be satisfied and assume that each λ_n, $n \geq 2$ is such that $\sum_n \sigma_n < \infty$. Let the probability distribution G_0 have a density g_0 and a finite entropy, namely*

$$\left| \int g_0(x) \log(g_0(x)) dx \right| < \infty.$$

Then, for all ε,

$$B_{\lambda_n, n \geq 1}\left(G : KL(G_0, G) < \varepsilon\right) > 0$$

Proof. We follow the same reasoning in [22]. We first write

$$\int g_0(x) \log \frac{g_0(x)}{g(x)} dx = \int g_0(x) \log g_0(x) dx - \int g_0(x) \log g(x) dx. \tag{3.6}$$

The first integral is the entropy of g_0 and is finite in absolute value by assumption. The second term can be written as:

$$\begin{aligned} -\int g_0(x) \log g(x) dx &= \int \log\left(\lim_{n\to\infty} \frac{\nu(I_{n,k(x)})}{G(I_{n,k(x)})} \right) g_0(x) dx \\ &= \int \log\left(\lim_n \left(\prod_{m=1}^n \frac{\nu(I_{m,k_m(x)} \mid I_{m-1,k_{m-1}(x)})}{G(I_{m,k_m(x)} \mid I_{m-1,k_{m-1}(x)})} \right) \right) g_0(x) dx, \end{aligned}$$

where the indices $k_n(x)$ of the sequence $I_{n,k_n(x)} = (q_{n,k_n(x)-1}, q_{n,k_n(x)}]$ satisfy $x \in I_{n,k_n(x)}$ for all n and ν is the Lebesgue measure [3]. Then, we can write

$$
\begin{aligned}
-\int g_0(x)\log g(x)dx &= \int \sum_{m=1}^{\infty} \log\left(\frac{\nu(I_{m,k_m(x)} \mid I_{m,k_{m-1}(x)})}{G(I_{m,k_m(x)} \mid I_{m-1,k_{m-1}(x)})}\right) g_0(x)dx \\
&= \sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}} \left[\log\left(\tfrac{\nu(I_{n,2k}|I_{n-1,k})}{G(I_{n,2k}|I_{n-1,k})}\right) G_0(I_{n,2k}) + \log\left(\tfrac{\nu(I_{n,2k-1}|I_{n-1,k})}{G(I_{n,2k-1}|I_{n-1,k})}\right) G_0(I_{n,2k-1})\right] \\
&\le \sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}} \max\left\{\log\left(\tfrac{\nu(I_{n,2k}|I_{n-1,k})}{G(I_{n,2k}|I_{n-1,k})}\right), \log\left(\tfrac{\nu(I_{n,2k-1}|I_{n-1,k})}{G(I_{n,2k-1}|I_{n-1,k})}\right)\right\} G_0(I_{n-1,k}) \\
&= \sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}} \max\left\{\log\left(\frac{1-X_{n,2k-1}}{1-\tau}\right), \log\left(\frac{X_{n,2k-1}}{\tau}\right)\right\} G_0(I_{n-1,k}).
\end{aligned}
$$

In the last equality we used the fact that $\nu(I_{n,2k} \mid I_{n-1,k}) = \frac{q_{n,2k}-q_{n,2k-1}}{q_{n-1,k}-q_{n-1,k-1}} = 1 - X_{n,2k-1}$.

We consider that

$$
\max\left\{\log\left(\frac{1-x}{1-\tau}\right), \log\left(\frac{x}{\tau}\right)\right\} = \log\left(\frac{1-x}{1-\tau}\right)\mathbf{1}\{x \le \tau\} + \log\left(\frac{x}{\tau}\right)\mathbf{1}\{x > \tau\}
$$

and we take the expectations, since if its expectation converges we obtain that the second term of (3.6) is bounded:

$$
\begin{aligned}
&\mathbb{E}\left[\mathbb{E}\left(\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}} \max\left\{\log\left(\frac{1-X_{n,2k-1}}{1-\tau}\right), \log\left(\frac{X_{n,2k-1}}{\tau}\right)\right\} \mid \{q_{n-1}\}\right) G_0(I_{n-1,k})\right] \\
&= \mathbb{E}\left\{\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}} \left[\mathbb{E}\left(\log\left(\frac{1-X_{n,2k-1}}{1-\tau}\right)\mathbf{1}\{X_{n,2k-1} \le \tau\} \mid \{q_{n-1}\}\right)\right.\right. \\
&\qquad \left.\left. + \mathbb{E}\left(\log\left(\frac{X_{n,2k-1}}{\tau}\right)\mathbf{1}\{X_{n,2k-1} > \tau\} \mid \{q_{n-1}\}\right)\right] G_0(I_{n-1,k})\right\} \\
&= \mathbb{E}\left\{\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}} \lambda_n((0,\tau])\left[\mathbb{E}\left(\log\left(\frac{1-X_{n,2k-1}}{1-\tau}\right)\middle| \{q_{n-1}\}, X_{n,2k-1} \le \tau\right)\right.\right. \\
&\qquad \left.\left. + \lambda_n((\tau,1])\mathbb{E}\left(\log\left(\frac{X_{n,2k-1}}{\tau}\right)\middle| \{q_{n-1}\}, X_{n,2k-1} > \tau\right)\right] G_0(I_{n-1,k})\right\}.
\end{aligned}
$$

[3]Here and in the following whenever this causes no ambiguity in the interpretation, G and G_0 denote either the cdf's and the associated probability measure.

We apply the conditional Jensen's inequality to the concave function $\log(x)$

$$\begin{aligned}
&\mathbb{E}\left\{\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}}\lambda_n((0,\tau])\left[\mathbb{E}\left(\log\left(\frac{1-X_{n,2k-1}}{1-\tau}\right)\middle|\{q_{n-1}\},X_{n,2k-1}\le\tau\right)\right.\right.\\
&\qquad\left.\left.+\lambda_n((\tau,1])\mathbb{E}\left(\log\left(\frac{X_{n,2k-1}}{\tau}\right)\middle|\{q_{n-1}\},X_{n,2k-1}>\tau\right)\right]G_0(I_{n-1,k})\right\}\\
&\le\mathbb{E}\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}}\left\{\lambda_n((0,\tau])\log\mathbb{E}\left(\frac{1-X_{n,2k-1}}{1-\tau}\middle|\{q_{n-1}\},X_{n,2k-1}\le\tau\right)\right.\\
&\qquad\left.+(1-\lambda_n((0,\tau]))\log\mathbb{E}\left(\frac{X_{n,2k-1}}{\tau}\middle|\{q_{n-1}\},X_{n,2k-1}>\tau\right)\right\}G_0(I_{n-1,k}).
\end{aligned}$$

Applying the inequality $\log(1+x)\le x$ to $\log\left(\frac{X_{n,2k-1}}{\tau}\right)=\log\left(1+\frac{X_{n,2k-1}-\tau}{\tau}\right)$ and $\log\left(\frac{1-X_{n,2k-1}}{1-\tau}\right)=\log\left(1+\frac{\tau-X_{n,2k-1}}{(1-\tau)}\right)$, we have

$$\begin{aligned}
&\mathbb{E}\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}}\left\{\lambda_n((0,\tau])\log\mathbb{E}\left(\frac{1-X_{n,2k-1}}{1-\tau}\middle|\{q_{n-1}\},X_{n,2k-1}\le\tau\right)\right.\\
&\qquad\left.+(1-\lambda_n((0,\tau]))\log\mathbb{E}\left(\frac{X_{n,2k-1}}{\tau}\middle|\{q_{n-1}\},X_{n,2k-1}>\tau\right)\right\}G_0(I_{n-1,k})\\
&\le\mathbb{E}\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}}\left\{\lambda_n((0,\tau])\mathbb{E}\left(\frac{\tau-X_{n,2k-1}}{1-\tau}\middle|\{q_{n-1}\},X_{n,2k-1}\le\tau\right)\right.\\
&\qquad\left.+(1-\lambda_n((0,\tau]))\mathbb{E}\left(\frac{X_{n,2k-1}-\tau}{\tau}\middle|\{q_{n-1}\},X_{n,2k-1}>\tau\right)\right\}G_0(I_{n-1,k}).
\end{aligned}$$

