ON APPROXIMATING LAW-INVARIANT COMONOTONIC COHERENT RISK MEASURES

BY

YUMIHARU NAKANO

ABSTRACT

The optimal quantization theory is applied for approximating law-invariant comonotonic coherent risk measures. Simple L^p-norm estimates for the risk measures provide the rate of convergence of that approximation as the number of quantization points goes to infinity.

KEYWORDS

Coherent risk measures, optimal quantization, average value-at-risk, comonotonicity.

1. INTRODUCTION

In last two decades, many computational methods for value-at-risk, the quantile of a loss distribution, have been proposed. To enumerate a few: historical and Monte Carlo method, Delta and Delta-Gamma approximation of a loss portfolio, and the saddle point approximation (see Feuerverger and Wong, 2000). We refer for a detailed account of this subject for instance to Chapter 2 in McNeil et al. (2005), Chapter 9 in Glasserman (2004) and references therein.

In insurance and credit risk management, average value-at-risk, also called as expected shortfall or conditional value-at-risk, serves as an important risk measure. This risk measure makes up for several drawbacks that value-at-risk has, and is a typical example of law-invariant comonotonic coherent risk measures. See Artzner et al. (1999), Delbaen (2002) and Föllmer and Schied (2004). As in the case of the value-at-risk, Monte Carlo methods are available for the computation of the average value-at-risk (see, e.g., Acerbi and Tasche, 2002 and Yamai and Yoshiba, 2002). Also, for a particular class of distributions, Landsman and Valdez (2003) obtains some explicit analytical formula for tail conditional expectation, which coincides with the average value-at-risk for continuous distributions.

In this paper, we propose the use of the quantization theory for the approximation of the average value-at-risk, or more generally of law-invariant

Astin Bulletin 42(1), 343-353. doi: 10.2143/AST.42.1.2160746

comonotonic coherent risk measures. The quantization means here the optimal approximation of a probability distribution by a discrete probability with a given number of supporting points, which originates from signal processing theory. It has several promising applications, including numerical integration (see Pagès, 1997) and stochastic control problems (see Pagès et al., 2004). Unlike usual integrals, the risk measures are nonlinear in general. So the situation is somewhat different from that of numerical integration. However, the comonotonicity property allows us to compute the risk measures of a quantized random variable. Moreover, simple L^p-norm estimates for the risk measures provide the rate of convergence of that approximation as the number of supporting points goes to infinity. In the next section we discuss these points in more detail.

2. Optimal quantization approach

Let $(\Omega, \mathcal{F}, \mathbb{P})$ be a probability space. Denote $L^p(\Omega, \mathcal{F}, \mathbb{P})$ by L^p for $p \in [1, \infty]$. We set $\|X\|_p := (\mathbb{E}|X|^p)^{1/p}$ for $X \in L^p$ and $p \in [1, \infty)$. Denote also by μ_X the distribution of a random variable X on $(\Omega, \mathcal{F}, \mathbb{P})$. Recall that a functional $\rho : L^1 \to \mathbb{R} \cup \{+\infty\}$ is said to be a *law-invariant coherent risk measure* if the following are satisfied:

(i) $\rho(X) \leq \rho(Y)$ if $X \geq Y$ a.s.;

(ii) $\rho(X + c) = \rho(X) - c$ for $c \in \mathbb{R}$;

(iii) $\rho(X + Y) \leq \rho(X) + \rho(Y)$;

(iv) $\rho(\lambda X) = \lambda \rho(X)$ for $\lambda > 0$;

(v) $\rho(X) = \rho(Y)$ if $\mu_X = \mu_Y$.

See, e.g., Artzner et al. (1999), Delbaen (2002), Föllmer and Schied (2004), and Kaina and Rüschendorf (2009). Moreover, we say that a coherent risk measure ρ is *comonotonic* if

$$\rho(X + Y) = \rho(X) + \rho(Y)$$

whenever $X, Y \in L^1$ satisfies

$$(X(\omega) - X(\omega'))\,(Y(\omega) - Y(\omega')) \geq 0 \tag{1}$$

for all $(\omega, \omega') \in \Omega \times \Omega$ except for a set of probability zero. We also say that two random variables X and Y are comonotone if they satisfy (1).

