Noname manuscript No.
(will be inserted by the editor)

A Generalized Neyman-Pearson Lemma for g-Probabilities

Shaolin Ji · Xun Yu Zhou

Received: date / Accepted: date

Abstract This paper is concerned with hypothesis tests for g-probabilities, a class of nonlinear probability measures. The problem is shown to be a special case of a general stochastic optimization problem where the objective is to choose the terminal state of certain backward stochastic differential equations so as to minimize a g-expectation. The latter is solved with a stochastic maximum principle approach. Neyman–Pearson type results are thereby derived for the original problem with both simple and randomized tests. It turns out that the likelihood ratio in the optimal tests is nothing else than the ratio of the adjoint processes associated with the maximum principle. Concrete examples, ranging from the classical simple tests, financial market modelling with ambiguity, to super- and sub-pricing of contingent claims and to risk measures, are presented to illustrate the applications of the results obtained.

Keywords backward stochastic differential equation · g-probability/expectation · hypothesis test · Neyman–Pearson lemma · stochastic maximum principle

Mathematics Subject Classification (2000) 60H10 · 60H30 · 93E20

The first author is supported in parts by the National Basic Research Program of China (973 Program, No. 2007CB814901 and the National Natural Science Foundation of China No. 10871118). The second author owes thanks for aid to RGC Grants #CUHK418605 and #CUHK418606, and a start-up fund of the University of Oxford.

S. Ji
School of Mathematics, Shandong University, Jinan, China, 250100
Tel.: +86-531-88364100
Fax: +86-531-88564100
E-mail: jsl@sdu.edu.cn

X. Zhou
Mathematical Institute, The University of Oxford, 24–29 St Giles, Oxford OX1 3LB, UK,
and
Department of Systems Engineering and Engineering Management, The Chinese University of Hong Kong, Shatin, Hong Kong
Tel.: +44-1865-280614
Fax: +44-1865-270515
E-mail: zhouxy@maths.ox.ac.uk

1 Introduction

A classical simple statistical hypothesis testing problem is to design a test so as to minimize the probability of Type II error while confining that of Type I error to a given acceptable significance level. The Neyman–Pearson lemma gives the most powerful test for such a problem via a critical level of the likelihood ratio; see, e.g., [9] and [35]. It is a natural yet challenging problem to extend this lemma to simple hypothesis testing for nonlinear probabilities. Huber and Strassen [23] studied an extended Neyman–Pearson lemma for the so-called 2-alternating capacities (or convex capacities as typically called in the economics literature; see for example [4], [10], [23] and [36]). But little has been known about the non 2-alternating case in the context of hypothesis tests.

The g-probability, induced by the g-expectation, a nonlinear expectation introduced by Peng [33] via a nonlinear backward stochastic differential equation (BSDE), is an example of the non 2-alternating capacity. Both the g-probability and g-expectation, apart from their own theoretical values, have found important applications in various areas especially in finance. For example, the super- and sub-pricing of contingent claims in an incomplete market can both be captured by the g-probability; see, e.g., Chen and Kulperger [7]. Ambiguity in financial modelling can be described by the g-expectation. Indeed, Chen and Epstein [6] introduced a k-ignorance model involving the g-probability to study ambiguity aversion. The g-probabilities/expectations have also been found to have intimate connection with the rapidly developed risk measure theory. Recently the so-called coherent, sublinear and convex risk measures have been proposed (see [1], [18], [19], [20], [21] and the references therein). In Gianin [22] interesting relations between the coherent/convex risk measures and the g-expectations are revealed. Coquet et al. [8] further showed that the relation can be extended to the dynamic setting; hence dynamic coherent or convex risk measures can be formulated via g-expectations. More applications of g-probabilities/expectations can be found in, e.g., [5], [6], [22], [34] and the references therein.

In view of its wide applications, it is an important problem to investigate the hypothesis tests for g-probabilities. In this paper, we study a simple hypothesis testing problem for g-probability measures. More precisely, we test a simple hypothesis H_0: $g = g^0$ versus a simple alternative hypothesis H_1: $g = g^1$ where g^0 and g^1 are generating functions of two possible g-probabilities. This problem is a generalization of the classical hypothesis testing involving the (Kolmogorov) probability measure; yet very different from that of Huber and Strassen [23] because a g-probability is inherently not 2-alternating.

A key idea of solving the problem is to "embed" it into an optimal stochastic control problem. This is very natural because the BSDEs defining the g-probabilities involved readily provide the state equations. The question is what the "control variable" should be. We take the test (the random variable serving as the terminal value of a BSDE) as the control variable, and then use a variational technique to obtain a stochastic maximum principle, i.e., a first-order necessary condition that characterizes the optimal test. This "terminal perturbation technique" was indeed first introduced by El Karoui, Peng and Quenez [14] in order to solve a recursive utility optimization problem, and employed by Ji and Peng [24] (along with Ekeland's variational principle) to obtain a necessary condition for a mean–variance portfolio selection problem with non-convex wealth equations. For its systematic application in stochastic control with state constraints, we refer the reader to a recent paper by Ji and Zhou [25]. The present paper, however, is the first to apply this technique to statistical hypothesis tests.

In our approach, we consider a general stochastic optimization problem where the objective is to choose the terminal state of certain BSDEs so as to minimize a g-expectation subject to another g-expectation bounded by a given level. This problem covers the testing problem for g-probabilities as a special case. We then employ a stochastic maximum principle approach to derive a necessary condition for the optimal randomized tests. Furthermore, under the convexity assumption, we prove that the established necessary condition is sufficient. We also show that the optimal tests can be solved by some forward-backward stochastic differential equations (FBSDEs) with constraints. These necessary/sufficient conditions resemble in form the classical Neyman–Pearson lemma. More interestingly, the likelihood ratio in the optimal tests is nothing else than the ratio of the adjoint processes associated with the maximum principle.

We finally present five concrete applications, some of which were indeed part of the original motivation of this research, to illustrate the general results obtained. These include the classical Neyman–Pearson lemma, the design of a financial portfolio to learn a financial market with ambiguity, super- and sub-pricing of contingent claims, and the minimization of the shortfall risk in hedging a risky position.

Parts of the results in this paper have been announced (without proofs) in Ji and Zhou [26].

This paper is organized as follows: In section 2, we formulate the hypothesis testing problem for g-probabilities. Some auxiliary stochastic optimization problems are introduced and solved in section 3. In section 4, we use the general results of section 3 to derive the Neyman–Pearson type lemmas for g-probabilities. Several examples are presented in section 5. Proofs of main results are put in Appendices A and B.

2 Problem Formulation

Let $W(\cdot) = \{W(t):\ 0 \leq t \leq T\}$ be a standard d-dimensional Brownian motion defined on a complete probability space $(\Omega, \mathcal{F}, P)$. The information structure is given by a filtration $F = \{\mathcal{F}_t\}_{0\leq t\leq T}$, which is the augmented $\sigma-$algebra generated by $W(\cdot)$, with $\mathcal{F}_T = \mathcal{F}$. Denote by $L^2_{\mathcal{F}}(0, T; R^n)$ the space of all $\mathcal{F}_t-$adapted processes $x(\cdot)$ with values in R^n such that $E\int_0^T \mid x(t) \mid^2 dt < \infty$, and by $L^2(\Omega, \mathcal{F}_T, P)$ the set of all $\mathcal{F}_T$-measurable random variables ξ with value in R^1 such that $E \mid \xi \mid^2 < \infty$.

Associated with a function

$$g = g(y, z, t)\colon R^1 \times R^{1\times d} \times [0, T] \to R^1$$

we introduce the following conditions:

(H1) g is uniformly Lipschitz in (y, z);

(H2) $g(y, z, t)$ is continuous in t and $\int_0^T g^2(0, 0, t)dt < \infty$;

(H3) $g(y, 0, t) \equiv 0\ \forall (y, t) \in R^1 \times [0, T]$.

For any $\xi \in L^2(\Omega, \mathcal{F}_T, P)$ and g satisfying (H1)-(H2), the following nonlinear backward stochastic differential equation (BSDE)

$$y(t) = \xi + \int_t^T g(y(s), z(s), s)ds - \int_t^T z(s)dW_s,\ \ 0 \leq t \leq T \tag{2.1}$$

has a unique solution $(y(\cdot), z(\cdot)) \in L^2_{\mathcal{F}}(0, T; R^1) \times L^2_{\mathcal{F}}(0, T; R^{1\times d})$; see [32] and [38]. This leads to the following definition of g-expectation and g-probability.

Definition 2.1 *([33]) Given g satisfying (H1)-(H3) and $\xi \in L^2(\Omega, \mathcal{F}_T, P)$, let $(y(\cdot), z(\cdot))$ be the solution of (2.1). The g-expectation of ξ, denoted by $\mathcal{E}_g[\xi]$, is defined as $\mathcal{E}_g[\xi] = y(0)$. Furthermore, for any $A \in \mathcal{F}_T$, the g-probability of A, denoted by $P_g(A)$, is defined as $P_g[A] = \mathcal{E}_g[1_A]$.*

Such a g-expectation is a nonlinear expectation and a g-probability is a non-additive probability measure. They certainly depend on the particular choice of the function g, which is called the *generating function* or *generator*, serving as a parameter, of the expectation and the probability. Note that under (H3), for any constant c one has $\mathcal{E}_g[c] = c$.