The identity $\mathbb{E}(X)=\Pr(X\in A)\mathbb{E}(X\mid A)+\Pr(X\in\bar{A})\mathbb{E}(X\mid\bar{A})$ implies $\mathbb{E}(X)\ge\Pr(X\in A)\mathbb{E}(X\mid A)$ for any $X\ge 0$:

$$\begin{aligned}
&\mathbb{E}\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}}\left\{\lambda_n((0,\tau])\mathbb{E}\left(\frac{\tau-X_{n,2k-1}}{1-\tau}\middle|\{q_{n-1}\},X_{n,2k-1}\le\tau\right)\right.\\
&\qquad\left.+(1-\lambda_n((0,\tau]))\mathbb{E}\left(\frac{X_{n,2k-1}-\tau}{\tau}\middle|\{q_{n-1}\},X_{n,2k-1}>\tau\right)\right\}G_0(I_{n-1,k})\\
&\le\mathbb{E}\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}}\left\{\mathbb{E}\left(\left|\tfrac{\tau-X_{n,2k-1}}{1-\tau}\right|\ \middle|\ \{q_{n-1}\}\right)+\mathbb{E}\left(\left|\tfrac{X_{n,2k-1}-\tau}{\tau}\right|\ \middle|\{q_{n-1}\}\right)\right\}G_0(I_{n-1,k})\\
&=\mathbb{E}\sum_{n=1}^{\infty}\sum_{k=1}^{2^{n-1}}\frac{\mathbb{E}(|\tau-X_{n,2k-1}|\mid\{q_{n-1}\})}{\tau(1-\tau)}G_0(I_{n-1,k})\\
&\le\sum_{n=1}^{\infty}\frac{\sqrt{\sigma_n^2+|\tau_n-\tau|^2}}{\tau(1-\tau)}<\infty,
\end{aligned}$$

under the assumptions of the theorem.

Since it converges in expected value, the term $\int \log g(x)g_0(x)dx$ is bounded with a positive probability. Finally, similarly to [22], we can say that, because of the assumption that all λ_n have full support in $(0,1)$, then $\int g_0(x)\log g(x)dx$ can be found to be at most different by a constant δ from $\int g_0(x)\log g_0(x)dx$ with a positive probability. □

3.3. Stopping rule

In building a random probability measure the problem of repeating (3.1) infinitely many times occurs. This is a common feature of most of the procedures based on dyadic expansions or similar ideas (including Polya trees, quantile pyramids, SBA-RPMs among the others). However, the SQA presents also the drawback that with a small τ the random distribution generated after a relatively small number of steps is extremely dense in the left tail and still sparse in the right and the lower is τ the higher the number of steps necessary to obtain a random distribution that is dense everywhere. This clearly causes the algorithm to become very slow for small values of τ, because a large number n of *levels* is necessary to cover in a reasonably dense way the whole support of the distribution and, for each n, $q_{n,2k-1}$ should be computed (from $X_{n,2k-1}$) for $k \leq 2^{n-1}$. However, similar as in [22] it is not necessary to compute any n but just as far as a certain level h and also not every $q_{n,k}$ for all $n = 1, ..., h$.

The SQA procedure can be implemented efficiently by fixing a threshold $\varepsilon > 0$, and by simulating only those values of $\{q_{n,2k-1}\}$ such that $G((q_{n,k-1}, q_{n,k}]) > \varepsilon$. This reduces drastically the number of iterations. As example, Fig. 2 plots for $n = 1, ..., 12$ the intervals for which $q_{n,k}$ must be computed according to the truncation procedure with $\varepsilon = 10^{-3}$ and $\tau = 0.01$. In this case the procedure stops after 134 levels but in the figure we report only the first 12 levels.

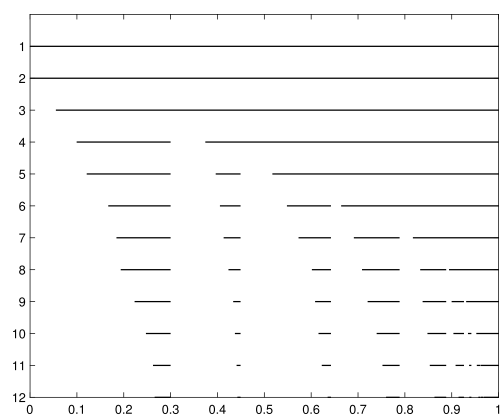

Fig 2. *Graphical representation of the intervals in which $q_{n,k}$ are simulated with $\varepsilon = 10^{-3}$, $\tau = 0.01$. The figure shows only 12 levels, while to cover in a reasonably dense way (according to ε) the whole intervals, 134 levels are needed.*

In practice, it is easy to implement the truncation procedure. Note that, given n, the probability of each interval $(q_{n,k-1}, q_{n,k}]$ is given by $G((q_{n,k-1}, q_{n,k}]) = \tau^{z_{n,k}}(1-\tau)^{n-z_{n,k}}$ where we recall that $z_{n,k}$ is the number of zeros in the binary representation of $k-1$ (adding zeros to the left if necessary in order to have exactly n cyphers). So, for example, the interval $(q_{6,k-1}, q_{6,k}]$ for $k = 13$ has probability $G((q_{n,k-1}, q_{n,k}]) = \tau^4(1-\tau)^2$, because the binary representation of 12 is 001100. Let us assume, without loss of generality, that $\tau < 1/2$ and fix a threshold ε. Then if we set $n_\varepsilon = \min_n\{\tau^n < \varepsilon\}$, we can limit to the simulation of the $X_{n,k}$ corresponding to the indices

$$K_{n,n_\varepsilon} = \{k \le 2^n : z_{n,k} < n_\varepsilon\} \supseteq \{k \le 2^n : \; G((q_{n,k-1}, q_{n,k}]) > \tau^{n_\varepsilon}\}.$$

Let moreover $N_\varepsilon = \min_n\{(1-\tau)^n < \varepsilon\}$. Because of the truncation mechanism, whenever $G_n(I_{n,k}) < \varepsilon$, no other quantiles inside $I_{n,k}$ are generated. Then, the numbers n_ε and N_ε are, respectively, the number of iterations before the first truncation and the last iteration before the procedure stops[4]. Once we have all q_{n,k_j}, $n \le N_\varepsilon, k_j \in K_{n,n_\varepsilon}$, the truncating measure is $G_\varepsilon(q_{n,k}) = G_n(q_{n,k_j})$ and it is extended to $(0,1)$ through linear interpolation.

Summing up this procedure is both a stopping rule because it provides the largest N_ε level to compute and a truncation procedure because for any level n provides the intervals needing to be still partitioned.

When defining a stopping rule, it is important to ascertain that the approximated random measure obtained by truncation, that we denote by G_ε to stress its dependence on the threshold ε, can be arbitrarily close to the random probability measure G generated by the procedure. Because of the definition of the stopping rule, it is easy to see that, under very mild assumptions, G_ε can be arbitrarily close to G with respect to the Kolmogorov-Smirnov distance: $d_{KS}(G_\varepsilon, G) = \sup_x |G_\varepsilon(x) - G(x)|$. If the generating measures are allowed to vary with the level n, then it is also possible to prove that ε can be chosen such that G lies in an arbitrarily small KL-neighborhood of G_ε.

