Errata and Comments for *Lectures on Stochastic Programming, Second Edition*

Pages 45, 330 and 464. For a further discussion of the interchangeability principle see [5].

Page 196, formula (5.127) should be

$$N \geq \frac{O(1)\lambda^2 \bar{D}^2_{a,\varepsilon}}{(\varepsilon-\delta)^2}\left[n \ln\left(\frac{O(1)LD^*_a}{\varepsilon-\delta}\right) + \ln\left(\frac{1}{\alpha}\right)\right].$$

Page 262, in eq. (5.410) and (5.411) replace γ_t and γ_1 with γ_{t+1} and γ_2, respectively.

Page 266, eq. (5.437). For more accurate derivation of this bound see [1]. Numerical experiments indicate that bias of this upper bound could be huge especially when the number of stages is large.

Page 337, eq. (6.251). It is assumed in Theorem 5.7 that the set over which the minimization is performed is compact, while in (6.248) it is the real line $\mathbb{R}$. It is possible to deal with this problem in the following way. The set of minimizers of the true problem here is the bounded interval $[t^*, t^{**}]$, where t^* and t^{**} are the respective quantiles. Consider a closed bounded interval I such that $[t^*, t^{**}]$ lies in the interior of I. Then $\theta^* = \inf_{t\in I} f(t)$, and let $\tilde{\theta}_N := \inf_{t\in I} \hat{f}_N(t)$. Since minimizers of $\hat{f}_N(t)$ over $t \in I$ converge w.p.1 to $[t^*, t^{**}]$, it follows that such a minimizer will be w.p.1 in the interior of the interval I for N large enough, and hence by convexity of $\hat{f}_N(t)$ will be also a minimizer of $\hat{f}_N(t)$ over $t \in \mathbb{R}$. It follows that $\tilde{\theta}_N = \hat{\theta}_N$ w.p.1 for N large enough. Thus $N^{1/2}(\tilde{\theta}_N - \hat{\theta}_N) = 0$ w.p.1 for N large enough, and hence $N^{1/2}(\tilde{\theta}_N - \hat{\theta}_N)$ tends to zero w.p.1 and thus tends to zero in probability. That is $\tilde{\theta}_N - \hat{\theta}_N = o_p(N^{-1/2})$. Applying now Theorem 5.7 to the respective minimization over the compact set I concludes the arguments.

Page 356. Dynamic programming equations (6.337) are sufficient for optimality of the respective policy but are not necessary unless risk mappings ρ_t are *strictly* monotone (cf., [5]).

Page 377, below eq. (6.440), change $\psi_t(\cdot, d_t)$ to $\psi_t(x_t, \cdot)$.

Page 459, line 5. Replace $P(A_1 \neq A_2) > 0$ by $P(A_1 \triangle A_2) > 0$, where $A_1 \triangle A_2 = (A_1 \setminus A_2) \cup (A_2 \setminus A_1)$.

Comments

In sections 6.3.3 - 6.3.5 of [4] the concept of measure-preserving transformations was used. For general probability spaces there are delicate measurability issues involved. We describe below an alternative approach which is more direct and intuitive.

For $\mathcal{Z} = L_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$, and nonempty set $\mathfrak{C} \subset \mathcal{Z}^*$ of *density functions* consider functional

$$\rho(Z) := \sup_{\zeta\in\mathfrak{C}} \langle \zeta, Z\rangle, \tag{1}$$

where $\langle \zeta, Z\rangle = \int \zeta Z dP$. Recall that (1) can be viewed as dual representation of coherent risk measure $\rho : \mathcal{Z} \to \mathbb{R}\cup\{+\infty\}$. If the set $\mathfrak{C}$ is convex and weakly* closed, then $\mathfrak{C}$ is referred to as the *dual set* of the corresponding coherent risk measure, and is denoted $\mathfrak{A}$.

