Part II: Multiperiod Functionals and Information Monotonicity

Georg Ch. Pflug

Winter School on Mathematical Finance, Lunteren 2009

Multi-period risk functionals

Let $Y = (Y_1, \ldots, Y_T)$ be an income process on some probability space $(\Omega, \mathcal{F}, P)$ and let $\boldsymbol{\mathcal{F}} = (\mathcal{F}_0, \ldots, \mathcal{F}_T)$ denote a filtration which models the available information over time, where $\mathcal{F}_0 = \{\emptyset, \Omega\}$, $\mathcal{F}_T = \mathcal{F}$, $\mathcal{F}_t \subseteq \mathcal{F}_{t+1} \subseteq \mathcal{F}$, and Y_t is $\mathcal{F}_t$ measurable for every $t = 1, \ldots, T$. Let $\mathcal{Y} \subseteq \times_{t=1}^{T} L_1(\Omega, \mathcal{F}, P)$ be a linear space of income processes $Y = (Y_1, \ldots, Y_T)$, which are all adapted to $\boldsymbol{\mathcal{F}}$.

Definition. A multi-period functional $\mathcal{A}$ with values $\mathcal{A}(Y; \boldsymbol{\mathcal{F}})$ is called *multi-period acceptability functional*, if satisfies

(MA0) **Information monotonicity.** If $Y \in \mathcal{Y}$ and $\mathcal{F}_t \subseteq \mathcal{F}'_t$, for all t, then

$$\mathcal{A}(Y; \mathcal{F}_0, \ldots, \mathcal{F}_{T-1}) \leq \mathcal{A}(Y; \mathcal{F}'_0, \ldots, \mathcal{F}'_{T-1}).$$

(MA1) **Predictable translation-equivariance.** If $W \in \mathcal{Y}$ such that W_t is $\mathcal{F}_{t-1}$ measurable for all t, then

$$\mathcal{A}(Y + W; \boldsymbol{\mathcal{F}}) = \sum_{t=1}^{T} \mathbb{E}(W_t) + \mathcal{A}(Y; \boldsymbol{\mathcal{F}}).$$

(MA2) **Concavity.** The mapping $Y \mapsto \mathcal{A}(Y;\boldsymbol{\mathcal{F}})$ is concave on $\mathcal{Y}$ for all filtrations $\boldsymbol{\mathcal{F}}$.

(MA3) **Monotonicity.** If $Y_t \leq \tilde{Y}_t$ holds a.s. for all t, then

$$\mathcal{A}(Y;\boldsymbol{\mathcal{F}}) \leq \mathcal{A}(\tilde{Y};\boldsymbol{\mathcal{F}}).$$

(MA1)* $(\pi, \mathcal{W})$-**translation-equivariance.** There exists a linear subspace $\mathcal{W}$ of $\times_{t=1}^{T} L_1(\Omega, \mathcal{F}_{t-1}, \mathbb{P})$ and a linear continuous functional $\pi : \mathcal{W} \to \mathbb{R}$ such that

$$\mathcal{A}(Y + W;\boldsymbol{\mathcal{F}}) = \pi(W) + \mathcal{A}(Y;\boldsymbol{\mathcal{F}})$$

holds for all $W \in \mathcal{W}$, $Y \in \mathcal{Y}$ and all filtrations $\boldsymbol{\mathcal{F}}$.

Special cases are the predictable translation equivariance (MA1), the *weak translation equivariance* ($\mathcal{W}$ consists only of constant functions) or
the *first period translation equivariance* ($\mathcal{W}$ consists of all $\mathcal{F}_1$ measurable functions).

Version independence in multi-period risk measuring

We identify filtrations with equivalence classes (by bijection) of tree processes.

Definition. ν is a tree process, iff the σ-fields generated by ν_t form a filtration (an increasing sequence of σ-fields).