We now introduce hypothesis tests for g-probabilities. Let G be a given set of generating functions satisfying (H1)-(H3), and $G_0 \subset G$ with both G_0 and $G \setminus G_0$ non-empty. G is called the set of *admissible hypotheses*, and G_0 (resp. $G \setminus G_0$) called a *null hypothesis* H_0 (resp. *alternative hypothesis* H_1). If G_0 (resp. $G \setminus G_0$) consists of only one element, then it is called *simple*, otherwise *composite*. In this paper, we are concerned with tests with a simple null hypothesis H_0: $g = g^0$ versus a simple alternative H_1: $g = g^1$. It should be noted that there is a rich theory on simple hypothesis tests in the classical statistics literature, including the Neyman-Pearson fundamental lemma.

Specifically, a statistical test is defined as follows.

Definition 2.2 *Given g^0 and g^1 satisfying (H1)-(H3). A test for a simple null hypothesis H_0: $g = g^0$ versus a simple alternative H_1: $g = g^1$ is an $\mathcal{F}_{\mathrm{T}}$-measurable random variable $\xi\colon \Omega \to \{0, 1\}$, which rejects H_0 on the event $\{\xi = 1\}$. For $0 < \alpha < 1$, the test is said to have a significance level α if $P_{g^0}(\xi = 1) \leq \alpha$.*

For an outcome of the test, i.e., a sample ω from the sample space Ω, the hypothesis g^0 is rejected (resp. accepted) if $\omega \in \{\xi = 1\}$ (resp. $\omega \in \{\xi = 0\}$). Thus, $P_{g^0}(\xi = 1)$ is the g-probability of *Type I error* (i.e. that of rejecting H_0 when it is true), whereas $P_{g^1}(\xi = 0)$ the g-probability of *Type II error* (i.e. that of accepting H_0 when it is false). Naturally, we need to design a test that keeps both types of error as small as possible. However, these two types of error are often conflicting in the sense that usually the smaller the one type the larger the other, and vice versa. Hence a common way is to find a test that minimizes the g-probability of Type II error, among all tests that keep the g-probability of Type I error below the given acceptable significance level $\alpha \in (0, 1)$. In this spirit, we formulate the following problem

$$\begin{aligned} &\underset{\xi \in \mathcal{L}}{\text{Minimize}} \quad P_{g^1}(\xi = 0), \\ &\text{subject to} \quad P_{g^0}(\xi = 1) \leq \alpha, \end{aligned} \tag{2.2}$$

where

$$\mathcal{L} = \left\{ \xi \mid \xi \in L^2(\Omega, \mathcal{F}_T, P) \text{ and } \xi \in \{0, 1\} \text{ a.s.} \right\}.$$

As with the classical Neyman-Pearson lemma, one needs also to consider the randomized tests (see [9], [15], [28] and [37]).

Definition 2.3 *Given g^0 and g^1 satisfying (H1)-(H3). A randomized test for a simple null hypothesis H_0: $g = g^0$ versus a simple alternative H_1: $g = g^1$ is an $\mathcal{F}_T$-measurable random variable $\xi\colon \Omega \to [0, 1]$, which rejects H_0 with probability $\xi(\omega)$ on outcome ω. For $0 < \alpha < 1$, the test is said to have a significance level α if $\mathcal{E}_{g^0}[\xi] \leq \alpha$.*

The requirement that $\mathcal{E}_{g^0}[\xi] \leq \alpha$ guarantees that the hypothesis H_0, if it is true, is rejected "on average" with at most α. In other words, with such a randomized test ξ, $\mathcal{E}_{g^0}[\xi]$ represents the averaged Type I error while $\mathcal{E}_{g^1}[1-\xi]$ represents the averaged Type II error.

Set

$$\bar{\mathcal{L}} = \left\{ \xi \mid \xi \in L^2(\Omega, \mathcal{F}_T, P) \text{ and } 0 \leq \xi \leq 1 \text{ a.s.} \right\}.$$

For a given acceptable significance level $\alpha \in (0,1)$, a randomized test is

$$\begin{array}{ll} \underset{\xi \in \bar{\mathcal{L}}}{\text{Minimize}} & \mathcal{E}_{g^1}[1-\xi], \\ \text{subject to} & \mathcal{E}_{g^0}[\xi] \leq \alpha. \end{array} \tag{2.3}$$

3 General Stochastic Optimization Problems

In this section, we introduce and solve some auxiliary and general stochastic optimization problems that include our hypothesis tests formulated in the previous section as special cases. These problems are interesting in their own rights.

3.1 A problem with convex constraints

We start with a problem corresponding to the randomized test problem. Set

$$U = \{\xi \mid \xi \in L^2(\Omega, \mathcal{F}_T, P) \text{ and } \xi \in K \quad a.s.\}$$

where K is a bounded closed convex subset in R^1. In fact, in this case, $K = [k_1, k_2]$ where $k_1 \leq k_2$. Notice that the boundary of K, $\partial K = \{k_1, k_2\}$, is nonconvex when $k_1 < k_2$.

For a given constant α such that $\mathcal{E}_{g^0}[k_1] < \alpha < \mathcal{E}_{g^0}[k_2]$, consider the following stochastic optimization problem:

$$\begin{array}{ll} \underset{\xi \in U}{\text{Minimize}} & \mathcal{E}_{g^1}[h(\xi)], \\ \text{subject to} & \mathcal{E}_{g^0}[\xi] \leq \alpha \end{array} \tag{3.1}$$

where $h(y) : R^1 \to R^1$ is a given function, and g^0 and g^1 two given generating functions leading to the following BSDEs:

$$\begin{cases} -dy_0(t) = g^0(y_0(t), z_0(t), t)dt - z_0(t)dW(t), \\ y_0(T) = \xi, \\ -dy_1(t) = g^1(y_1(t), z_1(t), t)dt - z_1(t)dW(t), \\ y_1(T) = h(\xi). \end{cases} \tag{3.2}$$

The following additional assumption, which is stronger than (H1), is introduced.

(H4) g^0 and g^1 (resp. h) are continuously differentiable in (y, z) (resp. y) and their derivatives are uniformly bounded.

We start with necessary conditions of an optimal solution to (3.1). Let ξ^* be optimal to (3.1) with $(y_0^*(\cdot), z_0^*(\cdot), y_1^*(\cdot), z_1^*(\cdot))$ the solution to (3.2) corresponding to $\xi = \xi^*$. Denote the derivatives $g_y^i(t) = g_y^i(y_i^*(t), z_i^*(t), t)$ and $g_z^i(t) = g_z^i(y_i^*(t), z_i^*(t), t)$, $i = 0, 1$. The following is the main result of this subsection.

Theorem 3.1 *Assume (H2)-(H4). If ξ^* is an optimal solution to (3.1), then ξ^* must be of the following form*

$$\xi^* = k_1 1_{\{h_1 m(T) > -h_0 h_y(\xi^*) n(T)\}} + b 1_{\{h_1 m(T) = -h_0 h_y(\xi^*) n(T)\}} + k_2 1_{\{h_1 m(T) < -h_0 h_y(\xi^*) n(T)\}} \tag{3.3}$$

where $h_1 \in R^1$, $h_0 \in R^1$ with $h_0 \geq 0$, b is a random variable with $b \in K$ a.s., and $(m(\cdot), n(\cdot))$ is the solution of the following adjoint equations:

$$\begin{cases} dm(t) = g_y^0(t)m(t)dt + g_z^0(t)'m(t)dW(t), & m(0) = 1, \\ dn(t) = g_y^1(t)n(t)dt + g_z^1(t)'n(t)dW(t), & n(0) = 1. \end{cases} \tag{3.4}$$

A proof of Theorem 3.1 is relegated to Appendix A.

Also we remark that from the proof in Appendix A the above theorem holds true when ξ is constrained in a random interval and/or when the function h is also random (i.e., $h = h(y, \omega)$ whereas $h(y, \cdot)$ is $\mathcal{F}_T$-measurable for each fixed y).

A natural question now is whether the form (3.3) is sufficient for an optimal solution. To address this we introduce the following convexity condition.

(H5) g^0 and g^1 (resp. h) are convex in (y, z) (resp. y), and h is a strictly decreasing function.

Remark: Under (H5) the following functionals defined on U by

$$\xi \to \mathcal{E}_{g^0}[\xi],$$
$$\xi \to \mathcal{E}_{g^1}[h(\xi)]$$

are both convex (see [14]). Note that in general if $\mathcal{E}_g[\cdot]$ is convex, then the corresponding generator g is independent of y (see [27, Theorem 3.2]) given Assumption (H3). On the other hand, Assumption (H3) is unnecessary for many results in this paper.

Let

$$V(\alpha) = \inf_{\xi \in \mathcal{A}(\alpha)} \mathcal{E}_{g^1}[h(\xi)] \tag{3.5}$$

where $\mathcal{A}(\alpha) = \{\xi \mid \xi \in U \text{ and } \mathcal{E}_{g^0}[\xi] \leq \alpha\}$. Applying the classical Lagrange approach in convex analysis (see [29]), we have

Lemma 3.2 *We assume (H1), (H2) and (H5). Then there exists a constant $v^* > 0$ such that*

$$V(\alpha) = \inf_{\xi \in U} \{\mathcal{E}_{g^1}[h(\xi)] + v^*(\mathcal{E}_{g^0}[\xi] - \alpha)\}. \tag{3.6}$$

Moreover, if the infimum is attained in (3.5) by ξ^*, then the infimum is attained in (3.6) by ξ^* with $\mathcal{E}_{g^0}[\xi^*] = \alpha$. Conversely, if there exist $v^o > 0$ and $\xi^o \in U$ such that the infimum is achieved in $\inf_{\xi \in U}\{\mathcal{E}_{g^1}[h(\xi)] + v^o(\mathcal{E}_{g^0}[\xi] - \alpha)\}$ with $\mathcal{E}_{g^0}[\xi^o] = \alpha$ then the infimum is achieved in (3.5) by ξ^o.