After applying the stopping rule, we therefore have a subsequence of quantiles $\{q_{N_\varepsilon,k_j}\}$ ($j = 1, \ldots, M$, with M depending on τ, ε and n) that satisfy $k_1 = 1$ and $k_{M-1} = 2^{n_\varepsilon} - 1$.

Theorem 3.7. *Let G be a SQA-RPM and G_ε its truncated version. (i) For every $\gamma > 0$ there is a $\varepsilon > 0$ such that $B_{\lambda_n, n\ge 1}(G : d_{KS}(G, G_\varepsilon) < \gamma) = 1$. (ii) Let G be a SQA-RPM obtained under the assumptions of Theorem 3.5 and the additional assumption that each λ_n have support $[l_n, 1-l_n]$, with $l_n \to 0$ and $|\tau_n - \tau| \log(1/l_n) = O(n^{-1-\zeta})$ for some $\zeta > 0$, then for all $\gamma > 0, \eta > 0$ there exists a $\varepsilon > 0$, such that*

$$B_{\lambda_n, n\ge 1}(G : KL(G, G_\varepsilon) < \gamma) > 1 - \eta.$$

Proof. (i) The Kolmogorov-Smirnov distance is given by:

$$d_{KS}(G_\varepsilon, G) = \sup_x |G_\varepsilon(x) - G(x)| \le \sup_x |G_\varepsilon(x) - G_n(x)| + \sup_x |G_n(x) - G(x)|, \quad (3.7)$$

[4]If $\tau > 1/2$ the definitions of N_ε and n_ε are inverted.

where $n > N_\varepsilon$. For the last term of (3.7) we have

$$\begin{aligned}\sup_x |G_n(x) - G(x)| \\ = \sup_x \left| \sum_{k=1}^{2^n} (G_n(I_{n,k})\mathbf{1}\{x \geq q_{n,k}\} - G(I_{n,k})\mathbf{1}\{x \geq q_{n,k}\}) - G((q_{n,k(x)}, x]) \right|\end{aligned}$$

where $q_{n,k(x)}$ is the largest quantile (of level n) smaller than x. Then, because of $G_n(q_{n,k}) = G(q_{n,k})$,

$$\sup_x |G_n(x) - G(x)| = \sup_k \sup_{x \in I_{n,k}} |G((q_{n,k}, x])| \leq \sup_k G(I_{n,k}) = (1-\tau)^n < \varepsilon.$$

The first term of (3.7) is equal to

$$\begin{aligned}\sup_j \sup_{x \in (q_{N_\varepsilon,k_{j-1}}, q_{N_\varepsilon,k_j}]} \left| G_{N_\varepsilon}((q_{N_\varepsilon,k_j}, x]) - G_n((q_{N_\varepsilon,k_j}, x]) \right| \\ < \sup_j \sup_{x \in (q_{N_\varepsilon,k_{j-1}}, q_{N_\varepsilon,k_j}]} \left| \frac{x - q_{N_\varepsilon,k_{j-1}}}{q_{N_\varepsilon,k_j} - q_{N_\varepsilon,k_{j-1}}} G_n((q_{N_\varepsilon,k_{j-1}}, G_n((q_{N_\varepsilon,k_j}, x])]) \right| < \varepsilon\end{aligned}$$

by construction.

(ii) To prove the second part, we write,

$$\begin{aligned}KL(G, G_\varepsilon) &= \int g(x) \log \frac{g(x)}{g_\varepsilon(x)} dx = \int g(x) \log \left(\lim_n \frac{G(I_{n,k(x)})/\nu(I_{n,k(x)})}{G_\varepsilon(I_{n,k(x)})/\nu(I_{n,k(x)})} \right) dx \\ &= \int \sum_{m=1}^{\infty} g(x) \log \left(\frac{G(I_{m,k(x)} \mid I_{m-1,k(x)})/\nu(I_{m,k(x)} \mid I_{m-1,k(x)})}{G_\varepsilon(I_{m,k(x)} \mid I_{m-1,k(x)})/\nu(I_{m,k(x)} \mid I_{m-1,k(x)})} \right) dx \\ &= \sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} \left[\log \tfrac{\tau}{G_\varepsilon(I_{n,2k-1} | I_{n-1,k})} G(I_{n,2k-1}) + \log \tfrac{1-\tau}{G_\varepsilon(I_{n,2k} | I_{n-1,k})} G(I_{n,2k}) \right].\end{aligned}$$

Note that, for $n < n_\varepsilon$, $G_\varepsilon(I_{n,k}) = G_n(I_{n,k})$, for all k, then $G_\varepsilon(I_{n,2k-1} \mid I_{n-1,k}) = \tau$ and $G_\varepsilon(I_{n,2k} \mid I_{n-1,k}) = 1 - \tau$, implies

$$\sum_{n=1}^{n_\varepsilon} \sum_{k=1}^{2^{n-1}} \left[\log \frac{\tau}{G_\varepsilon(I_{n,2k-1} \mid I_{n-1,k})} G(I_{n,2k-1}) + \log \frac{1-\tau}{G_\varepsilon(I_{n,2k} \mid I_{n-1,k})} G(I_{n,2k}) \right] = 0.$$

Then,

$$KL(G, G_\varepsilon) = \sum_{n=n_\varepsilon+1}^{\infty} \sum_{k=1}^{2^{n-1}} \left[\log \frac{\tau}{G_\varepsilon(I_{n,2k-1} \mid I_{n-1,k})} G(I_{n,2k-1}) + \log \frac{1-\tau}{G_\varepsilon(I_{n,2k} \mid I_{n-1,k})} G(I_{n,2k}) \right].$$

For $n > N_\varepsilon + 1$, the partition $\{I_{n,k}\}_{k \leq 2^n}$ is finer than the partition $\{I_{N_\varepsilon,k_j}\}_j$. For $n_\varepsilon + 1 < n \leq N_\varepsilon$, some of the intervals $I_{n,k}$ coincide with I_{N_ε,k_j} for some $j \leq M$, while for the others we can find j such that $I_{n,k} \subseteq I_{N_\varepsilon,k_j}$. In the first case, we have, by construction $G_\varepsilon(I_{n,2k-1} \mid I_{n-1,k}) = \tau$, while in the second case, since

$G_\varepsilon(x)$ is defined by linear interpolation, we have $G_\varepsilon(I_{n,2k-1} \mid I_{n-1,k}) = X_{n,2k-1}$ Then, we have the upper bound

$$KL(G, G_\varepsilon) \leq \sum_{n=n_\varepsilon+1}^{\infty} \sum_{k=1}^{2^{n-1}} \left[\log \frac{\tau}{X_{n,2k-1}} G(I_{n,2k-1}) + \log \frac{1-\tau}{1-X_{n,2k-1}} G(I_{n,2k}) \right].$$