We assume that the filtration $\mathcal{F}$ is generated by a tree process ν (with values in a Polish space) and that the scenario process $Y = (Y_1, \ldots, Y_T)$ is adapted to it. We call

$$(Y, \mathcal{F}) \quad \text{resp. } (Y, \nu)$$

a process-and-information pair. Notice that there are functions f_t such that

$$Y_t = f_t(\nu_t) \qquad \text{a.e.}$$

Artzner's Example

A fair coin is tossed three times. The payoff process $Y = (Y_1, Y_2, Y_3)$ is

$$Y_1 = 0; \quad Y_2 = 0;$$

$$Y_3 = \begin{cases} 1 & \text{if heads is shown at least two times} \\ 0 & \text{otherwise} \end{cases}$$

We compare this process to another payoff process

$$\tilde{Y}_1 = 0; \quad \tilde{Y}_2 = 0;$$

$$\tilde{Y}_3 = \begin{cases} 1 & \text{if heads is shown at the last throw} \\ 0 & \text{otherwise} \end{cases}$$

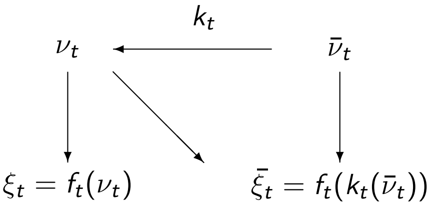
Equivalence

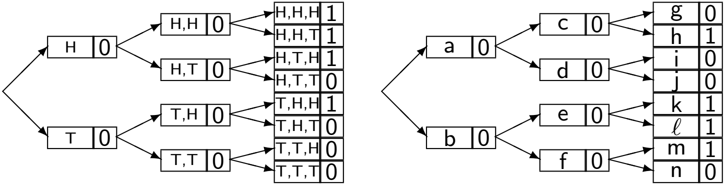
Definition. Two process-and-information pairs (ξ, ν) and $(\bar{\xi}, \bar{\nu})$ (which are defined on possibly different probability spaces) are *equivalent*, if there are bijective measurable functions k_t such that

(i) $k_t(\bar{\nu}_t)$ has the same distribution as ν_t.

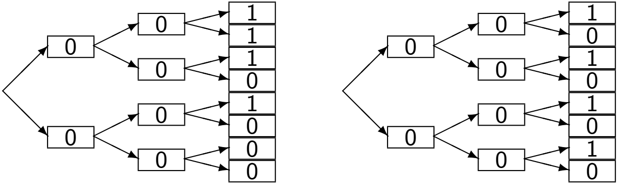
(ii) $\bar{\xi}_t = f_t(k_t(\bar{\nu}_t)) \qquad \text{a.s.} \qquad t = 1, \ldots, T,.$

The following diagram illustrates the notion of equivalence.





Equivalent process-and-information pairs.



Non-equivalent process and information pairs

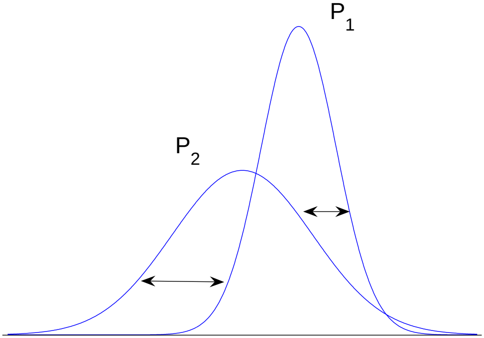
Nested distributions

Let (Ξ, d) be a Polish space, i.e. complete separable metric space and let $\mathcal{P}_1(\Xi, d)$ be the family of all Borel probability measures P on (Ξ, d) such that $\int d(u, u_0)\, dP(u) < \infty$ for some $u_0 \in \Xi$.
For two Borel probabilities, P and Q in $\mathcal{P}_1(\Xi, d)$, let $d(P, Q)$ denote the Kantorovich distance

$$
\begin{aligned}
d(P, Q) &= \inf\{\mathbb{E}[d(X, Y)] : X \sim P, Y \sim Q\} \\
&= \sup\{\int h(u)\, dP(u) - \int h(u)\, dQ(u) : |h(u) - h(v)| \leq d(u, v)\}
\end{aligned}
$$

d metrizises the weak topology on $\mathcal{P}_1$. On finite probability spaces, d can be found by solving a linear optimization problem.
$\mathcal{P}_1$ is a complete separable metric space (Polish space) under d.
Iterate the argument: $\mathcal{P}_1(\mathcal{P}_1(\Xi, d), d)$ is a Polish space, a space of distributions over distributions (i.e. what Bayesians would call a hyperdistribution).