Theorem 3.3 *We assume (H2), (H4) and (H5). If ξ^* is optimal for (3.1), then there exists a constant $v > 0$ and a random variable $b \in K$ a.s. such that*

$$\xi^* = k_1 1_{\{vm(T) > -h_y(\xi^*) n(T)\}} + b 1_{\{vm(T) = -h_y(\xi^*) n(T)\}} + k_2 1_{\{vm(T) < -h_y(\xi^*) n(T)\}} \tag{3.7}$$

and $\mathcal{E}_{g^0}[\xi^*] = \alpha$, where $m(\cdot)$ and $n(\cdot)$ are solutions of the adjoint equations (3.4). Conversely, if $\xi^* \in U$ has the form (3.7) with $\mathcal{E}_{g^0}[\xi^*] = \alpha$, then ξ^* is optimal.

Proof. See Appendix B. ■

Remark: If the generators of the g-expectations are not smooth, i.e. (H4) does not hold, we can show that (3.7) is still a sufficient form for optimality, where the derivatives involved should be replaced by the subdifferentials (in the sense of convex analysis). Specifically, one may choose a proper measurable version of $h_y(\xi^*)$ in $\partial h(\xi^*)$, where ∂h is the subdifferential of h. Similarly, in the adjoint equations (3.4) $(g_y^i(y_i^*(t), z_i^*(t), t), g_z^i(y_i^*(t), z_i^*(t), t))$ should be interpreted as a vector-valued adapted process which belongs to $\partial g^i(y_i^*(t), z_i^*(t), t)$ $dP \otimes dt$ almost surely. Here ∂g^i is the subdifferential of g^i with respect to (y, z).

3.2 A problem with nonconvex constraints

Now we move on to the problem with nonconvex constraints, which is a generalized version of the original simple test problem. Set

$$\tilde{U} = \{\xi \mid \xi \in L^2(\Omega, \mathcal{F}_T, P) \text{ and } \xi \in \partial K \quad a.s.\}$$

where $K = [k_1, k_2]$, $\partial K = \{k_1, k_2\}$.

We are interested in the following stochastic optimization problem with (nonconvex) constraint:

$$\begin{aligned} &\underset{\xi \in \tilde{U}}{\text{Minimize}} \quad \mathcal{E}_{g^1}[h(\xi)], \\ &\text{subject to} \;\; \mathcal{E}_{g^0}[\xi] \le \alpha. \end{aligned} \tag{3.8}$$

The main difficulty here, of course, is that the constraint set $\tilde{U}$ is nonconvex; hence the usual convex perturbation approach fails. Thus it is difficult to derive a necessary condition of an optimal solution to the above problem. Our purpose is to derive a sufficient condition instead.

Theorem 3.4 *Under Assumptions (H2), (H4) and (H5), if there exist a constant $v > 0$ and a random variable $b \in \partial K$ a.s. such that*

$$\xi^* = k_1 1_{\{vm(T) > -h_y(\xi^*)n(T)\}} + b 1_{\{vm(T) = -h_y(\xi^*)n(T)\}} + k_2 1_{\{vm(T) < -h_y(\xi^*)n(T)\}} \tag{3.9}$$

with $\mathcal{E}_{g^0}[\xi^*] = \alpha$, where $m(\cdot)$ and $n(\cdot)$ are the solutions of the adjoint equation (3.4), then ξ^* is optimal.

Proof. The proof is almost the same as that proving the sufficiency part of Theorem 3.3; hence is omitted. ■

Remark: Again, a non-smooth version of Theorem 3.4 holds when g^0 and g^1 are not smooth, where the derivatives are replaced by the subdifferentials.

4 Solving Hypothesis Testing Problems

In this section, we apply the general results derived in the previous section to our hypothesis testing problems (2.2) and (2.3) respectively. Before doing so we introduce the following assumption, which is weaker than (H3):

(H3)' $\mathcal{E}_g[0] = 0$ and $\mathcal{E}_g[1] = 1$.

This assumption guarantees that $0 \le \mathcal{E}_g[\xi] \le 1$ $\forall \xi \in \bar{\mathcal{L}}$, based on the backward comparison theorem (El Karoui et al. [13, Theorem 2.2]). The g-expectation whose generator g satisfies (H1)-(H2) and (H3)' can be regarded as a generalized g-expectation.

4.1 Simple tests

We start with the simple testing problem (2.2), which can be rewritten as

$$\begin{aligned} &\inf_{\xi\in\mathcal{L}} \ \mathcal{E}_{g^1}[1-\xi], \\ &\text{subject to } \ \mathcal{E}_{g^0}[\xi] \le \alpha \end{aligned} \tag{4.1}$$

where $0<\alpha<1$.

Theorem 4.1 *Under Assumptions (H2), (H3)', (H4) and (H5), if there exist a constant $v>0$ and a random variable $b\in\{0,1\}$ a.s. such that*

$$\xi^* = 1_{\{vm(T)<n(T)\}} + b1_{\{vm(T)=n(T)\}} \tag{4.2}$$

with $\mathcal{E}_{g^0}[\xi^*]=\alpha$, where $m(\cdot)$ and $n(\cdot)$ are the solutions of adjoint equation (3.4), then the test ξ^* must be optimal.

Proof. Applying Theorem 3.4 with $h(y)=1-y$, $\partial K=\{0,1\}$ and $0<\alpha<1$, we obtain the desired result immediately. ■

Remark: The form (4.2) specifies a class of random variables that may possibly serve as an optimal simple test. The constant v and the random variable b could be determined by the condition $\mathcal{E}_{g^0}[\xi^*]=\alpha$ if one further specifies certain distribution for b. In particular, if $n(T)/m(T)$ has no atom, then (4.2) reduces to $\xi^*=1_{\{n(T)/m(T)>v\}}$, which is indeed in exactly the same form as the classical Neyman–Pearson lemma (where $n(T)/m(T)$ is the likelihood ratio).

4.2 Randomized tests

Let us now consider the randomized problem (2.3). The following is the necessary condition without the convexity restriction on the generators.

Theorem 4.2 *We assume (H2)–(H4). Let ξ^* be an optimal randomized test. Then there exist $h_1\in R^1$, $h_0\in R^1$ with $h_0\ge 0$ and a random variable $b\in[0,1]$ a.s. such that*

$$\xi^* = 1_{\{h_1 m(T)<h_0 n(T)\}} + b1_{\{h_1 m(T)=h_0 n(T)\}} \tag{4.3}$$

with $\mathcal{E}_{g^0}[\xi^*]=\alpha$, where $m(\cdot)$ and $n(\cdot)$ are solutions of the adjoint equations (3.4).

Proof. Noting that $h(y)=1-y$, $K=[0,1]$ and $0<\alpha<1$, (4.3) follows immediately from Theorem 3.1. That $\mathcal{E}_{g^0}[\xi^*]=\alpha$ is due to the strict comparison theorem of BSDE. In fact, if ξ^* is optimal and $\mathcal{E}_{g^0}[\xi^*]<\alpha$, then we can construct a randomized test $\hat\xi$ such that $\hat\xi\ge\xi^*$, $P(\hat\xi>\xi^*)>0$ and $\mathcal{E}_{g^0}[\hat\xi]=\alpha$. Such a construction is feasible due to the continuity of the solutions of BSDEs on parameters (El Karoui et al. [13, Proposition 2.4]) and the strict comparison theorem of BSDEs (El Karoui et al. [13, Theorem 2.2]). For this $\hat\xi$, we have $\mathcal{E}_{g^1}[1-\xi^*]>\mathcal{E}_{g^1}[1-\hat\xi]$, again by virtue of the strict comparison theorem of BSDEs, which contradicts the optimality of ξ^*. ■

Under the additional convexity assumption on the generators, we have the following sufficient and necessary condition for the optimal randomized test.

Theorem 4.3 *Under Assumptions (H2), (H3)', (H4) and (H5), a randomized test ξ^* is optimal if and only if there exist a constant $v > 0$ and a random variable $b \in [0,1]$ a.s. such that*

$$\xi^* = 1_{\{vm(T)<n(T)\}} + b1_{\{vm(T)=n(T)\}} \tag{4.4}$$

and

$$\mathcal{E}_{g^0}[\xi^*] = \alpha$$

where $m(\cdot)$ and $n(\cdot)$ are the solutions of the adjoint equations (3.4).

Proof. This is a direct deduction of Theorem 3.3. ■

In addition to the characterization of an optimal test, a question of interest is whether we know a priori that such an optimal test exists. Such an existence problem is addressed in the following theorem.

Theorem 4.4 *Under Assumptions (H2), (H3)', (H4) and (H5), there must exist a unique optimal randomized test to problem (2.3).*

Proof. The following functional

$$J(\xi) = \mathcal{E}_{g^1}[1-\xi], \ \ \xi \in \bar{\mathcal{L}}$$

is convex. However, $\bar{\mathcal{L}}$ is bounded, closed and convex; hence it is weakly compact. Since $\xi \mapsto J(\xi)$ is strongly continuous (see [13] and [14]) and convex, by classical results in convex analysis (see [3]), J is weakly lower-semicontinuous. Thus the minimum of the problem (2.3) is attained (see, e.g., [3, Corollary 3.20]). The proof is complete. ■

The next question is what a general procedure is to compute the optimal randomized test ξ^*. Indeed, we can obtain ξ^* by solving the following forward–backward SDE (FBSDE) system

$$\begin{cases} dm(t) = g_y^0(y_0(t), z_0(t), t)m(t)dt + g_z^0(y_0(t), z_0(t), t)m(t)dW(t), \\ -dy_0(t) = g^0(y_0(t), z_0(t), t)dt - z_0(t)dW(t), \\ m(0) = 1, \qquad y_0(T) = \xi^*, \\ dn(t) = g_y^1(y_1(t), z_1(t), t)n(t)dt + g_z^1(y_1(t), z_1(t), t)n(t)dW(t), \\ -dy_1(t) = g^1(y_1(t), z_1(t), t)dt - z_1(t)dW(t), \\ n(0) = 1, \qquad y_1(T) = 1 - \xi^*, \end{cases} \tag{4.5}$$

together with the constraints $\xi^* = 1_{\{vm(T)<n(T)\}} + b1_{\{vm(T)=n(T)\}}$ and $\mathcal{E}_{g^0}[\xi^*] = \alpha$. Note that there has been now a rich theory on solving a FBSDE system (where the unknowns are the 6-turple $(m, n, y_0, z_0, y_1, z_1)$), both analytically and numerically, like (4.5); see, e.g., [11], [30], [31], [38] and [39].