A typical example of law-invariant comonotonic coherent risk measures is the average value-at-risk AVaR_α at level $\alpha \in (0,1]$ defined by

$$\mathrm{AVaR}_\alpha(X) = \frac{1}{\alpha}\int_0^\alpha \mathrm{VaR}_\lambda(X)\,d\lambda, \quad X \in L^1.$$

Here, $\mathrm{VaR}_\lambda(X)$ is the value-at-risk of the random variable X at level $\lambda \in (0,1)$, defined by

$$\mathrm{VaR}_\lambda(X) = \inf\{x \in \mathbb{R} : \mathbb{P}(-X > x) \leq \lambda\}.$$

In practice, the terms "conditional value-at-risk" and "expected shortfall" are used for AVaR_α. However, following Föllmer and Schied (2004), we prefer the term "average value-at-risk". Note that the average value-at-risk AVaR_α can be represented as

$$\mathrm{AVaR}_\alpha(X) = \max_{\mathbb{Q}\in\mathcal{Q}} \mathbb{E}_\mathbb{Q}[-X], \quad X \in L^1, \tag{2}$$

where $\mathcal{Q} = \{\mathbb{Q} \ll \mathbb{P} : d\mathbb{Q}/d\mathbb{P} \leq 1/\alpha\}$. See Inoue (2003), Kusuoka (2001), and Theorem 4.47 in Föllmer and Schied (2004).

Let μ be a fixed Borel probability measure on (0,1]. We consider the law-invariant comonotonic coherent risk measure ρ defined by

$$\rho(X) = \int_{(0,1]} \mathrm{AVaR}_\lambda(X)\,\mu(d\lambda), \; X \in L^1. \tag{3}$$

Notice that $\rho(X)$ takes a value in $(-\infty,\infty]$ for every $X \in L^1$ since $\mathrm{AVaR}_\lambda(X) \geq \mathrm{AVaR}_1(X) = \mathbb{E}[-X]$, $\lambda \in (0,1]$, follows from (2).

REMARK 1. Let $\psi : [0,1] \to [0,1]$ be the increasing concave function such that $\psi(0) = 0$, $\psi(1) = 1$, and

$$\psi'_+(t) := \lim_{u\downarrow t} \frac{\psi(u) - \psi(t)}{u - t} = \int_{(t,1]} s^{-1}\mu(ds), \; t \in (0,1).$$

Then, for any nonnegative random variable X, the risk measure $\rho(-X)$ can be represented as the so-called Choquet integral

$$\rho(-X) = \int X dc_\psi := \int_0^\infty c_\psi(X > x)\,dx \tag{4}$$

with submodular set function $c_\psi(A) = \psi(\mathbb{P}(A))$. We refer to Föllmer and Schied (2004) for more details.

REMARK 2. The functional H defined by

$$H(X) = \rho(-X), \quad X \in L^1,$$

can be seen as an insurance premium principle. In fact, it is easy to see that H satisfies several axiomatic properties for the insurance premium. See, e.g., Kaas et al. (2001) and Young (2004).

The risk measures of the form (3) are sufficiently general for the class of law-invariant comonotonic coherent risk measures. In fact, Theorem 4.87 in Föllmer and Schied (2004) tells us that on an atomless probability space and in the case where the space of random variables is restricted to L^∞, the risk measures of the form (3) exhaust real-valued law-invariant comonotonic coherent risk measures on L^∞.

In what follows, we are concerned with the problem of approximating $\rho(-X)$ for $X \in L^1$. To this end, first observe that for $A, B \in \mathcal{F}$, the two indicator functions 1_A and 1_B are comonotone if $A \supset B$. Thus, for $A_i \in \mathcal{F}$ and $a_i \in \mathbb{R}$, $i = 1, \ldots, n$, we have

$$\rho\left(-\sum_{i=1}^{n} a_i 1_{A_i}\right) = \sum_{i=1}^{n} a_i \rho(-1_{A_i}) \tag{5}$$

provided that $A_1 \subset \cdots \subset A_n$ and $a_i \geq 0$, $i = 1, \ldots, n$.