Consider the distributional equivalence relation, denoted "$\stackrel{\mathcal{D}}{\sim}$", between random variables. For $\zeta \in \mathcal{Z}^*$ define its orbit

$$\mathcal{O}(\zeta) := \{\eta \in \mathcal{Z}^* : \eta \stackrel{\mathcal{D}}{\sim} \zeta\}.$$

It is said that the set $\mathfrak{C}$ is *law invariant* if $\zeta \in \mathfrak{C}$ and $\zeta' \stackrel{\mathcal{D}}{\sim} \zeta$ imply that $\zeta' \in \mathfrak{C}$; in other words if $\zeta \in \mathfrak{C}$, then $\mathcal{O}(\zeta) \subset \mathfrak{C}$. We have the following result connecting law invariance of ρ and $\mathfrak{C}$ (cf., [6]). This can be compared with [4, Proposition 6.29 and Corollary 6.30].

Theorem 1 *(i) If the set $\mathfrak{C}$ is law invariant, then the corresponding risk measure ρ is law invariant. (ii) Conversely, if the risk measure ρ is law invariant and the set $\mathfrak{C}$ is convex and weakly* closed, then $\mathfrak{C}$ is law invariant.*

The above result holds for any probability space $(\Omega, \mathcal{F}, P)$, without assuming that the reference probability measure P is nonatomic. In fact it is shown in [6] that if a functional $\varrho : \mathcal{Z} \to \mathbb{R}\cup\{+\infty\}$ is law invariant, then its conjugate $\varrho^*(\zeta) = \sup_{Z\in\mathcal{Z}} \langle \zeta, Z\rangle - \varrho(Z)$ is law invariant. Thus if ϱ^* is law invariant and $\varrho^{**} = \varrho$, then ϱ is law invariant. Theorem 1 is a particular case of this, since if the set $\mathfrak{C}$ is law invariant, then its indicator function $\mathbb{I}_{\mathfrak{C}}$ is law invariant and the conjugate of $\mathbb{I}_{\mathfrak{C}}$ is ρ, and since the conjugate of the coherent risk measure ρ is the indicator function $\mathbb{I}_{\mathfrak{A}}$ where $\mathfrak{A}$ is the dual set of ρ.

In [4] the interval $[0, 1]$ equipped with its Borel sigma algebra $\mathcal{B}$ and uniform distribution U is called the standard uniform probability space. In literature on probability theory the term 'standard probability space' sometimes is used for a different concept. Anyway to simplify the notation we drop here the word 'uniform' and call this space 'standard probability space'. A random variable Z defined on the standard probability space is simply a measurable function $Z : [0, 1] \to \mathbb{R}$. For the standard probability space $([0, 1], \mathcal{B}, U)$ we simply write $\mathcal{L}_p$ for the space $L_p([0, 1], \mathcal{B}, U)$. Recall that $H_Z(z) = P(Z \le z)$ denotes the cdf of random variable Z and $H_Z^{-1}(t) = \inf\{z : H_Z(z) \ge t\}$. The function H_Z^{-1} is defined on the interval [0,1] (it can take value $-\infty$ at $t = 0$ and value $+\infty$ at $t = 1$), is monotonically nondecreasing and can be viewed as a random variable defined on the standard probability space. We can write the cdf of H_Z^{-1} as $U(H_Z^{-1} \le z) = H_Z(z)$, i.e., random variable Z, defined on a probability space $(\Omega, \mathcal{F}, P)$, and H_Z^{-1} defined on the standard probability space, have the same distribution. In particular if Z is also defined on the standard probability space, then $Z \stackrel{\mathcal{D}}{\sim} H_Z^{-1}$.

It follows that

$$\int_\Omega |Z(\omega)|^p dP(\omega) = \int_0^1 |H_Z^{-1}(t)|^p dt. \tag{2}$$

That is the mapping $L_p(\Omega, \mathcal{F}, P) \ni Z \mapsto H_Z^{-1}$ preserves the $\|\cdot\|_p$ norm in the respective spaces $L_p(\Omega, \mathcal{F}, P)$ and $\mathcal{L}_p$. We denote this mapping as $\mathbb{M}$, i.e.,

$$\mathbb{M}(Z) := H_Z^{-1}. \tag{3}$$

In particular, if ζ is a density function, then $1 = \int_\Omega \zeta(\omega)dP(\omega) = \int_0^1 H_\zeta^{-1}(t)dt$, and hence H_ζ^{-1} is a spectral function. We can assume that spectral functions are left-side or right-side continuous (in [4] spectral functions were assumed to be right-side continuous, while H_ζ^{-1} actually are left-side continuous).