In [23] an important representation theorem is derived which states that a test for capacities has a "representation" in terms of a test for usual probabilities in the sense that the two tests have the same likelihood ratio. By virtue of the explicit form in Theorem 4.3, $n(T)/m(T)$ can be considered as a (generalized) likelihood ratio for our test (see also Section 5.1 for elaboration on this point), we have the similar representation theorem for g-capacities.

Theorem 4.5 *There exist two probability measures Q_0 and Q_1 which are absolutely continuous with respect to the probability measure P such that the randomized test between P_{g^0} and P_{g^1} has the same likelihood ratio as a test between Q_0 and Q_1.*

Proof. In view of Theorem 4.3 the likelihood ratio of a randomized test between P_{g^0} and P_{g^1} is $\frac{n(T)}{m(T)}$. Thus we can define two probability measures Q_0 and Q_1 such that the Radon-Nikodym derivative dQ_0/dP (resp. dQ_1/dP) is equal to $m(T)/E_P[m(T)]$ (resp. $n(T)/E_P[n(T)]$). ■

Let us conclude this section by remarking that in this paper we assume that g is a deterministic function for simplicity. The g-expectation and probability can be defined for a more general random generator g (see [33]). All the results in this paper still hold when the generators involved are proper stochastic processes.

5 Applications

In this section, five examples are given to illustrate applications of the general results obtained.

5.1 Neyman–Pearson lemma for probabilities

We start with specializing our results to the classical Neyman-Pearson lemma (see [9] and [28]) for usual probabilities. Suppose there is a nominal probability measure μ and a standard d-dimensional Brownian Motion, $W^\mu(\cdot)$, defined on the complete probability space $(\Omega, \mathcal{F}, \mu)$. For two real vectors θ and ϕ, we define

$$m(T) = \exp\{\theta' W^\mu(T) - \frac{1}{2}T \parallel \theta \parallel^2\}, \ \ n(T) = \exp\{\phi' W^\mu(T) - \frac{1}{2}T \parallel \phi \parallel^2\}.$$

Suppose that a probability measure Q (resp. P) is absolutely continuous with respect to μ on $\mathcal{F}$ which admits the Radon-Nikodym derivative $m(T)$ (resp. $n(T)$). Our purpose is to test the 'hypothesis' Q against an 'alternative' P. In other words, we try to find a randomized test which solves the following problem:

$$\inf_{\xi \in \bar{\mathcal{L}}} E_P[1-\xi],$$
$$\text{subject to } E_Q[\xi] \le \alpha.$$

By Girsanov's theorem, we have that $E_Q[\xi] = E_\mu[m(T)\xi]$, $E_P[\xi] = E_\mu[n(T)\xi]$. On the other hand, E_Q and E_P are (very special) g-expectations: $E_Q[\xi] = \mathcal{E}_{g^0}[\xi]$, $E_P[\xi] = \mathcal{E}_{g^1}[\xi]$, with the following generators

$$g^0(y,z,t) = z\theta, \ \ g^1(y,z,t) = z\phi.$$

Hence, by Theorem 4.3, the optimal randomized test has the form

$$\xi^* = 1_{\{vm(T) < n(T)\}} + b 1_{\{vm(T) = n(T)\}}$$

where $v > 0$ is a constant and $b \in [0,1]$ is a random variable. However, θ and ϕ are deterministic vectors implying that $\mu(vm(T) = n(T)) = 0$. Thus, the optimal $\xi^* = 1_{\{vm(T) < n(T)\}}$, recovering the classical Neyman–Pearson lemma.

Since $n(T)/m(T)$ is known to be the likelihood ratio in this particular example, we have an interpretation of the adjoint processes governed by (3.4) for the general g-probability case: $n(T)/m(T)$ can be seen as a generalization of the "likelihood ratio". In Huber and Strassen [23], they derived a generalized Radon-Nikodym derivative of a capacity with respect to another capacity. Our results show that such a generalized Radon-Nikodym derivative in our context is nothing else than the ratio of the adjoint processes.

5.2 Learning a financial market with ambiguity

In this example, we consider a financial market model where the risk premium process is ambiguous to the investor, and the investor hopes to "learn" the market by choosing an appropriate portfolio and observing the resulting final wealth position.

Assume there are one bank account (risk free instrument) and d stocks (risky instruments) in a financial market. The respective prices $S_0(\cdot)$ and $S_1(\cdot), \cdots, S_d(\cdot)$ are governed by the equations

$$\begin{aligned} dS_0(t) &= rS_0(t)dt, & S_0(0) &= s_0, \\ dS_i(t) &= S_i(t)[b_i(t)dt + \sum_{j=1}^{d} \sigma_{ij} dW^\mu(t)], & S_i(0) &= s_i > 0; i = 1, \dots, d \end{aligned}$$

where $W^\mu(\cdot)$ is a standard d-dimensional Brownian Motion defined on a complete probability space $(\Omega, \mathcal{F}, \mu)$. We adopt the usual assumptions: the interest rate $r > 0$, and the stock-volatility matrix $\sigma = \{\sigma_{ij}\}_{1 \le i,j \le d}$ is assumed to be invertible. Set $B := (b_1 - r, \dots, b_m - r)'$. The risk premium η is defined by $\eta \equiv (\eta_1, \dots, \eta_m)' := \sigma^{-1}B$. However, η is ambiguous to the investor, which is usually caused by the so-called mean–blue problem, namely, the difficulty in estimating the appreciation rates $b_1, \dots, b_d$. For simplicity, we suppose that there are only two possibilities: $\eta(\cdot) = \theta$ (hypothesis) versus $\eta(\cdot) = \phi$ (alternative). Our question is: can the investor design a trading strategy so that s/he can learn the real risk premium by simply observing the corresponding final wealth position? In other words, if s/he wants to use the wealth $X(T)$ as a sample to test the above simple hypothesis, what should be the best form of $X(T)$? After obtaining the optimal $X^*(T)$, one can get the corresponding portfolio $\pi^*(\cdot)$ by replicating $X^*(T)$, since the market is complete in our setup.

We recall that the wealth process $X(\cdot)$ of the investor satisfies the following equation

$$\begin{cases} dX(t) = [rX(t) + \pi(t)'\sigma\eta]dt + \pi(t)'\sigma dW^\mu(t), \\ X(0) = x, \end{cases} \tag{5.1}$$

where x is the initial wealth and $\pi(\cdot) \equiv (\pi_1(\cdot), \dots, \pi_m(\cdot))'$ is a trading strategy (or portfolio). Here $\pi_i(t)$ is the amount of money held in the ith stock at time t.

Set $y(t) = e^{-rt}X(t)$ and $z(t) = e^{-rt}\pi(t)'\sigma$, then the dynamics of (5.1) can be rewritten as

$$-dy(t) = -z(t)\eta(t)dt - z(t)dW^\mu(t).$$

Denote by Q (resp. P) the probability measure which admits the Radon-Nikodym derivative $m(T)$ (resp. $n(T)$) with respect to μ on $\mathcal{F}$ where

$$m(T) = \exp\{\theta' W^\mu(T) - \frac{1}{2}T \parallel \theta \parallel^2\}, \;\; n(T) = \exp\{\phi' W^\mu(T) - \frac{1}{2}T \parallel \phi \parallel^2\}.$$

We are now in the framework of Section 5.1. Hence the optimal randomized test has the form $e^{-rT}X^*(T) = 1_{\{vm(T) < n(T)\}}$ or

$$X^*(T) = e^{rT} 1_{\{vm(T) < n(T)\}}.$$

This is the payoff of a digital (or binary) option, and one can use a standard Black–Scholes approach to find the corresponding replicating strategy. The details are left to the interested reader.