On the other hand, it is straightforward to see that $\mathrm{VaR}_\lambda(-1_A) = 1_{(0,\mathbb{P}(A))}(\lambda)$ for $A \in \mathcal{F}$ and $\lambda \in (0,1)$, so

$$\rho(-1_A) = \int_{(0,1]} \frac{1}{\lambda} \int_0^\lambda 1_{(0,\mathbb{P}(A))}(\gamma)\, d\gamma \mu(d\lambda) = \int_{(0,1]} \frac{1}{\lambda} (\mathbb{P}(A) \wedge \lambda) \mu(d\lambda). \tag{6}$$

Using (5) and (6), we have the following.

Lemma 3. *Let $n \in \mathbb{N}$, $x_1, \ldots, x_n \in \mathbb{R}^d$, $f : \mathbb{R}^d \to \mathbb{R}$ be Borel measurable, and $\{A_i\}_{i=1}^n$ be a partition of Ω. Moreover, let $\tau : \{1, \ldots, n\} \to \{1, \ldots, n\}$ be such that $f(x_{\tau(1)}) \leq \cdots \leq f(x_{\tau(n)})$. For the $\mathbb{R}^d$-valued random variable $X^{(n)}$ defined by $X^{(n)} = \sum_{i=1}^n x_i 1_{A_i}$, we obtain*

$$\rho(-f(X^{(n)})) = f(x_{\tau(1)}) + \sum_{i=2}^{n} (f(x_{\tau(i)}) - f(x_{\tau(i-1)})) \int_{(0,1]} \frac{1}{\lambda} (\mathbb{P}(\cup_{k=i}^n A_{\tau(k)}) \wedge \lambda) \mu(d\lambda). \tag{7}$$

Proof. We have

$$f(X^{(n)}) = \sum_{i=1}^{n} f(x_{\tau(i)}) 1_{A_{\tau(i)}} = f(x_{\tau(1)}) + \sum_{i=2}^{n} (f(x_{\tau(i)}) - f(x_{\tau(i-1)})) 1_{\cup_{k=i}^n A_{\tau(k)}}.$$

From this, the translation invariance of ρ, (5) and (6), the lemma follows. □

Now, define a possibly infinite constant c_p by

$$c_p = \begin{cases} \dfrac{p}{p-1} \displaystyle\int_{(0,1]} \lambda^{-1/p} \mu(d\lambda) & \text{if } p \in (1, \infty), \\ \displaystyle\int_{(0,1]} \lambda^{-1} \mu(d\lambda) & \text{if } p = 1. \end{cases}$$

REMARK 4. Consider the case where ρ is simply given by AVaR_α for $\alpha \in (0,1]$, i.e., the case $\mu = \delta_\alpha$. Then obviously $c_p < \infty$ for $p \in [1, \infty)$.

REMARK 5. Consider the case $\psi(x) = x^h$ with $h \in (0,1)$ in the Choquet integral representation (4) for ρ. Then, by Lemma 4.63 in Föllmer and Schied (2004) we have $\mu(dx) = -xd\psi_+'(x)$ and so

$$\int_{(0,1]} \lambda^{-1/p} \mu(d\lambda) = h(1-h) \int_0^1 \lambda^{h-(1/p)-1} d\lambda.$$

Thus, for $p \in [1, \infty)$, the constant c_p is finite if and only if $h > 1/p$.

We will use the following L^p-estimate for ρ. Hereafter, we write $|x|$ for the standard Euclidean norm of a vector x.

Lemma 6. *Let $p \in [1, \infty)$. Then, for $X, Y \in L^p$ with $\rho(X) < \infty$ and $\rho(Y) < \infty$,*

$$\left|\rho(X) - \rho(Y)\right| \le c_p \|X - Y\|_p. \tag{8}$$

Proof. If $c_p = \infty$ then (8) is trivial. So we assume that $c_p < \infty$.

By the subadditivity and monotonicity of ρ, for $X, Y \in L^p$ with $\rho(X) < \infty$ and $\rho(Y) < \infty$, we have

$$\rho(X) = \rho(X - Y + Y) \le \rho(X - Y) + \rho(Y) \le \rho(-|X - Y|) + \rho(Y).$$

Thus we see

$$|\rho(X) - \rho(Y)| \le \int_{(0,1]} \text{AVaR}_\lambda(-|X - Y|)\mu(d\lambda).$$

Now, consider the case $p = 1$. From (2) we find that $\text{AVaR}_\lambda(-|X - Y|) \le (1/\lambda)$ $\mathbb{E}|X - Y|$. Thus (8) follows.