Illustration of the Kantorovich distance



If (Ξ_1, d_1) and (Ξ_2, d_2) are Polish spaces then so is the Cartesian product $(\Xi_1 \times \Xi_2)$ with metric

$$d_{1,2}((u_1, u_2), (v_1, v_2)) = d_1(u_1, v_1) + d_2(u_2, v_2).$$

For a scenario process with values in $\mathbb{R}^{m_1} \times \mathbb{R}^{m_2} \times \cdots \times \mathbb{R}^{m_t}$, we consider some metric d_t on $\mathbb{R}^{m_t}$, which makes it Polish (it needs not be the Euclidean one). Then we define the following spaces

$$\begin{aligned}
\Xi_{T:T} &= (\mathbb{R}^{m_T}, d_T) \\
\Xi_{T-1:T} &= (\mathbb{R}^{m_{T-1}} \times \mathcal{P}_1(\Xi_{T:T}, d_T), d_{T-1,T}) = (\mathbb{R}^{m_{T-1}} \times \mathcal{P}_1(\mathbb{R}^{m_T}, d_T) \\
\Xi_{T-2:T} &= (\mathbb{R}^{m_{T-2}} \times \mathcal{P}_1(\Xi_{T-1:T}, d_{T-1,T}), d_{T-2,T-1,T}) \\
&= (\mathbb{R}^{m_{T-2}} \times \mathcal{P}_1(\mathbb{R}^{m_{T-1}} \times \mathcal{P}_1(\mathbb{R}^{m_T}, d_T), d_{T-1,T}), d_{T-2,T-1,T}) \\
&\;\;\vdots \\
\Xi_{1:T} &= (\mathbb{R}^{m} \times \mathcal{P}_1(\Xi_{2:T}, d_{2,\dots,T}), d_{1,\dots,T})
\end{aligned}$$

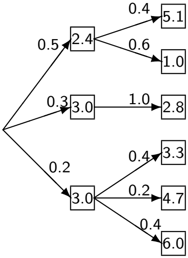
Definition. A Borel probability distribution $\mathbb{P}$ on $\Xi_{1:T}$ is called a *nested distribution of depth T*. (see also Vershik, 1995)
For any nested distribution $\mathbb{P}$, there is an embedded multivariate distribution P. The projection from the nested distribution to the embedded distribution is not injective. Notation for discrete distributions:

probabilities: / values:

$$\begin{bmatrix} 0.3 & 0.4 & 0.3 \\ \hline 3.0 & 1.0 & 5.0 \end{bmatrix} \qquad \begin{bmatrix} 0.4 & 0.3 & 0.3 \\ \hline 1.0 & 5.0 & 3.0 \end{bmatrix} \qquad \begin{bmatrix} 0.1 & 0.2 & 0.4 & 0.3 \\ \hline 3.0 & 3.0 & 1.0 & 5.0 \end{bmatrix}$$

Left: A valid distribution. Middle: the same distribution. Right: Not a valid distribution

Examples for nested distributions



$$\left[\begin{array}{ccc} 0.2 & 0.3 & 0.5 \\ \hline 3.0 & 3.0 & 2.4 \\ \left[\begin{array}{ccc} 0.4 & 0.2 & 0.4 \\ \hline 6.0 & 4.7 & 3.3 \end{array}\right] & \left[\begin{array}{c} 1.0 \\ \hline 2.8 \end{array}\right] & \left[\begin{array}{cc} 0.6 & 0.4 \\ \hline 1.0 & 5.1 \end{array}\right] \end{array}\right]$$

The embedded multivariate, but non-nested distribution of the scenario process can be gotten from it:

$$\left[\begin{array}{cccccc} 0.08 & 0.04 & 0.08 & 0.3 & 0.3 & 0.2 \\ \hline 3.0 & 3.0 & 3.0 & 3.0 & 2.4 & 2.4 \\ 6.0 & 4.7 & 3.3 & 2.8 & 1.0 & 5.1 \end{array}\right]$$