5.3 Super-pricing a contingent claim

In this subsection we consider the super-pricing

$$\sup_{Q\in\mathcal{P}} E_Q[\xi]$$

of a contigent claim ξ where

$$\mathcal{P}=\left\{Q \mid \frac{dQ}{dP}=\exp\{-\frac{1}{2}\int_0^T \mid v_t\mid^2 dt+\int_0^T v_t dW_t\}, \quad \sup_{t\in[0,T]} \mid v_t \mid\leq k\right\}.$$

The problem is addressed in [6], [7] and [13] via g-expectation. By Lemma 2 in [7], the above pricing problem can be described by a g-expectation with the generator $\bar{g}(y,z,t)=k\mid z\mid$ where k is a positive constant. Our simple hypothesis testing problem is to test

$$H_0:\bar{g}=\bar{g}^0 \text{ versus a simple alternative hypothesis } H_1:\bar{g}=\bar{g}^1$$

where

$$\bar{g}^0(y,z,t)=\theta\mid z\mid,\ \ \bar{g}^1(y,z,t)=\phi\mid z\mid.$$

Specifically, for a given $\alpha\in(0,1)$ we need to solve the following optimization problem

$$\begin{gathered}\inf_{\xi\in\bar{\mathcal{L}}}\mathcal{E}_{\bar{g}^1}[1-\xi],\\ \text{subject to } \mathcal{E}_{\bar{g}^0}[\xi]\leq\alpha.\end{gathered}\tag{5.2}$$

Note that in this case the generator is convex, yet inherently non-differentiable at $z=0$. As remarked earlier we have only a sufficient condition for the optimal randomized test. More precisely, if there exist a constant $v>0$ and a random variable $b\in[0,1]$ such that a randomized test $\xi=1_{\{vm(T)<n(T)\}}+b1_{\{vm(T)=n(T)\}}$ satisfies $\mathcal{E}_{\bar{g}^0}[\xi]=\alpha$, then ξ is an optimal test. Here $m(\cdot)$ and $n(\cdot)$ are the solutions of the following adjoint equations:

$$\begin{cases} dm(t)=m(t)\bar{g}^0_z(t)dW(t), & m(0)=1,\\ dn(t)=n(t)\bar{g}^1_z(t)dW(t), & n(0)=1,\end{cases}$$

with $\bar{g}^i_z(t)$, $i=0,1$, denoting an adapted process belonging $dP\otimes dt$ almost surely to $\partial g^i(z_i^*(t))$, where $\partial\bar{g}^i$ is the subdifferential of $\bar{g}^i$ with respect to z.

5.4 A k-ignorance model and sub-pricing a contingent claim

Chen and Epstein [6] formulates a so-called k-ignorance model to incorporate the notion of ambiguity aversion. It leads to a g-expectation whose generator is $\underline{g}(y,z,t)=-k\mid z\mid$, where k is a positive constant describing the degree of ambiguity aversion. On the other hand, the same generator describes the sub-pricing counterpart of the problem in Subsection 5.3:

$$\inf_{Q\in\mathcal{P}} E_Q[\xi].$$

Similarly, we consider the situation where k is unknown and we only know that it is either θ or ϕ. Then our simple hypothesis testing problem is to test

$$H_0:g=\underline{g}^0 \text{ versus } H_1:g=\underline{g}^1$$

where

$$\underline{g}^0(y,z,t) = -\theta \mid z \mid, \;\; \underline{g}^1(y,z,t) = -\phi \mid z \mid .$$

However, in this case, it is difficult to solve the optimization problem

$$\begin{gathered} \inf_{\xi \in \bar{\mathcal{L}}} \mathcal{E}_{\underline{g}^1}[1-\xi], \\ \text{subject to } \mathcal{E}_{\underline{g}^0}[\xi] \le \alpha, \end{gathered} \tag{5.3}$$

directly, since now $\underline{g}^0$ and $\underline{g}^1$ are *concave* in z, and hence the general theory developed in the previous sections (especially the sufficient condition) does not apply. To get around, we consider instead of (5.3) the following *dual problem*

$$\begin{gathered} \sup_{\xi \in \bar{\mathcal{L}}} \mathcal{E}_{\underline{g}^1}[\xi], \\ \text{subject to } \mathcal{E}_{\underline{g}^0}[1-\xi] \ge \beta \end{gathered} \tag{5.4}$$

where $\beta \in (0,1)$ can be interpreted as the power of the randomized test ξ.

Now we transform problem (5.4) into a problem like (5.2). For any constant c and $\xi \in \bar{\mathcal{L}}$, it is easy to check that

$$\begin{gathered} -\mathcal{E}_{\underline{g}}[-\xi] = \mathcal{E}_{\bar{g}}[\xi], \quad -\mathcal{E}_{\bar{g}}[-\xi] = \mathcal{E}_{\underline{g}}[\xi]; \\ \mathcal{E}_{\underline{g}}[c+\xi] = c + \mathcal{E}_{\underline{g}}[\xi], \quad \mathcal{E}_{\bar{g}}[c+\xi] = c + \mathcal{E}_{\bar{g}}[\xi]. \end{gathered}$$

Thus, problem (5.4) becomes

$$\begin{gathered} \inf_{\xi \in \bar{\mathcal{L}}} (\mathcal{E}_{\bar{g}^1}[1-\xi] - 1), \\ \text{subject to } \mathcal{E}_{\bar{g}^0}[\xi] \le 1-\beta \end{gathered} \tag{5.5}$$

which is problem (5.2) if $\alpha = 1-\beta$.

5.5 Minimizing shortfall risk

In a complete financial market, if the seller of a contingent claim does not have enough initial wealth, he may fail to hedge the contingent claim perfectly. Thus, he must face some shortfall risk. Föllmer and Leukert [17] use the expectation of the shortfall weighted by the loss function as a shortfall risk measure. In this subsection, we use a general measure, the convex risk measure, to evaluate the shortfall and consequently minimize such a shortfall risk.

We suppose that the complete financial market model is the same as in Section 5.2, except now that there is no ambiguity in the risk premium process. For simplicity, we assume that the interest rate $r \equiv 0$. Let H be a given contingent claim which is a nonnegative random variable in $L^2(\Omega, \mathcal{F}, P)$. It is well known that there is a unique equivalent martingale measure $P^* \approx P$ such that the price of H at time 0 is $H_0 = E_{P^*}[H]$. If the seller's initial wealth $\tilde{X}_0$ is smaller than H_0, he cannot perfectly hedge H and the shortfall is $-(H - X_T)^+$.

Now we need to measure appropriately the shortfall $-(H - X_T)^+$. Recall that a convex risk measure is a map from $\mathcal{X}$, a proper set of financial positions, to R, satisfying the monotonicity, translation invariance, and convexity; see Föllmer and Schied [19] for details. As in [22], we take $\mathcal{X} = L^2(\Omega, \mathcal{F}, P)$ in our case and introduce a risk measure $\rho_g : L^2(\Omega, \mathcal{F}, P) \to R$ by

$$\rho_g(X) = \mathcal{E}_g[-X], \quad \forall X \in L^2(\Omega, \mathcal{F}, P)$$

where the generator g is convex and satisfies (H2)-(H4). It follows from Proposition 11 in [22] that ρ_g is a convex risk measure.

Consequently, the seller of the contingent claim will solve the following optimization problem:

$$\begin{gathered} \inf_{\pi(\cdot)} \rho_g(-(H - X_T)^+), \\ \text{subject to } X_0 \leq \tilde{X}_0 \end{gathered}$$

where X_0 is the initial wealth used to hedge the claim. Since $X_0 = E_{P^*}[X_T]$, the above problem is equivalent to

$$\begin{gathered} \inf_{X_T} \mathcal{E}_g((H - X_T)^+), \\ \text{subject to } E_{P^*}[X_T] \leq \tilde{X}_0. \end{gathered} \tag{5.6}$$

Now we show that the optimal X_T^* must satisfy $0 \leq X_T^* \leq H$. In fact, if $P(X_T^* > H) > 0$, we can construct a feasible terminal wealth $\tilde{X}_T$ such that $0 \leq \tilde{X}_T \leq H$ and $(H - \tilde{X}_T)^+ < (H - X_T^*)^+$. Thus, $\mathcal{E}_g((H - \tilde{X}_T)^+) < \mathcal{E}_g((H - X_T^*)^+)$ by the strict comparison theorem of BSDEs. This leads to a contradiction.

Thus, without loss of generality we assume that $0 \leq X_T \leq H$ and (5.6) becomes

$$\begin{gathered} \inf_{0 \leq X_T \leq H} \mathcal{E}_g(H - X_T), \\ \text{subject to } E_{P^*}[X_T] \leq \tilde{X}_0. \end{gathered} \tag{5.7}$$

Note that E_{P^*} is a trivial g-expectation with a linear generator. Thus, in view of the remark immediately following Theorem 3.1, (5.7) is essentially a special case of problem (3.1) where $h(X_T) = H - X_T$ – except that now X_T is constrained in a random interval: $0 \leq X_T \leq H$ and h is a random function.

Following the proof in section 3.1, we have that there exist a positive number v and a random variable b $(0 \leq b \leq H)$ such that the optimal terminal wealth X_T^* satisfies

$$X_T^* = 1_{\{v < n(T)\}} + b 1_{\{v = n(T)\}}$$

and

$$E_{P^*}[X_T^*] = \tilde{X}_0$$

where $n(T)$ is the solutions of the following adjoint equation at time T.

$$\begin{cases} dn(t) = g_y(t) n(t) dt + g_z(t)' n(t) dW(t), \\ n(0) = 1. \end{cases}$$

Similar to Theorem 4.5, we can obtain the optimal terminal wealth by solving the following forward backward system:

$$\begin{cases} -dy(t) = g(y(t), z(t), t) dt - z(t) dW(t), \\ y(T) = H - X_T^*, \\ dn(t) = g_y(t) n(t) dt + g_z(t)' n(t) dW(t), \\ n(0) = 1, \\ E_{P^*}[X_T^*] = \tilde{X}_0, \\ X_T^* = 1_{\{v < n(T)\}} + b 1_{\{v = n(T)\}}. \end{cases}$$

After obtaining the optimal terminal wealth X_T^*, we can get the optimal strategy $\pi(\cdot)$ by replicating X_T^*.

Remark: Even when the wealth equation in question is nonlinear (e.g. when the risk-free borrowing rate is different from the lending rate), our method still works. This is because the approach developed in section 3 is not only suitable for g-expectations but also for BSDEs satisfying the usual assumptions.