For the case $p > 1$, we use Chebyshev's inequality to get $\mathbb{P}(|X - Y| > y) \le y^{-p}\|X - Y\|_p^p$. From this, we see $\{y : y^{-p}\|X - Y\|_p^p \le \alpha\} \subset \{y : \mathbb{P}(|X - Y| > y) \le \alpha\}$. Thus,

$$\text{VaR}_\alpha(-|X - Y|) \le \inf\{y > 0 : y^{-p}\|X - Y\|_p^p \le \alpha\} = \alpha^{-1/p}\|X - Y\|_p, \ \alpha \in (0,1).$$

Therefore,

$$\text{AVaR}_\alpha(-|X - Y|) \le \frac{\|X - Y\|_p}{\alpha} \int_0^\alpha \lambda^{-1/p} d\lambda = \frac{p\alpha^{-1/p}}{p - 1} \|X - Y\|_p.$$

Thus (8) follows. □

Our plan is to obtain the best discretization $X^{(n)}$ of X with respect to L^p-norm and then to approximate $\rho(-f(X))$ by $\rho(-f(X^{(n)}))$ with formula (7). We shall use the *optimal quantization theory* for probability distributions to obtain such discretization $X^{(n)}$. We will briefly explain this theory below, and refer to Graf and Luschgy (2000) and the references therein for more details.

Let $p \in [1, \infty)$ be fixed and $X = (X_1, \dots, X_d)$ be an $\mathbb{R}^d$-valued random variable with each component belonging to L^p. The random variable $X^{(n)}$ of the form $X^{(n)} = h(X)$ is called an *n-quantizer* of X if $h : \mathbb{R}^d \to \mathbb{R}^d$ is Borel measurable and $\# h(\mathbb{R}^d) \leq n$, where $\# A$ denotes the number of elements of a set A. Thus $X^{(n)}$ is an approximation of X with at most n-discretizing points. The n-th quantization error for X of order p is defined by

$$V_{n,p}(X) = \inf\{\mathbb{E}[|X - h(X)|^p] \,|\, h : \mathbb{R}^d \to \mathbb{R}^d \text{ is Borel measurable with } \# h(\mathbb{R}^d) \leq n\}.$$

An n-quantizer $\hat{X}^{(n)} = \hat{h}(X)$ is said to be *optimal* if it satisfies

$$V_{n,p}(X) = \mathbb{E}[|X - \hat{h}(X)|^p].$$

By Lemma 3.1 in Graf and Luschgy (2000), the n-th quantization error $V_{n,p}$ is given by

$$V_{n,p}(X) = \inf\{Q_n^p(\{x_1, \dots, x_n\}) | x_i \in \mathbb{R}^d, \, i = 1, \dots, n\},$$

where

$$Q_n^p(\{x_1, \dots, x_n\}) = \mathbb{E}\Big[\min_{1 \leq i \leq n} |X - x_i|^p\Big].$$

If $\Gamma = \{\hat{x}_1, \dots, \hat{x}_n\}$ is a minimizer of $\{x_1, \dots, x_n\} \mapsto Q_n^p(\{x_1, \dots, x_n\})$, then an optimal n-quantizer $\hat{X}^{(n)}$ of X is given by

$$\hat{X}^{(n)} = \sum_{i=1}^{n} \hat{x}_i \, 1_{C_i(\Gamma)}(X), \tag{9}$$

where $C_i(\Gamma)$, $i = 1, \dots, n$, are Borel measurable subsets in $\mathbb{R}^d$ such that

$$\mu_X(C_i(\Gamma) \cap C_j(\Gamma)) = 0 \quad (i \neq j), \quad \mu_X\Big(\mathbb{R}^d \setminus \cup_{i=1}^n C_i(\Gamma)\Big) = 0,$$

$$C_i(\Gamma) \subset \Big\{x \in \mathbb{R}^d : |x - \hat{x}_i| = \min_{1 \leq j \leq n} |x - \hat{x}_j|\Big\} \;\; \mu_X\text{-a.e., } i = 1, \dots, n.$$

The partition $\{C_i(\Gamma)\}_{i=1}^n$ is usually called a *Voronoi tessellation* of $\mathbb{R}^d$ with respect to Γ and μ_X.