It is possible to show that if the reference probability measure P is nonatomic, then for $Z \in \mathcal{Z}$ and $\zeta \in \mathcal{Z}^*$ it follows that (cf., [2, Lemma 4.55])

$$\sup_{\eta \in \mathcal{O}(\zeta)} \langle \eta, Z \rangle = \int_0^1 H_Z^{-1}(t) H_\zeta^{-1}(t) dt. \tag{4}$$

Hence assuming that the reference probability measure P is nonatomic, we can write the dual representation (1) for law invariant coherent ρ as

$$\rho(Z) = \sup_{\zeta \in \mathfrak{C},\, \eta \in \mathcal{O}(\zeta)} \langle \eta, Z \rangle = \sup_{\zeta \in \mathfrak{C}} \int_0^1 H_Z^{-1}(t) H_\zeta^{-1}(t) dt.$$

and hence

$$\rho(Z) = \sup_{\sigma \in \Upsilon} \int_0^1 \sigma(t) H_Z^{-1}(t) dt, \tag{5}$$

where

$$\Upsilon := \mathbb{M}(\mathfrak{C}) = \{\sigma = H_\zeta^{-1} : \zeta \in \mathfrak{C}\} \tag{6}$$

is a set of spectral functions. Note again that by (2) if $\mathfrak{C}$ is a subset of $L_q(\Omega, \mathcal{F}, P) = \mathcal{Z}^*$, then $\mathbb{M}(\mathfrak{C})$ is a subset of $\mathcal{L}_q$.

- Assuming that the reference probability measure P is nonatomic, we have that if a law invariant risk measure $\rho : \mathcal{Z} \to \mathbb{R} \cup \{+\infty\}$ has the dual representation (1) for some set $\mathfrak{C} \subset \mathcal{Z}^*$ of density functions, then it can be written in the form (5) with the corresponding set $\Upsilon = \mathbb{M}(\mathfrak{C})$ of spectral functions.

The representation (5) can be viewed as a dual representation on the standard probability space. In [4, Definition 6.33] the set Υ is called *generating set*. Since the set $\mathfrak{C}$ in the dual representation (1) is not unique, the corresponding generating set is not defined uniquely by ρ. Therefore it makes sense to look for, in some sense, minimal generating set. In that respect we have the following[1], [3], [4, Proposition 6.34].

Theorem 2 *Suppose that the reference probability measure P is nonatomic and $\mathcal{Z} = L_p(\Omega, \mathcal{F}, P)$, $p \in [1, \infty)$. Let $\rho : \mathcal{Z} \to \mathbb{R}$ be a real valued law invariant coherent risk measure and* $\mathrm{Exp}(\mathfrak{A})$ *be the set of exposed points of its dual set $\mathfrak{A}$. Then the representation* (1) *holds with* $\mathfrak{C} := \mathrm{Exp}(\mathfrak{A})$. *Moreover, if the representation* (1) *holds for some weakly* closed set $\mathfrak{C}$, then* $\mathrm{Exp}(\mathfrak{A}) \subset \mathfrak{C}$.

The above theorem shows that, in a certain sense, the set $\mathfrak{C} := \mathrm{Exp}(\mathfrak{A})$ is minimal in the dual representation (1), and hence the corresponding generating set $\Upsilon = \mathbb{M}(\mathrm{Exp}(\mathfrak{A}))$ is minimal in the representation (5). In particular if $\Upsilon = \{\sigma\}$ is a singleton, then the corresponding risk measure ρ is called spectral.

Kusuoka representation (cf., [4, Section 6.3.5]).

By using transformation

$$(\mathbb{T}\mu)(t) = \int_0^t (1-\alpha)^{-1} d\mu(\alpha),\ t \in [0,1),$$

[1]Recall that $\mathrm{Exp}(\mathfrak{A})$ denotes the set of exposed point of the dual set $\mathfrak{A}$.