A

This appendix is devoted to a proof of Theorem 3.1. First of all, for each $0 \le \rho \le 1$ and $\xi \in U$, define $\xi^\rho = \xi^* + \rho(\xi - \xi^*)$. Let $(y_0^\rho(\cdot), z_0^\rho(\cdot), y_1^\rho(\cdot), z_1^\rho(\cdot))$ be the solution to (3.2) corresponding to $\xi = \xi^\rho$, and $(\hat{y}_0(\cdot), \hat{z}_0(\cdot), \hat{y}_1(\cdot), \hat{z}_1(\cdot))$ be the solution of the following variational equations:

$$\begin{cases} -d\hat{y}_0(t) = (g_y^0(t)\hat{y}_0(t) + g_z^0(t)\hat{z}_0(t))dt - \hat{z}_0(t)dW(t), \\ \quad \hat{y}_0(T) = \xi - \xi^*, \\ -d\hat{y}_1(t) = (g_y^1(t)\hat{y}_1(t) + g_z^1(t)\hat{z}_1(t))dt - \hat{z}_1(t)dW(t), \\ \quad \hat{y}_1(T) = h_y(\xi^*)(\xi - \xi^*). \end{cases} \tag{A0}$$

For $i = 0, 1$, set

$$\begin{array}{l} \tilde{y}_i^\rho(t) = \rho^{-1}(y_i^\rho(t) - y_i^*(t)) - \hat{y}_i(t) \\ \tilde{z}_i^\rho(t) = \rho^{-1}(z_i^\rho(t) - z_i^*(t)) - \hat{z}_i(t) \end{array}$$

Lemma A1 Assume (H2)-(H4). Then

$$\begin{array}{l} \lim\limits_{\rho\to 0} E(\sup\limits_{0\le t\le T}[\tilde{y}_i^\rho(t)^2]) = 0, \\ \lim\limits_{\rho\to 0}(E[\int_0^T \mid \tilde{z}_i^\rho(t) \mid^2 dt]) = 0, \quad i = 0, 1. \end{array}$$

Proof. We only prove for $i = 0$. From (3.2) and (A0) it follows

$$\begin{cases} -d\tilde{y}_0^\rho(t) & = \rho^{-1}[g^0(y_0^\rho(t), z_0^\rho(t), t) - g^0(y_0^*(t), z_0^*(t), t) - \rho g_y^0(y_0^*(t), z_0^*(t), t)\hat{y}_0(t) \\ & \quad -\rho g_z^i(y_0^*(t), z_0^*(t), t)\hat{z}_0(t)]dt - \tilde{z}_0^\rho(t)'dW(t), \\ \tilde{y}_0^\rho(T) & = 0. \end{cases}$$

Let

$$\begin{aligned} A^\rho(t) &= \int_0^1 g_y^0(y_0^*(t) + \lambda\rho(\hat{y}_0(t) + \tilde{y}_0^\rho(t)), z_0^*(t) + \lambda\rho(\hat{z}_0(t) + \tilde{z}_0^\rho(t)), t)d\lambda, \\ B^\rho(t) &= \int_0^1 g_z^0(y_0^*(t) + \lambda\rho(\hat{y}_0(t) + \tilde{y}_0^\rho(t)), z_0^*(t) + \lambda\rho(\hat{z}_0(t) + \tilde{z}_0^\rho(t)), t)d\lambda, \\ C^\rho(t) &= [A^\rho(t) - g_y^0(y_0^*(t), z_0^*(t), t)]\hat{y}_0(t) + [B^\rho(t) - g_z^0(y_0^*(t), z_0^*(t), t)]\hat{z}_0(t). \end{aligned}$$

Then

$$\begin{cases} -d\tilde{y}_0^\rho(t) = (A^\rho(t)\cdot\tilde{y}_0^\rho(t) + B^\rho(t)\cdot\tilde{z}_0^\rho(t) + C^\rho(t))dt - \tilde{z}_0^\rho(t)'dW(t), \\ \tilde{y}_0^\rho(T) = 0. \end{cases}$$

A standard estimation on the above BSDE (see, e.g., [38, p.349]) yields

$$\begin{aligned} & E(\sup_{0\le t\le T}[\tilde{y}_i^\rho(t)^2]) + E\int_t^T \mid \tilde{z}_0^\rho(s) \mid^2 ds \\ & \le KE\int_t^T \mid \tilde{y}_0^\rho(s) \mid^2 ds + KE\int_t^T \mid C^\rho(s) \mid^2 ds \end{aligned}$$

where $K > 0$ is a constant. However, the Lebesgue dominated convergence theorem implies

$$\lim_{\rho\to 0} E\int_0^T \mid C^\rho(t) \mid^2 dt = 0.$$

The desired result follows by applying Grownwall's inequality. ■

Let $d(\cdot,\cdot)$ be the metric in U naturally induced by its norm, and introduce a mapping $F_\varepsilon(\cdot): U \to R$ by

$$F_\varepsilon(\xi) = \Big((\mathcal{E}_{g^0}[\xi] - \alpha)^2 + (\max(0, \mathcal{E}_{g^1}[h(\xi)] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon))^2 \Big)^{\frac{1}{2}}$$

where ε is a given positive constant. It is easy to verify that $F_\varepsilon(\cdot)$ is a continuous functional on U.

Remark: In the following proof, we only prove the case where the optimal ξ^* satisfies $\mathcal{E}_{g^0}[\xi^*] = \alpha$ (i.e. the constraint is binding). If $\mathcal{E}_{g^0}[\xi^*] < \alpha$, then we need only to redefine $F_\varepsilon(\xi)$ as

$$F_\varepsilon(\xi) = \Big((\max(0, \mathcal{E}_{g^1}[h(\xi)] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon))^2 \Big)^{\frac{1}{2}}$$

and the proof is similar.

Lemma A2 Under the assumptions of Theorem 3.1, there must exist $h_1 \in R^1$, $h_0 \in R^1$ with $h_0 \geq 0$ and $|h_0| + |h_1| = 1$ such that the following variational inequality holds

$$h_1 \hat{y}_0(0) + h_0 \hat{y}_1(0) \geq 0 \tag{A1}$$

where $\hat{y}_i(0)$, $i = 0, 1$, are the solutions of (A0) valued at time 0.

Proof. It is easy to check that

$$F_\varepsilon(\xi^*) = \varepsilon; \quad F_\varepsilon(\xi) > 0 \ \ \forall \xi \in U,$$

which leads to $F_\varepsilon(\xi^*) \leq \inf\limits_{\xi \in U} F_\varepsilon(\xi) + \varepsilon$. Thus by Ekeland's variational principle ([12, Theorem 1.1]), there exists $\xi^\varepsilon \in U$ such that

$$\begin{array}{ll} \text{(i)} & F_\varepsilon(\xi^\varepsilon) \leq F_\varepsilon(\xi^*), \\ \text{(ii)} & d(\xi^\varepsilon, \xi^*) \leq \sqrt{\varepsilon}, \\ \text{(iii)} & F_\varepsilon(\xi) + \sqrt{\varepsilon} d(\xi, \xi^\varepsilon) \geq F_\varepsilon(\xi^\varepsilon) \quad \forall \xi \in U. \end{array}$$

For any $\xi \in L^2$ and $0 \leq \rho \leq 1$ introduce the following notation

$$\hat{\xi} = \xi - \xi^*, \qquad \hat{\xi}^\varepsilon = \xi - \xi^\varepsilon, \qquad \xi^\varepsilon_\rho = \xi^\varepsilon + \rho \hat{\xi}^\varepsilon.$$

Then we have

$$F_\varepsilon(\xi^\varepsilon_\rho) + \sqrt{\varepsilon} d(\xi^\varepsilon_\rho, \xi^\varepsilon) - F_\varepsilon(\xi^\varepsilon) \geq 0 \tag{A2}$$

where

$$d(\xi^\varepsilon_\rho, \xi^\varepsilon) = (E \mid \rho \hat{\xi}^\varepsilon \mid^2)^{\frac{1}{2}} = \rho (E \mid \hat{\xi}^\varepsilon \mid^2)^{\frac{1}{2}}.$$

Consider the following variational equation

$$\begin{cases} -d\hat{y}^\varepsilon_0(t) = [g^0_y(y^\varepsilon_0(t), z^\varepsilon_0(t), t)\hat{y}^\varepsilon_0(t) + g^0_z(y^\varepsilon_0(t), z^\varepsilon_0(t), t)\hat{z}^\varepsilon_0(t)]dt - \hat{z}^\varepsilon_0(t) dW(t), \\ \quad \hat{y}^\varepsilon_0(T) = \hat{\xi}^\varepsilon, \\ -d\hat{y}^\varepsilon_1(t) = [g^1_y(y^\varepsilon_1(t), z^\varepsilon_1(t), t)\hat{y}^\varepsilon_1(t) + g^1_z(y^\varepsilon_1(t), z^\varepsilon_1(t), t)\hat{z}^\varepsilon_1(t)]dt - \hat{z}^\varepsilon_1(t) dW(t), \\ \quad \hat{y}^\varepsilon_1(T) = h_y(\xi^\varepsilon)\hat{\xi}^\varepsilon \end{cases} \tag{A3}$$

where $(y^\varepsilon_0(\cdot), z^\varepsilon_0(\cdot), y^\varepsilon_1(\cdot), z^\varepsilon_1(\cdot))$ is the solution to (3.2) corresponding to $\xi = \xi^\varepsilon$. A similar result to Lemma A1 has

$$\lim_{\rho \to 0} \mid \rho^{-1}(\mathcal{E}_{g^i}[\xi^\varepsilon_\rho] - \mathcal{E}_{g^i}[\xi^\varepsilon]) - \hat{y}^\varepsilon_i(0) \mid = 0, \qquad i = 0, 1,$$

namely,

$$\mathcal{E}_{g^i}[\xi^\varepsilon_\rho] - \mathcal{E}_{g^i}[\xi^\varepsilon] = \rho \hat{y}^\varepsilon_i(0) + o(\rho).$$