It should be mentioned that the function Q_n^p is not convex in general. This means that a global minimum is not easily constructed by the differentiation of Q_n^p. However, we have from Theorem 4.12 in Graf and Luschgy (2000) the existence of a global minimum for $\{x_1, \dots, x_n\} \mapsto Q_n^p(\{x_1, \dots, x_n\})$ in the case that μ_X has an infinite support.

In addition, Theorem 6.2 in Graf and Luschgy (2000) tells us that if μ_X is not singular with respect to the Lebesgue measure and if $\mathbb{E}[|X|^{p+\delta}] < \infty$ for some $\delta > 0$, then

$$V_{n,p}(X) \sim J_{p,d} K_{p,d,X} n^{-p/d}, \quad n \to \infty,$$

where

$$K_{p,d,X} = \left(\int_{\mathbb{R}^d} \varphi^{d/(d+p)}(u) du\right)^{1+p/d}. \tag{10}$$

Here $\varphi(x)$ is the Radon-Nikodym density of the absolutely continuous part of μ_X with respect to the Lebesgue measure, and $J_{p,d}$ is a positive constant depending only on p and d. For examples, it is known that $J_{p,1} = 2^{-p}(p+1)^{-1}$ and $J_{2,2} = 5/(18\sqrt{3})$.

Simply applying the results above, Lemmas 3 and 6, we have the following.

Theorem 7. *Let $f : \mathbb{R}^d \to \mathbb{R}$ be a Lipschitz continuous function and $X = (X_1, \dots, X_d)$ be an $\mathbb{R}^d$-valued random variable. Suppose that μ_X is not singular with respect to the Lebesgue measure and that $\mathbb{E}[|X|^{p+\delta}] < \infty$ for some $\delta > 0$. Let $\hat{X}^{(n)}$ be an optimal n-quantizer as in* (9) *and let $\tau : \{1, \dots, n\} \to \{1, \dots, n\}$ be such that $f(\hat{x}_{\tau(1)}) \le \cdots \le f(\hat{x}_{\tau(n)})$. Then we have*

$$\rho(-f(\hat{X}^{(n)})) = f(\hat{x}_{\tau(1)}) + \sum_{i=2}^{n} (f(\hat{x}_{\tau(i)}) - f(\hat{x}_{\tau(i-1)})) \int_{(0,1]} \frac{1}{\lambda} \left(\mu_X \left(\cup_{k=i}^{n} C_{\tau(k)}(\Gamma)\right) \wedge \lambda\right) \mu(d\lambda)$$

and

$$\overline{\lim_{n\to\infty}}\, n^{1/d} \left| \rho(-f(X)) - \rho(-f(\hat{X}^{(n)})) \right| \le K_f c_p J_{p,d}^{1/p} K_{p,d,X}^{1/p}$$

where K_f is a Lipschitz coefficient of f and $K_{p,d,X}$ is given by (10).

In the one-dimensional case, i.e., the case $d = 1$, a minimizer

$$\Gamma = \{\hat{x}_1, \dots, \hat{x}_n\} = \operatorname{argmin}\{Q_n^p(\{x_1, \dots, x_n\}) : x_i \in \mathbb{R},\ i = 1, \dots, n\}$$

exists uniquely if μ_X is a strongly unimodal (see Graf and Luschgy, 2000). It is straightforward to see that a Voronoi tessellation $\{C_j(\Gamma)\}$ is given by

$$C_{\tau(1)}(\Gamma) = \left(-\infty, \frac{\hat{x}_{\tau(1)} + \hat{x}_{\tau(2)}}{2}\right], \quad C_{\tau(N)}(\Gamma) = \left(\frac{\hat{x}_{\tau(N-1)} + \hat{x}_{\tau(N)}}{2}, \infty\right),$$

$$C_{\tau(j)}(\Gamma) = \left(\frac{\hat{x}_{\tau(j-1)} + \hat{x}_{\tau(j)}}{2}, \frac{\hat{x}_{\tau(j)} + \hat{x}_{\tau(j+1)}}{2}\right] \quad (j = 2, \dots, N-1),$$

provided that $\hat{x}_{\tau(1)} \le \cdots \le \hat{x}_{\tau(n)}$.