Version independence

Let $Y = (Y_1, \ldots, Y_T)$ be a stochastic process adapted to the filtration $\boldsymbol{\mathcal{F}} = (\mathcal{F}_1, \ldots, \mathcal{F}_T)$. The functional $\mathcal{A}(Y; \boldsymbol{\mathcal{F}})$ is called version-independent, if it depends only on the nested distribution $\mathbb{P}$ generated by the pair $(Y; \boldsymbol{\mathcal{F}})$.

$$
\begin{array}{ccc}
(Y; \boldsymbol{\mathcal{F}}) & \longrightarrow & \mathbb{P} \\
 & \searrow & \downarrow \\
 & & \mathcal{A}(Y; \boldsymbol{\mathcal{F}}) = \mathcal{A}\{\mathbb{P}\}
\end{array}
$$

Construction of multi-period risk functionals

(a) Separable multi-period acceptability functionals:

$$\mathcal{A}(Y;\boldsymbol{\mathcal{F}}) := \sum_{t=1}^{T} \mathcal{A}_t(Y_t),$$

where $\mathcal{A}_t$ are single-period acceptability functionals. They (MA1)', (MA2) and (MA3), but do not depend on $\boldsymbol{\mathcal{F}}$.

(b) Scalarization:

$$\mathcal{A}(Y;\boldsymbol{\mathcal{F}}) := \mathcal{A}_0(s(Y))$$

where $\mathcal{A}_0$ is a (single-period) acceptability functional and $s : \mathcal{Y} \to L_1(\Omega, \mathcal{F}, \mathbb{P})$ a mapping satisfying concavity, monotonicity and $s(Y_1 + r, Y_2, \ldots, Y_T) := s(Y_1, \ldots, Y_T) + r$ for all $Y \in \mathcal{Y}$ and $r \in \mathbb{R}$. Such functionals satisfy (MA1)$''$, (MA2) and (MA3), but do not depend on $\boldsymbol{\mathcal{F}}$ (Eichhorn and Roemisch). Examples:

(i) $s(Y) = \sum_{t=1}^{T} Y_t$.

(ii) $s(Y) := \min_{t=1,\ldots,T} \sum_{\tau=1}^{t} Y_\tau$.

(c) Separable expected conditional (SEC) multi-period acceptability functionals:

$$\mathcal{A}(Y;\boldsymbol{\mathcal{F}}) := \sum_{t=1}^{T} \mathbb{E}(\mathcal{A}_t(Y_t|\mathcal{F}_{t-1}))$$

where $\mathcal{A}_t(\cdot\,|\mathcal{F}_{t-1})$, $t = 1, \dots, T$, are conditional (single-period) acceptability functionals. They satisfy (MA0)–(MA3). The dual representation is

$$\begin{aligned}
&\textstyle\sum_{t=1}^{T} \mathbb{E}[\mathcal{A}_t(Y_t|\mathcal{F}_{t-1})] \\
&= \inf\{\textstyle\sum_{t=1}^{T} \mathbb{E}[Y_t Z_t] - \sum_{t=1}^{T} \mathbb{E}[\mathcal{A}_t^{+}(Z_t|\mathcal{F}_t)] : Z_t \lhd \mathcal{F}_t, \\
&\qquad Z_t \geq 0, \mathbb{E}(Z_t|\mathcal{F}_{t-1}) = 1\} \\
&= \textstyle\sum_{t=1}^{T} \inf\{\mathbb{E}[Y_t Z_t] - \mathbb{E}[\mathcal{A}_t^{+}(Z_t|\mathcal{F}_t) : Z_t \lhd \mathcal{F}_t, \\
&\qquad Z_t \geq 0, \mathbb{E}(Z_t|\mathcal{F}_{t-1}) = 1\}.
\end{aligned}$$

Remark. If $\mathcal{A}(Y;\boldsymbol{\mathcal{F}})$ is SEC functional, then also its conjugate of $\mathcal{A}^{+}(\cdot;\boldsymbol{\mathcal{F}})$ is SEC. $\mathcal{A}(Y;\boldsymbol{\mathcal{F}})$ is information-monotone, iff $\mathcal{A}^{+}(\cdot;\boldsymbol{\mathcal{F}})$ is information-monotone. $\mathcal{A}(Y;\boldsymbol{\mathcal{F}})$ is information-monotone, if all $\mathcal{A}_t$'s are compound convex.