We now prove that the series

$$\sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} \left[\log \frac{\tau}{X_{n,2k-1}} G(I_{n,2k-1}) + \log \frac{1-\tau}{1-X_{n,2k-1}} G(I_{n,2k}) \right] \tag{3.8}$$

is convergent in mean (and therefore in probability). Note that in the following we exploit the fact that, although $I_{n,k}$ is random, $G(I_{n,k})$ is a nonrandom function of τ, n and k:

$$\begin{aligned}
&\mathbb{E}\left(\sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} \left[\log \frac{\tau}{X_{n,2k-1}} G(I_{n,2k-1}) + \log \frac{1-\tau}{1-X_{n,2k-1}} G(I_{n,2k}) \right] \right) \\
&\quad = \sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} \left[G(I_{n,2k-1}) \mathbb{E} \log \frac{\tau}{X_{n,2k-1}} + G(I_{n,2k}) \mathbb{E} \log \frac{1-\tau}{1-X_{n,2k-1}} \right] \\
&\quad \text{(using concavity of } \log(x) \text{ and Jensen's inequaliy)} \\
&\quad = \sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} \left[G(I_{n,2k-1}) \log \mathbb{E} \frac{\tau}{X_{n,2k-1}} + G(I_{n,2k}) \log \mathbb{E} \frac{1-\tau}{1-X_{n,2k-1}} \right].
\end{aligned}$$

Now, we use the sharp Kantorovich's inequality (see eq. (4) in [8]), with $M = 1 - l_n$ and $m = l_n$: we have that

$$\mathbb{E}\left(\frac{\tau}{X_{n,2k-1}} \right) \cdot \left(\frac{\tau}{\mathbb{E} X_{n,2k-1}} \right)^{-1} \leq \frac{\left(\frac{1-l_n}{1-2l_n} \log \frac{1-l_n}{l_n} - \frac{1}{2(1-l_n)} \right)^2}{2 \left(\log \frac{1-l_n}{l_n} - \frac{1-2l_n}{1-l_n} \right)} \leq -\frac{3}{2} \log(l_n),$$

where the last inequality holds for all $l_n < 0.5$. The same inequality holds for $\mathbb{E}\left(\frac{1-\tau}{1-X_{n,2k-1}} \right)$. Then, we can write the following upper bound for the sum above:

$$\begin{aligned}
&\frac{3}{2} \sum_{n=1}^{\infty} \sum_{k=1}^{2^{n-1}} \left(G(I_{n,2k-1}) \log \frac{\tau}{\tau_n} + G(I_{n,2k}) \log \frac{1-\tau}{1-\tau_n} \right) |\log(l_n)| \\
&\quad = \frac{3}{2} \sum_{n=1}^{\infty} \left(\tau \log \frac{\tau}{\tau_n} + (1-\tau) \log \frac{1-\tau}{1-\tau_n} \right) |\log(l_n)|,
\end{aligned}$$

where the last identity follows from $\sum_{k=1}^{2^{n-1}} G(I_{n,2k-1}) = 1 - \sum_{k=1}^{2^{n-1}} G(I_{n,2k}) = \tau$ (this can be seen by induction). The right hand side converges under the assumptions of the theorem which implies convergence in mean of (3.8).

This implies that

$$\mathbb{E}\left(\sum_{n=m}^{\infty}\sum_{k=1}^{2^{n-1}}\left[\log\frac{\tau}{X_{n,2k-1}}G(I_{n,2k-1})+\log\frac{1-\tau}{1-X_{n,2k-1}}G(I_{n,2k})\right]\right)\overset{m\to\infty}{\longrightarrow}0$$

and this yields convergence in probability, that is: for every $\gamma, \eta > 0$, $\exists \varepsilon > 0$ such that

$$\mathcal{B}_{\lambda_n,\,n\geq 1}\left\{\sum_{n=n_\varepsilon+1}^{\infty}\sum_{k=1}^{2^{n-1}}\left[\log\frac{\tau}{X_{n,2k-1}}G(I_{n,2k-1})+\log\frac{1-\tau}{1-X_{n,2k-1}}G(I_{n,2k})\right]<\gamma\right\}>1-\eta.$$

□

4. Generalizations

We dedicate this section to some generalizations. First we extend SQA procedure to unbounded set as $\mathbb{R}$. Then we show that the SQA can be used to generate random probability distributions F, whose expectiles of level τ (or more generally, whose generalized M-quantiles, following the definition of [5]) have distribution λ_1. Finally we consider the case of generating a random probability distribution with two given quantiles.

4.1. Extension to RPM's on unbounded sets

The SQA procedure described in Section 3.2 can be extended for the construction of random measures in $\mathbb{R}$ (the same reasoning applies to any unbounded subset of $\mathbb{R}$) without losing their main properties.

Let $H : \mathbb{R} \mapsto [0,1]$ be an absolutely continuous cdf. Then, any SQA-RPM G on $[0,1]$ can be transformed into a RPM on $\mathbb{R}$ by using the distribution H as a *link function*: $F(x) = G(H(x))$. Then, if G is differentiable, so is F, with density $f(x) = g(H(x))h(x)$.

Moreover, for any absolutely continuous distribution F_0 in $\mathbb{R}$, the distribution $G_0(u) = F_0(H^{-1}(u))$ is an absolutely continuous distribution function on $[0,1]$ with density $g_0(u) = f_0(H^{-1}(u))/h(H^{-1}(u))$ and we have

$$\begin{aligned}KL(F_0,F) &= \int f_0(x)\log\frac{f_0(x)}{f(x)}dx\\ &\quad\text{by applying the transform } x = H^{-1}(u)\\ &= \int \frac{f_0(H^{-1}(u))}{h(H^{-1}(u))}\log\frac{f_0(H^{-1}(u))}{f(H^{-1}(u))}du\\ &= \int g_0(u)\log\frac{g_0(u)}{g(u)}du = KL(G_0,G).\end{aligned}$$

From this remark we can conclude that Theorem 3.6 can be also extended to SQA-RPM on the real line. A small issue to take care of is to ascertain that F_0

is such that the entropy of G_0 is finite. This can be checked by noting that

$$\int g_0(u)\log g_0(u)du = \int \frac{f_0(H^{-1}(u))}{h(H^{-1}(u))}\log\frac{f_0(H^{-1}(u))}{h(H^{-1}(u))}du$$
$$= \int \frac{f_0(x)}{h(x)}\log\frac{f_0(x)}{h(x)}\; h(x)dx = KL(F_0, H).$$

Thus, the condition of bounded entropy of G_0 is equivalent to finite KL-divergence between H and F_0.

Also the truncation procedure defined in Section 3.3 is easily adapted. Given the continuous distribution H, the truncated SQA random distribution function F_ε may be defined from G_ε: $F_\varepsilon(x) = G_\varepsilon(H(x))$: for any $x \in (H^{-1}(q_{N_\varepsilon,k_{j-1}}), H^{-1}(q_{N_\varepsilon,k_j})]$

$$F_\varepsilon(x) = G(q_{N_\varepsilon,k_{j-1}}) + \frac{H(x) - q_{N_\varepsilon,k_{j-1}}}{q_{N_\varepsilon,k_j} - q_{N_\varepsilon,k_{j-1}}} G(I^\varepsilon_j),$$

where $I^\varepsilon = (q_{N_\varepsilon,k_{j-1}}, q_{N_\varepsilon,k_j}]$.

Then, while the density of G_ε is stepwise constant, the density of F_ε is not constant and its behavior depends on the link H:

$$f_\varepsilon(x) = \frac{G(I^\varepsilon_j)}{q_{N_\varepsilon,k_j} - q_{N_\varepsilon,k_{j-1}}} h(x), \quad \text{for} \quad H(x) \in I^\varepsilon_j.$$

This definition of the truncated distribution allows us to approximate F with arbitrary precision (according to the measure of discrepancy defined by the KL-divergence). In fact, Theorem 3.7(ii) applies, because of

$$KL(F, F_\varepsilon) = \sum_j \int_{x:H(x)\in I^\varepsilon_j} h(x)g(H(x))\log\frac{g(H(x))h(x)}{g_\varepsilon(H(x))h(x)}\,dx = KL(G, G_\varepsilon).$$

We point out that, given λ_1, the choice of H might be determined by the constraint on the distribution of the random τ-quantile. In fact, from

$$\Pr\left(F^{-1}(\tau) \le u\right) = \Pr\left(F(u) \ge \tau\right) = \Pr\left(G(H(u)) \ge \tau\right) = \lambda_1\left((0, H(u)]\right),$$

then if we wish for the τ-quantile of F to have a given distribution, say π, we need H to satisfy $H(u) = \lambda_1^{-1}((0, \pi(u)])$.