Corollary 8. *Suppose that* $d = 1$ *and* $c_p < \infty$. Let $X \in L^{p+\delta}$ *for some* $\delta > 0$. *If* X *has a non-singular distribution with respect to the Lebesgue measure, then*

$$\left|\rho(-X) - \hat{x}_{\tau(1)} - \sum_{i=2}^{n} (\hat{x}_{\tau(i)} - \hat{x}_{\tau(i-1)}) \int_{(0,1]} \frac{1}{\lambda} \left(\mathbb{P}\left(X > \frac{\hat{x}_{\tau(i-1)} + \hat{x}_{\tau(i)}}{2} \right) \wedge \lambda \right) \mu(d\lambda) \right| = O\left(\frac{1}{n}\right),$$

as $n \to \infty$.

To implement the above approximation method, we need to solve the minimization problem of $\{x_1, \ldots, x_n\} \mapsto Q_n^p(\{x_1, \ldots, x_n\})$, and then to compute $\mu_X(C_j(\Gamma))$, $j = 1, \ldots, n$. Following Pagès (1997) and Pagès et al. (2004), we describe a stochastic approximation algorithm for the implementation. Let $\{\xi^\nu\}_{\nu=1}^{\infty}$ be an i.i.d. sequence with common distribution μ_X. We take an initial grid $\Gamma^0 = \{x^{1,0}, \ldots, x^{d,0}\}$ consisting of pairwise distinct components, and then construct the sequences $\{i(\nu)\}_{\nu=1}^{\infty}$ and $\{x^{i,\nu}\}_{\nu=1}^{\infty}$, $i = 1, \ldots, n$, as follows:

$$i(\nu+1) = \operatorname*{argmin}_{1 \le i \le n} \left| x^{i,\nu} - \xi^{\nu+1} \right|,$$

$$x^{i,\nu+1} = \begin{cases} x^{i,\nu} - a_{\nu+1} \dfrac{x^{i,\nu} - \xi^{\nu+1}}{\left| x^{i,\nu} - \xi^{\nu+1} \right|^{2-p}} & \text{if } i = i(\nu+1), \\ x^{i,\nu} & \text{if } i \neq i(\nu+1). \end{cases}$$

Here $\{a_\nu\}_{\nu=1}^{\infty}$ is a sequence of positive numbers such that $\sum_{\nu=1}^{\infty} a_\nu = +\infty$ and $\sum_{\nu=1}^{\infty} a_\nu^2 < +\infty$. Set $\Gamma^\nu = \{x^{1,\nu}, \ldots, x^{n,\nu}\}$ for $\nu \in \mathbb{N}$. Then we notice that $i(\nu+1)$ satisfies $\xi^{\nu+1} \in C_{i(\nu+1)}(\Gamma^\nu)$. Under some conditions, the sequences $\{Q_n^{p,\nu}\}_{\nu=1}^{\infty}$ and $\{\pi_i^\nu\}_{\nu=1}^{\infty}$, $i = 1, \ldots, n$, constructed by

$$Q_n^{p,\nu+1} = Q_n^{p,\nu} - a_{\nu+1}\left(Q_n^{p,\nu} - \left| x^{i(\nu+1),\nu} - \xi^{\nu+1} \right|^p \right), \; Q_n^{p,0} = 0,$$

$$\pi_i^{\nu+1} = \pi_i^\nu - a_{\nu+1}(\pi_i^\nu - 1_{\{i(\nu+1)\}}(i)), \; \pi_i^0 = \frac{1}{n}, \quad 1 \le i \le n,$$

satisfy

$$Q_n^{p,\nu} \to Q_n^p(\Gamma^*), \; \pi_i^\nu \to \mu_X(C_i(\Gamma^*)), \quad i = 1, \ldots, d,$$

as $\nu \to \infty$ almost surely on the event that $\Gamma^\nu(\omega)$ converges to a local minimum $\Gamma^*(\omega)$. A sufficient condition for this convergence is the compactness of the support of μ_X. This is a very strong condition because we are usually interested with a distribution with unbounded support in risk measurement. However, in the computer simulation, we are necessarily restricted to bounded random variables, so the compactness condition is not essential from numerical view point. Also, when $p = 2$, almost sure convergence results can be obtained under additional regularity conditions for μ_X. We refer to Pagès and Printems (2003) for the numerical heuristics of the implementation methods here with $p = 2$ for the case of the Gaussian random variables.