and its inverse

$$(\mathbb{T}^{-1}\sigma)(\alpha) = (1-\alpha)\sigma(\alpha) + \int_0^\alpha \sigma(t)dt, \ \alpha \in [0,1),$$

the Kusuoka representation

$$\rho(Z) = \sup_{\mu \in \mathfrak{M}} \int_0^1 \mathsf{AVaR}_{1-\alpha}(Z) d\mu(\alpha), \ Z \in \mathcal{Z}, \tag{7}$$

is obtained from (5) with $\mathfrak{M} = \mathbb{T}^{-1}(\Upsilon)$. Recall that

$$\mathsf{AVaR}_{1-\alpha}(Z) = \inf_{\tau \in \mathbb{R}} \left\{ \tau + (1-\alpha)^{-1} \mathbb{E}[Z-\tau]_+ \right\}. \tag{8}$$

The mapping $\mathbb{T}$ is one-to-one from the set of probability measures on the interval[2] $[0,1)$ onto the set of spectral functions. Consider the set $\mathfrak{P}_q$ of probability measures on $[0,1)$ such that $\mathbb{T}\mu \in \mathcal{L}_q$, $q \in (1,\infty]$. We have the following result, [3, Proposition 3.4].

Proposition 1 *The set $\mathfrak{P}_q$ is closed and the mapping $\mathbb{T}$ is continuous on the set $\mathfrak{P}_q$ with respect to the weak topology of measures and the weak* topology of $\mathcal{Z}^* = \mathcal{L}_q$.*

It follows that if $\mathfrak{C}$ is a weakly* closed subset of the dual set $\mathfrak{A} \subset \mathcal{Z}^*$ and $\Upsilon = \mathbb{M}(\mathfrak{C})$, then $\mathbb{T}^{-1}(\Upsilon)$ is a weakly closed subset of $\mathfrak{P}_q$.

Let $\mathcal{Z} = L_p(\Omega, \mathcal{F}, P)$, $p \in [1,\infty)$, and $\rho : \mathcal{Z} \to \mathbb{R}$ be a law invariant coherent risk measure. Since ρ is law invariant, it can be considered as a function $\rho(H)$ of cdfs $H = H_Z$, $Z \in \mathcal{Z}$. It is possible to write the Kusuoka representation (7) of ρ in the following minimax form. Using variational representation (8) of the Average Value-at-Risk, we can write

$$\rho(H) = \sup_{\mu \in \mathfrak{M}} \int_0^1 \inf_{\tau \in \mathbb{R}} \left\{ \int_{-\infty}^{+\infty} \psi_\alpha(z,\tau) dH(z) \right\} d\mu(\alpha), \tag{9}$$

where

$$\psi_\alpha(z,\tau) := \tau + (1-\alpha)^{-1}[z-\tau]_+, \ \alpha \in [0,1). \tag{10}$$

For $\alpha \in (0,1)$ the minimizer inside the integral in (9) is given by $\bar{\tau}(\alpha) = H^{-1}(\alpha)$. For $H^{-1} \in \mathcal{L}_p$, by interchanging the integral and minimization operators in the right-hand side of (9) we can write

$$\begin{aligned} \rho(H) &= \sup_{\mu \in \mathfrak{M}} \inf_{y \in \mathcal{L}_p} \int_0^1 \int_{-\infty}^{+\infty} \psi_\alpha(z, y(\alpha)) dH(z) d\mu(\alpha) && (11) \\ &= \sup_{\mu \in \mathfrak{M}} \inf_{y \in \mathcal{L}_p} \int_{-\infty}^{+\infty} \int_0^1 \psi_\alpha(z, y(\alpha)) d\mu(\alpha) dH(z). && (12) \end{aligned}$$

Note that minimization in (11) and (12) is performed over functions $y \in \mathcal{L}_p$. By interchanging the 'sup' and 'inf' operators we can write the dual of problem (12):

$$\inf_{y \in \mathcal{L}_p} \sup_{\mu \in \mathfrak{M}} \int_{-\infty}^{+\infty} \int_0^1 \psi_\alpha(z, y(\alpha)) d\mu(\alpha) dH(z). \tag{13}$$

For $Z \in \mathcal{Z} = L_p(\Omega, \mathcal{F}, P)$ and $H = H_Z$ consider the set

$$\bar{\mathfrak{A}} := \arg\max_{\zeta \in \mathfrak{A}} \langle \zeta, Z \rangle.$$

[2] We assume that these measures have mass zero at $\alpha = 1$.