4.2. Generating random distributions from M-quantiles

One way to define the quantile of level τ of a continuous distribution with density f, is as that value q that satisfies

$$\int \rho_1(x - q)f(x)dx = 0,$$

where $\rho_1(u) = \tau - \mathbf{1}\{u \leq 0\}$. This corresponds to solving the optimization:

$$q^\tau(F) = \arg\min_q \mathbb{E}_F \left[\rho_1(X-q)(X-q)\right], \tag{4.1}$$

where $\mathbb{E}_F X = \int x f(x) dx$.

If one replaces to the function ρ_1 the function $\rho_2(u) = (\tau - \mathbf{1}\{u \leq 0\})|u|$, a different function, called expectile function (Newey and Powell (1987) [27]), is defined. [5] proposed a natural generalization of quantiles and expectiles, by replacing the constant and modulus functions by the modulus of a more general odd function, ψ:

$$\int (\tau - \mathbf{1}\{x \leq y\})|\psi(x-y)| f(x) dx = 0. \tag{4.2}$$

The function ψ is related to the choice of the function in an M-estimation procedure. In particular, if $\tau = .5$, then the corresponding quantity is equal to the M-estimator of the location parameter. For this reason [5] proposed the name M-quantiles for the solution to (4.2).

[20] proved that, if ψ is odd, monotone nondecreasing, continuous and piecewise differentiable, the M-quantile corresponds to the quantile of level τ of the distribution function

$$G(y) = \frac{\int^y \psi(y-x) f(x) dx}{2\int^y \psi(y-x) f(x) dx - \int \psi(y-x) f(x) dx}. \tag{4.3}$$

Thus, in particular, one obtains the standard quantiles by taking ψ equal to the odd function $\psi(u) = sign(u)$, and in that case it is easy to see that $G(y) = F(y)$.

From $\psi(x) = x$, Jones proves in particular that expectiles are quantiles of the distribution function

$$G(y) = \frac{\mu(y) - yF(y)}{2(\mu(y) - yF(y)) + y - \mu}, \tag{4.4}$$

where $\mu(y) = \int^y x f(x) dx$.

Expectiles have been experiencing a growing interest in the recent financial literature (as well as in all fields where it is crucial to manage extreme events, such as actuarial sciences), in particular since the introduction of the concept of elicitability ([14]) and the proof by [35], that expectiles give the only risk measure that is both coherent and elicitable[5]. See [2] and references therein for their use as risk measure in comparison with VaR and ES. However, expectiles are known to be less robust to extreme events relative to the VaR. [10] consider

[5]A risk measure is elicitable if there exists a scoring function such that the risk under a given distribution is obtained by minimizing the expected value of the score under that distribution. All M-quantiles are elicitable. A risk measure is coherent if it is simultaneously (i) translation invariant, (ii) monotonic, (iii) positively homogeneous, (iv) subadditive (see [1] for more details). While the ES is coherent but not elicitable, the VaR is elicitable but not coherent.

L^p-quantile based risk measures, (see [7]), obtained by taking the function ψ in (4.2) equal to $|x|^{p-1}$ with $1 < p < 2$.

The SQA procedure can be used to generate RPMs controlling for the distribution of a given M-quantile. We illustrate the procedure by focusing on the case of expectiles, namely to the case $\psi = |x|$. This clearly includes the case, if $\tau = 1/2$, of random distributions with mean distributed according to λ_1.

The idea behind this procedure is straightforward: exploiting the fact that M-quantiles of a distribution F are ordinary quantiles of the distribution G in (4.3), we first generate a random distribution from a SQA sequence of level τ as described in the previous section. Then, we invert the transformation (4.3) to get the distribution $F := F_G$.

In particular, the inverse transformation of (4.4) can be easily found once we notice that the median of the G distribution and the mean of F_G coincide. Then, for all y, from

$$\frac{G(y)}{1-2G(y)} = \frac{\mu(y) - yF(y)}{y - \mu},$$

we can get:

$$F_G(y) = \frac{\partial}{\partial y} \frac{G(y)(\mu - y)}{1 - 2G(y)}. \tag{4.5}$$

The link between quantiles of G and expectiles of F_G guarantees that, if the τ-level quantile of G has distribution λ_1 $B_{\lambda_n, n\geq 1}$-a.s., then also the τ-expectile of F_G follows the same distribution. We underline that equation (4.5) is independent on the τ level.

If in particular we choose $\tau = 1/2$, this procedure gives a distribution F_G whose mean has distribution λ_1. Besides the already mentioned SBA by [17], the problem of generating RPMs whose mean follow a given distribution, has been studied with a focus on Dirichlet means, motivated by applications in nonparametric statistics and in combinatorics (see [23], [30] and [31]). Our procedure thus also permits to define an alternative approach to tackle this issue.

4.3. SQA with two given quantiles

We can consider an extension of the proposed SQA procedure with more given quantiles. Consider the case we want construct a RPM G with given quantiles at level τ_1 and τ_2, with $\tau_1 < \tau_2$, denoting q^{τ_1} and q^{τ_2} respectively and assuming we need $q^{\tau_1} \sim \lambda_1^1$ and $q^{\tau_2} \sim \lambda_1^2$. The construction in (3.1) must be slightly changed in the number of points in each partition and two initial distributions must be introduced; indeed, in redefining the (3.1), we have that $\{X_{n,2k-1}\}_{n,k}$, $n \geq 1$ and $k = 1, \dots, 3 \cdot 2^{n-2}$ is a triangular array of independent random variables, with $X_{1,1} \sim \lambda_1^1$, $X_{1,2} \sim \lambda_1^2$ and $X_{n,2k-1} \sim \lambda_n$, for all $n \geq 2$. Then the first two lines of the recursive construction (3.1) change in

$$\begin{aligned} &q_{1,1} = X_{1,1} \qquad q_{1,2} = X_{1,2} \\ &q_{n,2k} = q_{n-1,k}, \ \ n \geq 1, \ k = 1, \dots, 3 \cdot 2^{n-2} \end{aligned} \tag{4.6}$$

and the rest remains unchanged. Supposing to adopt a generic τ after the first partition, the result in Theorem 2.1 still holds, except for $G(q_{n,0}) = 0 = 1 - G(q_{n,2^n})$ that must be transformed in $G(q_{n,0}) = 0 = 1 - G(q_{n,3\cdot 2^{n-3}})$ and the fact that $G(q_{11}) = \tau_1$ and $G(q_{1,2}) = \tau_2$.

Finally Theorem 3.1 still holds with

$$B_{\lambda_n, n\geq 1}\left\{G : G^{-1}(\tau_i) \leq z\right\} = \lambda_1^i([0,z]) \quad i = 1, 2.$$

5. Simulations

In the following we report some examples of random probabilities obtained via our SQA construction. The truncation procedure has been used to reduce the computational time with $\varepsilon = 10^{-3}$.