This leads to the following expansions:

$$\begin{aligned} &\mid \mathcal{E}_{g^0}[\xi^\varepsilon_\rho] - \alpha \mid^2 - \mid \mathcal{E}_{g^0}[\xi^\varepsilon] - \alpha \mid^2 = 2\rho(\mathcal{E}_{g^0}[\xi^\varepsilon] - \alpha)\hat{y}^\varepsilon_0(0) + o(\rho); \\ &\mid \mathcal{E}_{g^1}[\xi^\varepsilon_\rho] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon \mid^2 - \mid \mathcal{E}_{g^1}[\xi^\varepsilon] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon \mid^2 \\ &= 2\rho \hat{y}^\varepsilon_1(0)(\mathcal{E}_{g^1}[\xi^\varepsilon] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon) + o(\rho). \end{aligned}$$

Now consider two cases:

Case 1: For sufficiently small $\rho > 0$, $\mathcal{E}_{g^1}[h(\xi^\varepsilon_\rho)] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon > 0$. In this case,

$$\begin{aligned}
&\lim_{\rho\to 0} \frac{F_\varepsilon(\xi^\varepsilon_\rho) - F_\varepsilon(\xi^\varepsilon)}{\rho} \\
&= \lim_{\rho\to 0} \frac{1}{F_\varepsilon(\xi^\varepsilon_\rho) + F_\varepsilon(\xi^\varepsilon)} \frac{[F_\varepsilon(\xi^\varepsilon_\rho)]^2 - [F_\varepsilon(\xi^\varepsilon)]^2}{\rho} \\
&= \frac{1}{F_\varepsilon(\xi^\varepsilon)} \left\{ (\mathcal{E}_{g^0}[\xi^\varepsilon] - \alpha)\hat{y}^\varepsilon_0(0) + (\mathcal{E}_{g^1}[\xi^\varepsilon] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon)\hat{y}^\varepsilon_1(0) \right\}.
\end{aligned}$$

Set

$$h^\varepsilon_1 = \frac{\mathcal{E}_{g^0}[\xi^\varepsilon] - \alpha}{F_\varepsilon(\xi^\varepsilon)},\ h^\varepsilon_0 = \frac{1}{F_\varepsilon(\xi^\varepsilon)} [\mathcal{E}_{g^1}[\xi^\varepsilon] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon] \geq 0.$$

It then follows from (A2) that

$$h^\varepsilon_1 \hat{y}^\varepsilon_0(0) + h^\varepsilon_0 \hat{y}^\varepsilon_2(0) \geq -\sqrt{\varepsilon}(E \mid \hat{\xi}^\varepsilon \mid^2)^{\frac{1}{2}}. \tag{A4}$$

Case 2: There exists $\{\rho_n\} \downarrow 0$ such that $\mathcal{E}_{g^1}[h(\xi^\varepsilon_{\rho_n})] - \mathcal{E}_{g^1}[h(\xi^*)] + \varepsilon \leq 0$. In this case, $F_\varepsilon(\xi^\varepsilon_{\rho_n}) = \mid \mathcal{E}_{g^0}[\xi^\varepsilon_{\rho_n}] - \alpha \mid$. Hence

$$\lim_{\rho_n\to 0} \frac{F_\varepsilon(\xi^\varepsilon_{\rho_n}) - F_\varepsilon(\xi^\varepsilon)}{\rho_n} = \frac{1}{F_\varepsilon(\xi^\varepsilon)} (\mathcal{E}_{g^0}[\xi^\varepsilon] - \alpha)\hat{y}^\varepsilon_0(0).$$

Setting

$$h^\varepsilon_1 = \frac{\mathcal{E}_{g^0}[\xi^\varepsilon] - \alpha}{F_\varepsilon(\xi^\varepsilon)},\ \ h^\varepsilon_0 = 0,$$

then clearly (A4) holds as well for this case.

In summary, for a given ε, we have (i) A4) holds, (ii) $h^\varepsilon_0 \geq 0$, and (iii) $\mid h^\varepsilon_0 \mid^2 + \mid h^\varepsilon_1 \mid^2 = 1$ for both cases. Consequently, there exists a converging subsequence of $(h^\varepsilon_0, h^\varepsilon_1)$ (still denoted by $(h^\varepsilon_0, h^\varepsilon_1)$) with the limit (h_0, h_1). Since $h^\varepsilon_0 \geq 0$, we have $h_0 \geq 0$. On the other hand, it is easy to check that $\hat{y}^\varepsilon_1(0) \to \hat{y}_0(0)$, $\hat{y}^\varepsilon_1(0) \to \hat{y}_1(0)$ as $\varepsilon \to 0$. This proves (A1). ■

Proof of Theorem 3.1. Applying Itô's lemma to $h_0\hat{y}_1(t)n(t) + h_1\hat{y}_0(t)m(t)$ and using (A0) and (3.4), we have

$$E[(h_0\hat{y}_1(T)n(T) + h_1\hat{y}_0(T)m(T)) - (h_0\hat{y}_1(0)n(0) + h_1\hat{y}_0(0)m(0))] = 0.$$

However,

$$E[h_0\hat{y}_1(T)n(T) + h_1\hat{y}_0(T)m(T)] = E\langle h_0 h_y(\xi^*)n(T) + h_1 m(T), \xi - \xi^*\rangle.$$

It then follows from (A1) that

$$E\langle h_0 h_y(\xi^*)n(T) + h_1 m(T), \xi - \xi^*\rangle \geq 0.$$

Since the above is true for any $\xi \in U$, we have

$$\langle h_1 m(T) + h_0 h_y(\xi^*)n(T), u - \xi^*\rangle \geq 0 \quad \text{a.s.}, \ \ \forall u \in K.$$

Set

$$\bar{\Omega} \triangleq \{\omega \in \Omega \mid \xi^*(\omega) \in \partial K\}.$$

Thus, for each $u \in K$,

$$\begin{aligned}
&\langle h_1 m(T) + h_0 h_y(\xi^*)n(T), u - \xi^*\rangle \geq 0 \ \ a.s. \quad on \quad \bar{\Omega}, \\
&h_1 m(T) + h_0 h_y(\xi^*)n(T) = 0 \ \ a.s. \quad \text{on } \bar{\Omega}^c.
\end{aligned}$$

This leads to

$$\begin{aligned}
h_0 n(T) &= h_1 m(T) \quad a.s. \quad \text{on } \{k_1 < \xi^* < k_2\}; \\
h_0 n(T) &\leq h_1 m(T) \quad a.s. \quad \text{on } \{\xi^* = k_1\}; \\
h_0 n(T) &\geq h_1 m(T) \quad a.s. \quad \text{on } \{\xi^* = k_2\}.
\end{aligned}$$

It is easy to see that (3.3) holds. The proof is complete. ■

B

In this appendix we prove Theorems 3.3.

Proof of Theorem 3.3. (1) To prove the necessary condition. Let ξ^* be optimal and (y_0^*, z_0^*) and (y_1^*, z_1^*) be the corresponding solutions with respect to g^0 and g^1. Take $\xi \in U$. Then, for each $0 \le \rho \le 1$, $\xi^\rho = \xi^* + \rho(\xi - \xi^*) \in U$ since K is a convex set.. Let $(\hat{y}_i, \hat{z}_i)$ be the solution of the variational equation (A0), and (y_i^ρ, z_i^ρ) be the solution to (3.2) associated with generator g^i and terminal condition ξ^ρ, $i = 1, 2$. Finally, set

$$\tilde{y}_i^\rho(t) = \rho^{-1}(y_i^\rho(t) - y_i^*(t)) - \hat{y}_i(t)$$
$$\tilde{z}_i^\rho(t) = \rho^{-1}(z_i^\rho(t) - z_i^*(t)) - \hat{z}_i(t).$$

Similar to Lemma A1, we have the following convergence results:

$$\lim_{\rho \to 0}(\sup_{0 \le t \le T} E[\tilde{y}_i^\rho(t)^2]) = 0, \quad \lim_{\rho \to 0}(E[\int_0^T \mid \tilde{z}_i^\rho(t) \mid^2 dt]) = 0.$$

Since ξ^* is optimal, for each $0 \le \rho \le 1$, we have

$$\mathcal{E}_{g^1}[(\xi^\rho)] + v(\mathcal{E}_{g^0}[\xi^\rho] - \alpha) \ge \mathcal{E}_{g^1}[h(\xi^*)] + v(\mathcal{E}_{g^0}[\xi^*] - \alpha).$$

Dividing the inequality by ρ and letting $\rho \to 0$, we have

$$\hat{y}_1(0) + v\hat{y}_0(0) \ge 0 \tag{B1}$$

where $\hat{y}_i(0)$ denotes the solution of (A0) at time 0. Applying Itô's lemma to $\hat{y}_1(t)n(t) + v\hat{y}_0(t)m(t)$ and noting (3.4) we have s

$$E[(\hat{y}_1(T)n(T) + v\hat{y}_0(T)m(T)) - (\hat{y}_1(0)n(0) + v\hat{y}_1(0)m(0))] = 0.$$

It follows from (B1) that

$$\begin{aligned} &E[\hat{y}_1(T)n(T) + v\hat{y}_0(T)m(T)] \\ &= E[(h_y(\xi^*)n(T) + vm(T))(\xi - \xi^*)] \ge 0. \end{aligned}$$

A standard analysis shows that

$$\langle vm(T) + h_y(\xi^*)n(T), u - \xi^* \rangle \ge 0, \quad a.s., \quad \forall u \in K.$$

Hence, similar to the proof of Theorem 3.1, (3.7) holds.
$\mathcal{E}_{g^0}[\xi^*] = \alpha$ is due to the strict backward comparison theorem.