Proof. If $\mathcal{A}_t$ is compound convex and $\mathcal{F}_t \subseteq \mathcal{F}'_t$, then

$$\mathcal{A}(Y|\mathcal{F}_t) \le \mathbb{E}[\mathcal{A}(Y|\mathcal{F}'_t)|\mathcal{F}_t].$$

Taking the expectation on both sides, one gets

$$\mathbb{E}[\mathcal{A}(Y|\mathcal{F}_t)] \le \mathbb{E}[\mathcal{A}(Y|\mathcal{F}'_t)].$$

Example. (Multi-period Average Value-at-Risk)

$$m\mathbb{A}\mathrm{V@R}_\alpha(Y;\boldsymbol{\mathcal{F}}) := \sum_{t=1}^{T} \mathbb{E}(\mathbb{A}\mathrm{V@R}_\alpha(Y_t|\mathcal{F}_{t-1}))$$

$$= \inf\left\{\sum_{t=1}^{T} \mathbb{E}(Y_t Z_t) : 0 \le Z_t \le \frac{1}{\alpha}, \mathbb{E}(Z_t|\mathcal{F}_{t-1}) = 1, t = 1,\ldots,T\right\}$$

The multi-period $\mathbb{A}$V@R is information-monotone.

Multiperiod polyhedral risk measures

It is a natural idea to introduce acceptability and risk functionals as optimal values of certain stochastic programs.

Definition. (Eichhorn and Roemisch, 2005)
A multi-period functional $\mathcal{A}$ on $\times_{t=1}^{T} L_p(\mathcal{F}_t)$ is called *polyhedral* if there are $k_t \in \mathbb{N}$, $c_t \in \mathbb{R}^{k_t}$, $t = 1, \ldots, T$, $w_{t\tau} \in \mathbb{R}^{k_{t-\tau}}$, $t = 1, \ldots, T$, $\tau = 0, \ldots, t-1$, (convex) polyhedral sets $V_t \subset \mathbb{R}^{k_t}$, $t = 1, \ldots, T$, such that

$$\mathcal{A}(Y) = \sup \left\{ \mathbb{E}\left[\sum_{t=1}^{T} \langle c_t, v_t \rangle \right] \middle| \begin{array}{l} v_t \in L_p(\mathcal{F}_t; \mathbb{R}^{k_t}),\ v_t \in V_t, \\ \sum_{\tau=0}^{t-1} \langle w_{t,\tau}, v_{t-\tau} \rangle = Y_t,\ \ t = 1, \ldots, T \end{array} \right\}.$$

There exist multi-period polyhedral acceptability functionals satisfying (MA0), (MA1)* (first period translation equivariance), (MA2), (MA3), (strictness, positive homogeneity).

Multi-period polyhedral acceptability functionals preserve linearity, decomposition structures and stability properties of multi-stage stochastic programming models.

Examples.

(a) *Multi-period average value-at-risk m$\mathbb{A}$V@R.*

(b) $\mathcal{A}(Y) := \mathbb{A}\mathrm{V@R}_\alpha(\sum_{\tau=1}^{t(\cdot)} Y_\tau)$, where $t(\cdot)$ is uniformly distributed on $\{1, \ldots, T\}$ and independent of $(Y_\tau)_{\tau=1}^{T}$, is polyhedral (Eichhorn).

(c) $\mathcal{A}(Y) := \mathbb{A}\mathrm{V@R}_\alpha(\min_t \sum_{\tau=1}^{t} Y_\tau)$

These acceptability mappings satisfy (MA0), (MA1*), (MA2), (MA3) and positive homogeneity.

A stylized dynamic optimization problem: multi-stage inventory control

Suppose that the demand (say for grapefruits) at times $t = 1, \ldots, T$ is given by a random process $\xi_1, \ldots, \xi_T$. The grocery shop has to place regular orders one period ahead. The costs for ordering one piece 1. If the demand exceeds the inventory plus the newly arriving order, the demands has to be fulfilled by rapid orders (which are immediately delivered), for a price of $u_t > 1$ per piece. Unsold grapefruits may be stored in the inventory, but a fraction $1 - \ell_t$ is storage loss. The selling price is s_t ($s_t > 1$) and the final inventory K_T has a value of $\ell_T K_T$.
Let K_t be the inventory volume right after all sales have been effectuated at time t. Let x_t be the order size at time t. We have that $K_0 = 0$ and $K_t = [\ell_{t-1} K_{t-1} + x_{t-1} - \xi_t]^+; t = 1, \ldots, T$. The shortage at time t is $M_t = [\ell_{t-1} K_{t-1} + x_{t-1} - \xi_t]^-; t = 1, \ldots, T$.