In Figure 3 five examples of random distributions obtained with $q_{1,1} = 0.1$ and $\tau = 0.05$ are reported; these five examples are derived with the same λ distribution equal to a $Beta(3, 30)$, plotted on the left up corner. In contrast, in Figure 4, five examples of random distributions obtained with the same $q_{1,1} = 0.1$ and $\tau = 0.05$, but with five different λs, are reported. In this case the λ distributions, plotted on the left up corner, are $Beta(3, 3a^2)$, with $a = 1, .., 5$. The more the distribution is symmetric the higher the probability assigned to one. Indeed a symmetric λ tends to select more likely the mid point of each interval leaving τ probability on the left and $1 - \tau$ on the right. Given a fixed

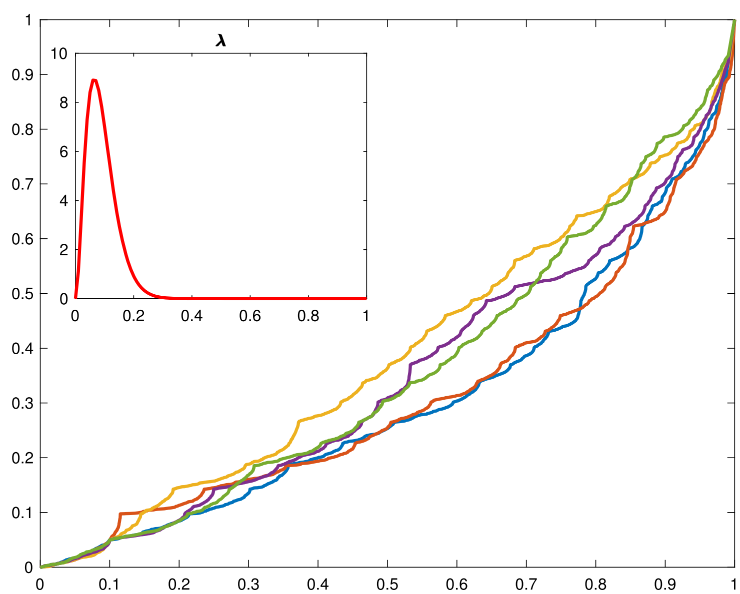

FIG 3. *Five random SQA cumulative distribution function with the same* λ *that is a* $Beta(3, 30)$. *The quantile* $q_{1,1}$ *is* 0.1 *with* $\tau = 0.05$.

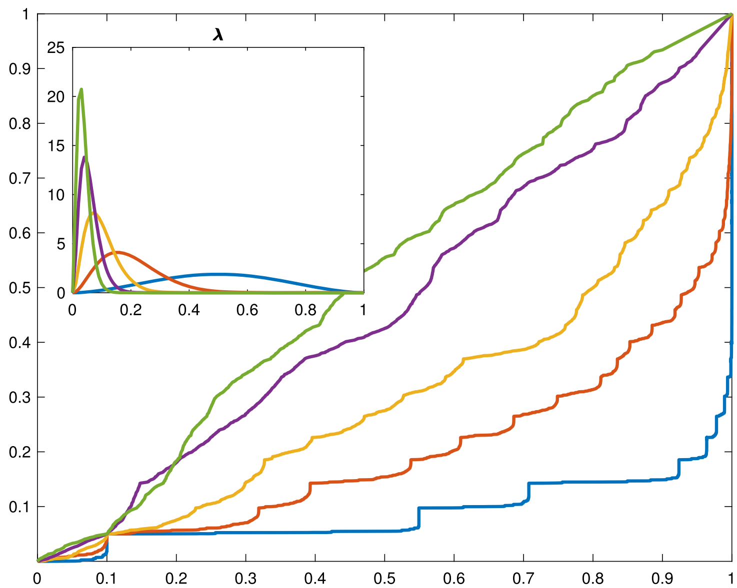

FIG 4. *Five random SQA cumulative distribution function with different* λ. *The distributions* λ *are* $Beta(3, 3a^2)$ *with* $a = 1, .., 5$. *The quantile* $q_{1,1}$ *is* 0.1 *with* $\tau = 0.05$.

quantile $q_{1,1}$ at level τ all the distributions pass necessarily through the point $(q_{1,1}, \tau)$[6], see Figure 3 and 4.

Differently from [6] the SQA procedure allows to build random distributions with quantile following a given distribution λ_1; in Fig 5 five examples are reported. In this case λ is $Beta(3, 15)$ and is the same for all the five examples, while the given quantile is drawn by a λ_1 equal to a uniform from 0 and 0.2.

Differently from [17] SQA-RPM can be continuous even if λ assigns some probability to 0. Fig 6 shows on the left hand side five SQA-RPMs and on the right hand side five SBA-RPM. Both SQA and SBA are generated with λ that gives 0.1 probability to 0 and Uniform(0,1) elsewhere; the quantile and the mean are both fixed to 0.4 and for an easier comparison τ is 0.5. This figure offers a graphical intuition about the difference in continuity of the two constructions as formalized in our Theorem 3.3 and Theorem 3.6 by [17]. Indeed in Theorem 3.6 by [17] for any distribution μ (the equivalent of our λ) giving positive mass in zero they get discrete distributions, while we can have also continuous distributions.

In Figure 7 (lhs) five simulations of SQA with λ_n changing with n are represented. The distributions λ_n are chosen to satisfy conditions of Theorems 3.5, 3.6 and 3.7, namely λ_n are $Beta(a_n, b_n)$ in $[l_n, 1 - l_n]$ with $l_n = \frac{0.8\tau}{n\sqrt{n}}$, $a_n = \tau n^2$, $b_n = \frac{1-l_n-\tau_n}{\tau_n-l_n} a_n$ and $\tau_n = \tau + \frac{1}{n^2}$. The quantile $q_{1,1} = 0.2$ and $\tau = 0.2$. Finally in Figure 8 five simulations of RPM on $\mathbb{R}$ are reported, indeed they are the trans-

[6]This is equivalent to choosing $\lambda_1 = \delta_{q_{11}}$

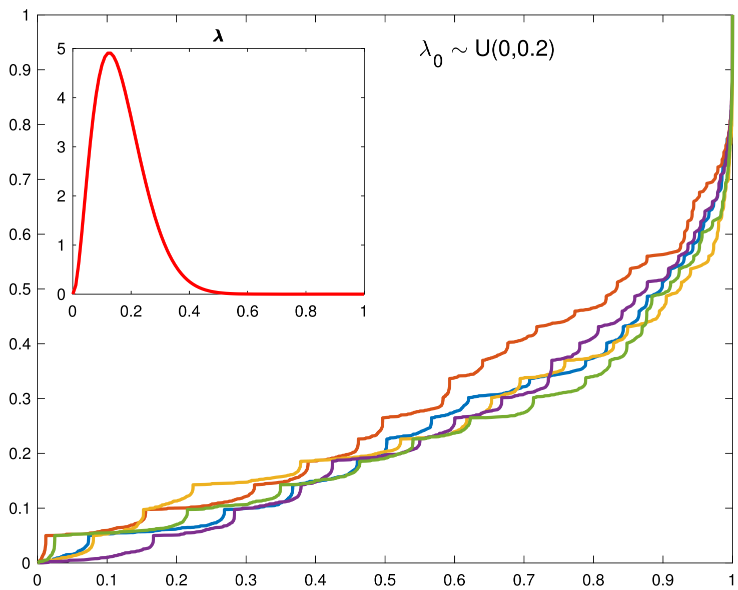

FIG 5. *Five random SQA cumulative distribution function with* λ_1 *uniform from* 0 *and* 0.2. λ *is Beta*$(3, 15)$ *and is the same for all the five examples*