(2) To prove the sufficient condition. For any $\xi \in U$, we want to prove

$$\mathcal{E}_{g^1}[h(\xi)] + v(\mathcal{E}_{g^0}[\xi] - \alpha) \ge \mathcal{E}_{g^1}[h(\xi^*)] + v(\mathcal{E}_{g^0}[\xi^*] - \alpha)$$

or

$$\mathcal{E}_{g^1}[h(\xi)] - \mathcal{E}_{g^1}[h(\xi^*)] + v(\mathcal{E}_{g^0}[\xi] - \mathcal{E}_{g^0}[\xi^*]) \ge 0.$$

Set

$$f_1(y, z, t) = g^0(y_0^*(t) + y, z_0^*(t) + z, t) - g^0(y_0^*(t), z_0^*(t), t),$$
$$f_2(y, z, t) = g_y^0(y_0^*(t), z_0^*(t), t)y + g_z^1(y_0^*(t), z_0^*(t), t)z.$$

Consider (A0) and the following equation

$$\begin{cases} -d(y_0(t) - y_0^*(t)) = [g^0(y_0(t), z_0(t), t) - g^0(y_0^*(t), z_0^*(t), t)]dt - (z_0(t) - z_0^*(t))dW(t), \\ \qquad\qquad\qquad\quad\; = [f_1(y_0(t) - y_0^*(t), z_0(t) - z_0^*(t), t)dt - (z_0(t) - z_0^*(t))dW(t), \\ \quad y_0(T) - y_0^*(T) = \xi - \xi^*. \end{cases}$$

By assumption (H5),

$$f_1(y,z,t) \geq f_2(y,z,t) \qquad \forall y, z, \quad dP \otimes dt - a.s.$$

Using the comparison theorem, we have

$$y_0(t) - y_0^*(t) \geq \hat{y}_0(t) \quad \forall t, \quad P - a.s.$$

where $\hat{y}_0(t)$ is the solution of (A0). Using the similar analysis, we obtain

$$y_1(t) - y_1^*(t) \geq \hat{y}_1(t).$$

Thus, we have

$$\begin{aligned} &\mathcal{E}_{g^1}[h(\xi)] - \mathcal{E}_{g^1}[h(\xi^*)] + v(\mathcal{E}_{g^0}[\xi] - \mathcal{E}_{g^0}[\xi^*]) \\ &= y_1(0) - y_1^*(0) + v[y_0(0) - y_0^*(0)] \\ &\geq \hat{y}_1(0) + v\hat{y}_0(0) \\ &= E[\hat{y}_1(T)n(T) + v\hat{y}_0(T)m(T)] \\ &= E[(\xi - \xi^*)(vm(T) + h_y(\xi^*)n(T))]. \end{aligned}$$

Since we assume that (3.7) holds, it is easy to prove

$$E[(\xi - \xi^*)(vm(T) + h_y(\xi^*)n(T))] \geq 0.$$

By Lemma 3.2, we obtain the result. ■

Acknowledgements

The authors would like to thank the referee for his/her comments, and S. Peng and Z. Chen for their helpful discussions.

References

1. Artzner, P., Delbaen, F., Eber, J. M. and Heath, D.: Coherent measures of risk. Math. Finance **4**, 203-228 (1999)
2. Boyle, P. and Tian, W.: Portfolio management with constraints. Math. Finance **17**, 319-343 (2007)
3. Brezis, H.: Analyse fonctionnelle. Masson Paris (1983)
4. Buja, A.: Simulataneously least favourable experiments I: Upper standard functional and sufficeincy. Z. Wahrsch. Verw. Gebiete **69**, 367-384 (1984)
5. Chen, Z., Chen, T. and Davison, M.: Choquet expectation and Peng's g-expectation. Ann. of Prob. **33**, 1179-1199 (2005)
6. Chen, Z. and Epstein, L.: Ambiguity, risk and asset returns in continuous time. Econometrica **70**, 1403-1443 (2002)
7. Chen, Z. and Kulperger, R.: Minimax pricing and Choquet pricing. Insurance: Math. and Econom. **38**, 518-528 (2006)
8. Coquet, F., Hu, Y., Memin, J. and Peng, S.: Filtration consistent nonlinear expectations and related g-expectations. Probab. Theory and Rel. Fields **123**, 1-27 (2002)
9. Cvitanic, J. and Karatzas, I.: Generalized Neyman-Pearson lemma via convex duality. Bernoulli **7**, 79-97 (2001)
10. Waegenaere, A. De and Wakker, P.: Nonmonotonic Choquet integrals. J. Math. Econom. **36**, 45-60 (2001)
11. Douglas, Jr. J, Ma, J. and Protter, P.: Numerical methods for forward-backward stochastic differctential equations. Ann. Appl. Prob., **6**, 940–968 (1996)
12. Ekeland, I.: On the variational principle. J. Math. Anal. Appl., **47**, 324-353 (1974)
13. Karoui, N. El , Peng, S. and Quenez, M. C.: Backward stochastic differential equations in finance. Math. Finance **7**, 1-71 (1997)
14. Karoui, N. El , Peng, S. and Quenez, M. C.: A dynamic maximum principle for the optimization of recursive utilities under constraints.Ann. Appl. Prob. **11**, 664-693 (2001)

15. Ferguson, T.: Mathematical statistics: a decision-theoretic approach. Academic Press, New York and London (1967)
16. Föllmer, H. and Leukert, P.: Quantile hedging. Fin. and Stoch. **3**, 251-273 (1999)
17. Föllmer, H. and Leukert, P.: Efficient hedging: cost versus shortfall risk. Fin. and Stoch.**4**, 117-146 (2000)
18. Föllmer, H. and Schied, A.: Convex measures of risk and trading constraints. Fin. and Stoch. **6**, 429-447 (2002)
19. Föllmer, H. and Schied, A.: Robust preferences and convex measures of risk. In: Sandmann, K., Schonbucher, P.J. (Eds.), Advances in Finance and Stochastics. Springer-Verlag, 39-56 (2002)
20. Frittelli, M.: Representing sublinear risk measures and pricing rules. Working paper n. **10**, Universit'a di Milano Bicocca, Italy (2000)
21. Frittelli, M. and Gianin, E.: Putting order in risk measures, J. Banking and Fin. **26(7)**, 1473-1486 (2002)
22. Gianin, E.: Risk measures via g-expectations. Insurance: Math. and Econom. **39**, 19-34 (2006)
23. Huber, P. and Strassen, V.: Minimax tests and the Neyman-Pearson lemma for capacities. Ann. Statist. **1(2)**, 251-263 (1973)
24. Ji, S. and Peng, S.: Terminal perturbation method for the backward approach to continuous-time mean-variance portfolio selection. Stoch. Proc. and Their Appl. **118**, 952-967 (2008)
25. Ji, S. and Zhou, X. Y.: A maximum principle for stochastic optimal control with terminal state constraints, and its applications. Comm. in Inform. and Syst. **6** (A special issue dedicated to Tyrone Duncan on the occasion of his 65th birthday), 321-337 (2006)
26. Ji, S. and Zhou, X. Y.: The Neyman-Pearson lemma under g-probabilities. C. R. Acad. Sci. Paris, Ser. I **346**, 209-212 (2008)
27. Jiang, L.: Convexity, translation invariance and subadditivity for g-expectations and related risk measures. Ann. Appl. Prob. **18**, 245-285 (2008)
28. Lehmann, E.: Testing statistical hypotheses, 2nd edition. Wiley, New York (1986)
29. Luenberger, D.: Optimization by vector space methods. Wiley, New York (1969)
30. Ma, J. and Yong, J.: Forward-backward stochastic differential equations and their applications. Lecture Notes in Math. **1702**, Springer-Verlag, New York (1999)
31. Milstein, G. and Tretyakov, M.: Numerical algorithms for forward-backward stochastic differential equations. SIAM J. Sci. Comput. **28(2)**, 561-582 (2006)
32. Pardoux, E. and Peng, S.: Adapted solution of a backward stochastic differential equation. Syst. and Contr. lett. **14**, 55-61 (1990)
33. Peng, S.: Backward stochastic differential equations and related g-expectation. in: Karoui,N. El. and Mazliak, L. (Eds.), Backward stochastic differential equations, in: Pitman Res. Notes Math. Ser., Vol. **364**, 141-159 (1997)
34. Peng, S.: Nonlinear expectations, nonlinear evaluations and risk measures. in: Frittelli, M. and Runggaldier, W. (Eds.), Stochastic methods in finance, Lecture Notes in Math. **1856**, 143-217 (2004)
35. Shao, J.: Mathematical statistics 2nd ed. Springer-Verlag, New York (2003)
36. Wasserman, L. and Kadane, J.: Bayes's theorem for Choquet capacities. Ann. Statist. **18**, 1328-1339 (1990)
37. Witting, H.: Mathematische statistik I. Stuttgart: B.G. Teubner (1985)
38. Yong, J. and Zhou, X. Y.: Stochastic controls: Hamiltonian systems and HJB equations. Springer, New York (1999)
39. Zhao, W., Chen, L. and Peng, S.: A new kind of accurate numerical method for backward stochastic differential equations. SIAM J. Sci. Comput. **28(4)**, 1563-1581 (2006)