Example 9. Consider the case where the loss random variable X has the log-normal distribution with scale parameter one and location parameter zero, i.e., the case that $\log(X)$ follows the standard normal distribution. As a simple illustration of our method, we compute $\mathrm{AVaR}_{0.01}(-X)$. Denote by e_n the relative error between the analytical value of $\mathrm{AVaR}_{0.01}(-X)$ and the approximation of this risk measure given in Corollary 8 via the n-th quantization with respect to L^2-norm. Then, the implementation method above with $a_v = 1/(1+v)$ gives $e_{10} \approx 0.432262$, $e_{20} \approx 0.106573$, $e_{50} \approx 0.010449$ and $e_{100} \approx 0.004964$, for examples.

REMARK 10. Monte Carlo approach for computing $\mathrm{AVaR}_\alpha(-X)$ is based on the fact

$$\lim_{N\to\infty} \frac{\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{i,N}}{\lfloor N\alpha \rfloor} = \mathrm{AVaR}_\alpha(-X) \quad \text{a.s.},$$

where $\{X_i\}$ is an i.i.d sequence with common distribution μ_X, $\lfloor x \rfloor$ denotes the integer part of x, and $X_{i,N}$'s are the order statistics such that $X_{1,N} \geq \cdots \geq X_{N,N}$. This result can be found in Chapter 2 in McNeil et al. (2005) for examples. As is usual in Monte Carlo methods, the estimator $\sum_{i=1}^{\lfloor N\alpha \rfloor} X_{i,N} / \lfloor N\alpha \rfloor$ depends on replications of a simulation. Thus variance is usually adopted as a figure of merit for Monte Carlo approach.

On the other hand, for the quantization approach, once quantization points for a targeted distribution are obtained, these values can be used in every computation. So the approximation error of a risk measure is independent of the number of trials, and does depend on the number of quantization points. Therefore, the comparison of the quantization approach with Monte Carlo one is less straightforward. This situation is similar to the one in the comparison of quasi Monte Carlo and Monte Carlo methods.

REMARK 11. It should be noticed that the comonotonicity and the L^p-estimate are essential for our approach. The proposed method can be applicable only to comonotonic coherent risk measures with L^p-Lipschitz continuity. To employ a quantization method for other classes of risk measures, we need to use different criteria for the quantization error. For examples, every coherent risk measure is Lipschitz continuous with respect to L^∞-norm: $\operatorname{ess\,sup}_{\omega\in\Omega} |X(\omega) - X^{(n)}(\omega)|$. Thus, we can use this norm to obtain a quantized random variable. However, this approach of course excludes unbounded random variables. Another possibility is to use Ky Fan metric: $\inf\{\varepsilon > 0 : \mathbb{P}(|X - X^{(n)}| > \varepsilon) \leq \varepsilon\}$. Anyway, we will face numerical difficulties in the quantization.

REMARK 12. Let us discuss the case of the value-at-risk. It is well-known that this risk measure is also comonotonic but not subadditive (see, e.g., Example 4.41, Lemma 4.84, and Remark 4.85 in Föllmer and Schied, 2004). Thus we cannot rely on the estimate as in Lemma 6. However, for a fixed small $\varepsilon > 0$ we observe

$$\left|\mathbb{P}(X > x) - \mathbb{P}(X^{(n)} > (1-\varepsilon)x)\right| \le \mathbb{P}\left(\left|X - X^{(n)}\right| > \varepsilon x\right) \le \frac{1}{(\varepsilon x)^p} \| X - X^{(n)} \|_p^p.$$

So if we know *a priori* a compact set A that contains $\mathrm{VaR}_\alpha(-X)$, $\mathrm{VaR}_{\alpha-\varepsilon'}(-X^{(n)})/(1-\varepsilon)$ and $\mathrm{VaR}_{\alpha+\varepsilon'}(-X^{(n)})/(1-\varepsilon)$ for sufficiently large n and for some small $\varepsilon' > 0$, then there exists n_0 such that

$$\sup_{x\in A} \left|\mathbb{P}(X > x) - \mathbb{P}(X^{(n)} > (1-\varepsilon)x)\right| \le \varepsilon', \quad n \ge n_0.$$