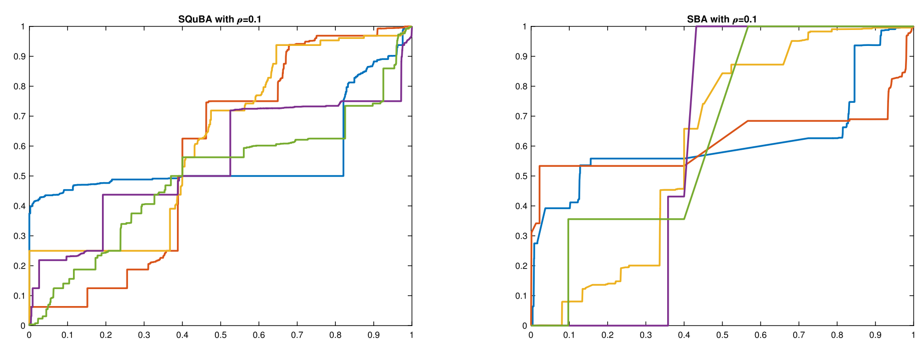

FIG 6. *Five random SQA cumulative distribution function on the left and five random SBA on the right. The SQA is derived with* $\tau = 0.5$ *for an easier comparison with SBA that consider the mean. The median on the left and the mean on the right are assigned equal to* 0.4. λ *gives* 0.1 *mass to* 0 *and is uniform elsewhere.*

formation of those reported in Figure 7 (lhs) with the link function H chosen as the standard normal distribution.

5.1. Comparison with similar approaches

Polya tree and mixture of Polya tree, are widely used to construct random probability distributions. [24] showed that Polya trees can give probability 1 to the

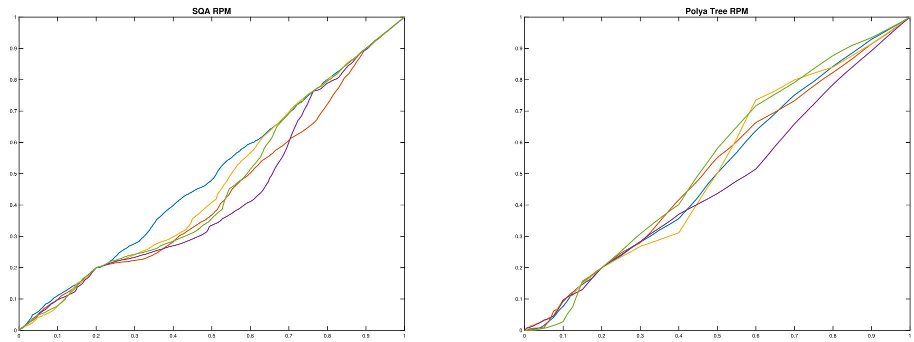

Fig 7. *Five random SQA cumulative distribution functions on the left with λ_n changing with n. Five random Polya tree cumulative distribution function on the right. Both are derived with $\tau = 0.2$ and quantile 0.2. The λ_n in SQA are such that to satisfy conditions of Theorems 3.5, 3.6 and 3.7 while PT is simulated as in [34] using a Beta(n^2, n^2).*

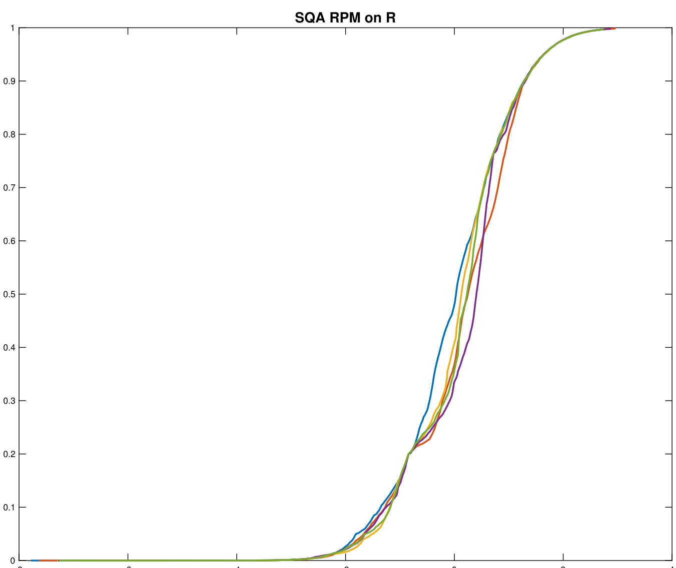

Fig 8. *Five random SQA cumulative distribution function on $\mathbb{R}$ obtained by those reported in 7(rhs) with the link function H chosen as the inverse of a standard normal.*

set of continuous distribution functions, they constitute a conjugate family with an easy update to get the posterior distribution and they have full support. However a well known drawback is the dependence of the model on the partition specified to construct the tree, indeed the partition can affect the posterior distribution. [28] addressed this issue “jittering” the partition. Similarly [18] proposed quantile pyramids to build random partition with fixed mass instead of having random mass in a fixed partition. Our approach builds a random partition assigning back the distribution, in this sense it looks more alike to quantile pyramids. In terms of Polya tree, SQA is like a realization from a Polya tree distribution with partition given by our SQA realization and degenerate Beta. In terms of quantile pyramids instead SQA can be seen as a generalization that uses τ quantiles instead of dyadic quantiles. With both methods we share the property of producing continuous distributions with probability one, whenever the partition has not degenerate intervals, and we share also the property of large support on the space of probability measures.

Polya trees and Mixtures of Polya trees, that are less sensible to the partition choice, can be used also to generate RPMs with given median or quantiles and this finds application in Bayesian quantile regression (see [34], [16]). Relative to those methods, the SQA procedure permits an even easier simulation of prior distributions, with any distributional constraint on one or more quantiles. As quantile pyramides, however, SQA distribution is less analitically tractable than Polya tree and requires a computational effort to derive the posterior distribution. In Figure 7 (rhs) five simulations of Polya tree RPM are represented jointly with five simulations of SQA with λ_n changing with n. In both the distribution generating the partition in SQA and the weights of the partition in PT are chosen to satisfy conditions that ensures continuity and differentiability. Polya Tree RPMs are simulated according to [34] with the Beta distributions with parameters $a_n = b_n = n^2$. A comparison of the graphs produced by the two methods (also considering different choices of the parameters), shows that SQA-RPMs can be a valid substitute for the Polya tree RPMs.

6. Concluding remarks

In this paper we studied a procedure for the generation of RPMs based on a sequence of conditional quantiles. This procedure offers an alternative with respect to the popular approaches based on Polya trees, and can be seen as a generalization of the more recent quantile pyramids ([18]). The SQA procedure is able to produce families of both continuous and discrete random distributions, and also with full support, under the appropriate conditions.

As the Polya tree or other procedures for the generation of random probability measures, the SQA procedure can be used in Bayesian nonparametric statistics, and it represents a suitable choice especially in cases when it is necessary to control for one or more quantiles (see for instance the approach in [29]). A thorough analysis of the application of SQA-RPM to Bayesian nonparametric statistics is worth doing and will be object of future research.

Further, this procedure can be used in other types of applications as those addresses by [17]. We here mention some examples for which a procedure fixing one (or more) quantile distribution is convenient.