This set is nonempty since the dual set $\mathfrak{A}$ is weakly* compact. The respective sets $\bar{\Upsilon} := \mathbb{M}(\bar{\mathfrak{A}})$ and $\bar{\mathfrak{M}} := \mathbb{T}^{-1}(\bar{\Upsilon})$ are also nonempty, and $\bar{\mathfrak{M}}$ is the set of maximizers in (11) and the right hand side of (9).

Proposition 2 *Suppose that $H^{-1} \in \mathcal{L}_p$, $p \in [1, \infty)$, and every $\mu \in \mathfrak{M}$ has mass zero at $\alpha = 0$. Then for any $\bar{y}(\alpha) := H^{-1}(\alpha)$, $\alpha \in (0,1)$, and $\bar{\mu} \in \bar{\mathfrak{M}}$, the point $(\bar{y}, \bar{\mu})$ is a saddle point of problems* (12) *and* (13), *and there is no duality gap between problems* (12) *and* (13).

Proof. Suppose that every $\mu \in \mathfrak{M}$ has mass zero at $\alpha = 0$ (as it was pointed above every $\mu \in \mathfrak{M}$ has mass zero at $\alpha = 1$). Consider $\bar{y}(\alpha) := H^{-1}(\alpha)$, $\alpha \in (0,1)$. Recall that $\bar{y}(\cdot)$ is a minimizer of $\int_{-\infty}^{+\infty} \psi_\alpha(z, y(\alpha)) dH(z)$. It follows that for every $\mu \in \mathfrak{M}$, and in particular for $\mu \in \bar{\mathfrak{M}}$, and all $y \in \mathcal{L}_p$,

$$\int_0^1 \int_{-\infty}^{+\infty} \psi_\alpha(z, y(\alpha)) dH(z) d\mu(\alpha) \geq \int_0^1 \int_{-\infty}^{+\infty} \psi_\alpha(z, \bar{y}(\alpha)) dH(z) d\mu(\alpha).$$

Since $\bar{\mu} \in \bar{\mathfrak{M}}$ is a maximizer in the right hand side of (9), it follows that

$$\int_0^1 \int_{-\infty}^{+\infty} \psi_\alpha(z, \bar{y}(\alpha)) dH(z) d\bar{\mu}(\alpha) \geq \int_0^1 \int_{-\infty}^{+\infty} \psi_\alpha(z, \bar{y}(\alpha)) dH(z) d\mu(\alpha)$$

for any $\mu \in \mathfrak{M}$. This shows that $(\bar{y}, \bar{\mu})$ is a saddle point. Existence of a saddle point implies that there is no duality gap between problems (12) and (13). ■

If $H(z) = 0$ for some $z \in \mathbb{R}$, then $H^{-1}(0)$ is finite. In that case in order to conclude that $(\bar{y}, \bar{\mu})$ is a saddle point there is no need to assume that every $\mu \in \mathfrak{M}$ has mass zero at $\alpha = 0$. Also if ρ can be written as the convex combination

$$\rho(H) = w\mathbb{E}_H[Z] + (1 - w) \sup_{\mu \in \mathfrak{M}} \int_0^1 \mathsf{AVaR}_{1-\alpha}(H) d\mu(\alpha), \; w \in [0,1], \tag{14}$$

with every $\mu \in \mathfrak{M}$ having mass zero at $\alpha = 0$, then the conclusions of Proposition 2 follow.

References

[1] Ding, L. and Shapiro, A., Upper bound for optimal value of risk averse multistage problems, http://www2.isye.gatech.edu/people/faculty/Alex_Shapiro/UB-paper.pdf

[2] Föllmer, H. and Schied, A., *Stochastic Finance: An Introduction in Discrete Time*, Walter de Gruyter, Berlin, 2nd ed., 2004.

[3] Pichler, A. and Shapiro, A., Minimal Representation of Insurance Prices, *Insurance: Mathematics and Economics*, 62, 184–193, 2015.

[4] Shapiro, A., Dentcheva, D. and Ruszczyński, A., *Lectures on Stochastic Programming: Modeling and Theory*, second edition, SIAM, Philadelphia, 2014.

[5] Shapiro, A., Interchangeability principle and dynamic equations in risk averse stochastic programming, *Operations Research Letters*, 45, 377–381, 2017.

[6] Shapiro, A., Distributionally robust stochastic programming, *SIAM J. Optimization*, 2017, to appear.