From this we find that

$$\left\{x \in A : \mathbb{P}(X^{(n)} > (1-\varepsilon)x) \le \alpha - \varepsilon'\right\} \subset \left\{x \in A : \mathbb{P}(X > x) \le \alpha\right\}$$
$$\subset \left\{x \in A : \mathbb{P}(X^{(n)} > (1-\varepsilon)x) \le \alpha + \varepsilon'\right\},$$

leading to

$$\frac{1}{1-\varepsilon}\,\mathrm{VaR}_{\alpha+\varepsilon'}(-X^{(n)}) \le \mathrm{VaR}_\alpha(-X) \le \frac{1}{1-\varepsilon}\,\mathrm{VaR}_{\alpha-\varepsilon'}(-X^{(n)}),$$

for $n \ge n_0$.

ACKNOWLEDGEMENT

The author would like to thank the anonymous referees for their useful comments on previous versions of this paper.

REFERENCES

ACERBI, C. and TASCHE, D. (2002) On the coherence of expected shortfall. *J. Banking Finance* **26**, 1487-1503.

ARTZNER, P., DELBAEN, F., EBER, J.-M. and HEATH, D. (1999) Coherent measurement of risk. *Math. Finance* **9**, 203-228.

DELBAEN, F. (2002) Coherent measures of risk on general probability spaces. In *Advances in Finance and Stochastics, Essays in Honor of Dieter Sondermann* (eds. K. Sandmann and P. J. Schönbucher), pp. 1-37, Springer, Berlin.

FEUERVERGER, A. and WONG, A.C. (2000) Computation of Value at Risk for nonlinear portfolios. *J. Risk* **3**, 37-55.

FÖLLMER, H. and SCHIED, A. (2004) *Stochastic Finance: An Introduction in Discrete Time*. 2nd ed., Walter de Gruyter, Berlin.

GLASSERMAN, P. (2004) *Monte Carlo methods in financial engineering*. Springer, New York.

GRAF, S. and LUSCHGY, H. (2000) *Foundation of quantization for probability distributions*. Springer, Berlin.

INOUE, A. (2003) On the worst conditional expectation. *J. Math. Anal. Appl.* **286**, 237-247.

KAAS, R., GOOVAERTS, M., DHAENE, J. and DENUIT, M. (2001) *Modern actuarial risk theory*. Kluwer, Boston.

KAINA, M. and RÜSCHENDORF, L. (2009) On convex risk measures on L^p-spaces. *Math. Methods Oper. Res.* **69**, 475-495.

KUSUOKA, S. (2001) On law invariant coherent risk measures. *Adv. Math. Econ.* **3**, 83-95.

LANDSMAN, Z.M. and VALDEZ, E.A. (2003) Tail conditional expectations for elliptical distributions. *N. Amer. Actuarial J.* **7**, 55-71.

MCNEIL, A.J., FREY, R. and EMBRECHTS, P. (2005) *Quantitative risk management: concepts, techniques and tools*. Princeton University Press, Princeton.

PAGÈS, G. (1997) A space quantization method for numerical integration. *J. Comput. Appl. Math.* **89**, 1-38.

PAGÈS, G., PHAM, H. and PRINTEMS, J. (2004) An optimal Markovian quantization algorithm for multidimensional stochastic control problems. *Stoch. Dyn.* **4**, 501-545.

PAGÈS, G. and PRINTEMS, J. (2003) Optimal quadratic quantization for numerics: the Gaussian case. *Monte Carlo Methods Appl.* **9**, 135-168.

YAMAI, Y. and YOSHIBA, T. (2002) Comparative analyses of expected shortfall and value-at-risk: their estimation error, decomposition, and optimization. *Monet. Econ. Stud.* **20**, 87-121.

YOUNG, V.R. (2004) Premium principles. In *Encyclopedia of actuarial sciences*. Wiley, New York.

YUMIHARU NAKANO
Graduate School of Innovation Management
Tokyo Institute of Technology
2-12-1 W9-117 Ookayama 152-8552 Tokyo
Japan