Example 5 (Approximation of Universal Constants)**.** *The SQA procedure can be used to derive an experimental method to estimate the supremum or the infimum of a continuous functional* $f : \mathcal{P}([0,1]) \to \mathbb{R}$*, under constraints on an* M*-quantile.*

As a particular case, we can consider the problem of finding the KL projection of a probability distribution onto a set of probability distributions defined by nonlinear constraints[7]*. Given a two distributions* F_0, F *with Radon-Nikodym derivative* dF/dF_0*, the KL-divergence* $KL(F, F_0) = \int \log(dF/dF_0)\, dF$*. For any subset* $\Omega \subset \mathcal{P}([0,1])$*, a KL projection of* F_0 *onto* Ω *is the pdf* $F^* \in \Omega$ *satisfying* $KL(F^*, F_0) \leq KL(F, F_0)$*, for every* $F \in \Omega$*. Suppose that* Ω *is the set of distribution functions with a constraint on a quantile (i.e.* $F^{-1}(\tau) = c$*). This is a nonlinear constraint, therefore the minimum KL-divergence between* F_0 *and* Ω *can be approximated by taking the minimum over a finite set of simulated SQA distributions* $\{F_i, i = 1, \ldots, N\}$*:* $KL(F^*, F_0) = \min_i KL(F_i, F_0)$*.*

Example 6. *The SQA procedure can be used to generate scenarios in which the distributions must satisfy some condition on the risk measure if this is defined by quantiles (VaR for example) or expectiles.*

SQA could be also applied in generating RPMs for which the First degree Stochastic Dominance (FSD) holds. Indeed given that, according to the quantile formulation of FSD, F *dominates* G *if* $q^\tau(F) \geq q^\tau(G)\ \forall \tau \in [0,1]$*, fixing one of the two distributions, a specific construction using SQA and involving a sequence of quantiles can generate randomly the other distribution such that the FSD holds.*

Acknowledgements

The authors wish to thank the anonymous referee and the associate editor, whose comments and suggestions decidedly contributed to the significant improvement of the paper over the initial submission.

References

[1] Artzner, P., Delbaen, F., Eber, J.-M., and Heath, D. (1999). Coherent measures of risk. *Mathematical Finance*, 9:203–228. MR1850791

[2] Bellini, F. and Bernardino, E. D. (2017). Risk management with expectiles. *The European Journal of Finance*, 23(6):487–506.

[3] Billingsley, P. (1968). *Convergence of Probability Measures*. Wiley Series in Probability and Mathematical Statistics, Wiley, 1968. MR0233396

[4] Bloomer, L. and Hill, T. (2002). Random probability measures with given mean and variance. *Journal of Theoretical Probability*, 15:919–937. MR1937779

[5] Breckling, J. and Chambers, R. (1988). M-quantiles. *Biometrika*, 75(4):761–771. MR0995118

[6] Cascos, I. and Molchanov, I. (2013). Choosing a random distribution with prescribed risks. *Insurance: Mathematics and Economics*, 52(3):599–605. MR3054751

[7] In case of linear constraints (that is, of the form $\int g(x)dF(x) = a$), there is no need to approximate the minimum of KL divergence because a characterisation of the Radon-Nikodym density of the projection is available (see for example [32]).

[7] Chen, Z. (1996). Conditional l^p-quantiles and their application to testing of symmetry in non-parametric regression. *Statistics & Probability Letters*, 29:107–115. MR1411408

[8] Clausing, A. (1982). Kantorovich-type inequalities. *The American Mathematical Monthly*, 89(5):314–330. MR0653510

[9] Crauel, H. (2002). *Random Probability Measures on Polish Spaces*. CRC Press. MR1993844

[10] Daouia, A., Girard, S., and Stupfler, G. (2017). Extreme m-quantiles as risk measures: From l_1 to l_p optimization. TSE Working Papers 17-841, Toulouse School of Economics (TSE). MR3892320

[11] Dubins, L. E. and Freedman, D. A. (1967). Random distribution functions. In *Proceedings of the Fifth Berkeley Symposium on Mathematical Statistics and Probability, Volume 2: Contributions to Probability Theory, Part 1*, pages 183–214, Berkeley, Calif. University of California Press. MR0214109

[12] Ferguson, T. S. (1973). A bayesian analysis of some nonparametric problems. *Ann. Statist.*, 1(2):209–230. MR0350949

[13] Ferguson, T. S. (1974). Prior distributions on spaces of probability measures. *Ann. Statist.*, 2(4):615–629. MR0438568

[14] Gneiting, T. (2011). Making and evaluating point forecasts. *Journal of the American Statistical Association*, 106:746–762. MR2847988

[15] Graf, S., Mauldin, R., and Williams, S. (1986). Random homeomorphisms. *Advances in Mathematics*, 60:239–359. MR0848153

[16] Hanson, T. and Johnson, W. O. (2002). Modeling regression error with a mixture of polya trees. *Journal of the American Statistical Association*, 97(460):1020–1033. MR1951256

[17] Hill, T. and Monticino, M. (1998). Construction of random distributions via sequential barycenters. *The Annals of Statistics*, pages 1242–1253. MR1647649

[18] Hjort, N. L. and Walker, S. G. (2009). Quantile pyramids for bayesian nonparametrics. *Ann. Statist.*, 37(1):105–131. MR2488346

[19] Ishwaran, H. and James, L. F. (2001). Gibbs sampling methods for stick-breaking priors. *Journal of the American Statistical Association*, 96(453):161–173. MR1952729

[20] Jones, M. (1994). Expectiles and m-quantiles are quantiles. *Statistics & Probability Letters*, 20(2):149–153. MR1293293

[21] Kraft, C. H. (1964). A class of distribution function processes which have derivatives. *Journal of Applied Probability*, 1(2):385–388. MR0171296

[22] Lavine, M. (1994). More aspects of polya tree distributions for statistical modelling. *Ann. Statist.*, 22(3):1161–1176. MR1311970

[23] Lijoi, A. and Prünster, I. (2009). Distributional properties of means of random probability measures. *Statist. Surv.*, 3:47–95. MR2529667

[24] Mauldin, R. D., Sudderth, W. D., and Williams, S. C. (1992). Polya trees and random distributions. *Ann. Statist.*, 20(3):1203–1221. MR1186247

[25] Monticino, M. (2001). How to construct random probability measures. *International Statistical Review*, 69(1):153–167.

[26] Müller, P., Quintana, F., Jara, A., and Hanson, T. (2015). *Bayesian Non-*

parametric Data Analysis. Springer Series in Statistics. MR3309338

[27] Newey, W. K. and Powell, J. L. (1987). Asymmetric Least Squares Estimation and Testing. *Econometrica*, 55(4), 819–847. MR0906565

[28] Paddock, S. M., Ruggeri, F., Lavine, M., and West, M. (2003). Randomized polya tree models for nonparametric bayesian inference. *Statistica Sinica*, 13:443–460. MR1977736

[29] Rodrigues, T., Dortet-Bernadet, J.-L., and Fan, Y. (2019). Pyramid quantile regression. *Journal of Computational and Graphical Statistics*, 28(3):732–746. MR4007754

[30] Romik, D. (2004). Explicit formulas for hook walks on continual young diagrams. *Advances in Applied Mathematics*, 32(4):625–654. MR2053837

[31] Romik, D. (2005). Roots of the derivative of a polynomial. *The American Mathematical Monthly*, 112(1):66–69. MR2110114

[32] Ruschendorf, L. (1984). On the minimum discrimination information theorem. *Statist. & Decisions Suppl. Issue*, 1:263–283. MR0785212

[33] Shiryaev, A. (1995). *Probability*. Graduate Texts in Mathematics. Springer New York. MR3467826

[34] Walker, S. G. and Mallick, B. K. (1997). Hierarchical generalized linear models and frailty models with bayesian nonparametric mixing. *Journal of the Royal Statistical Society: Series B (Statistical Methodology)*, 59(4):845–860. MR1483219

[35] Ziegel, J. (2016). Coherence and elicitability. *Mathematical Finance*, 26:901–918. MR3551510