These two equations can be merged into

$$\ell_{t-1}K_{t-1} + x_{t-1} - \xi_t = K_t - M_t; \; K_t \geq 0, \; M_t \geq 0. \qquad (1)$$

The profit of the whole operation is

$$H(x_0, \xi_1, \ldots, x_{T-1}, \xi_T) = \sum_{t=1}^{T} s_t \xi_t - \sum_{t=0}^{T-1} x_t - \sum_{t=1}^{T} u_t M_t + \ell_T K_T.$$

The problem is to maximize the expected profit

$$\text{Maximize } \mathbb{E}\Big[\sum_{t=1}^{T}(s_t \xi_t - x_{t-1} - u_t M_t) + \ell_T K_T\Big]$$

subject to $x_t \lhd \mathcal{F}_t$ for $t = 1, \ldots, T$;

and subject to (1).

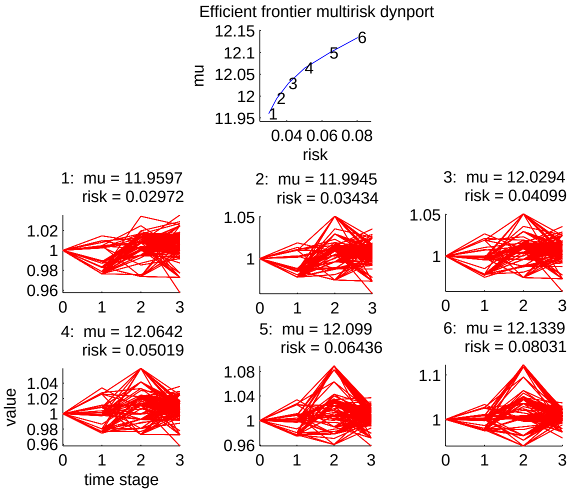
Notice that $\mathbb{E}[\sum_{t=1}^{T} s_t \xi_t]$ does not depend on the decisions and can be removed from the optimization problem.

The optimum value m^* of this problem is

$$m^* = \sum_{t=1}^{T} u_t \mathbb{E}(-\xi_t) + \sum_{t=1}^{T} (u_t - 1)\mathbb{E}[\mathbb{A}\mathsf{V@R}_{\beta_t}(\xi_t|\mathcal{F}_{t-1})]$$

with $\beta_t = (u_t - 1)/(\ell_t - 1)$. The optimal order sizes x_t are given by the Values-at-Risk $x_t = \mathbb{V}\mathsf{@R}_{\beta_{t+1}}(\xi_{t+1}|\mathcal{F}_t) - \ell_t K_t$. with $K_t = [\mathbb{V}\mathsf{@R}_{\beta_t}(\xi_t|\mathcal{F}_{t-1}) - \xi_t]^+$.

Example: dynamic portfolio management



An efficient frontier using the (negative) multiperiod $\mathbb{A}$V@R as risk functional

Minimal filtrations and randomized decisions

Not all process and information pairs allow a representation as a nested distribution.

$$\left[\begin{array}{cc} 0.5 & 0.5 \\ \hline 0 & 0 \\ \left[\begin{array}{cc} 0.5 & 0.5 \\ \hline 0.0 & 1.0 \end{array}\right] & \left[\begin{array}{cc} 0.5 & 0.5 \\ \hline 0.0 & 1.0 \end{array}\right] \end{array}\right] \qquad \left[\begin{array}{c} 1.0 \\ \hline 1.0 \\ \left[\begin{array}{cc} 0.5 & 0.5 \\ \hline 0.0 & 1.0 \end{array}\right] \end{array}\right]$$

Left: Not a valid nested distribution. Right: A valid one

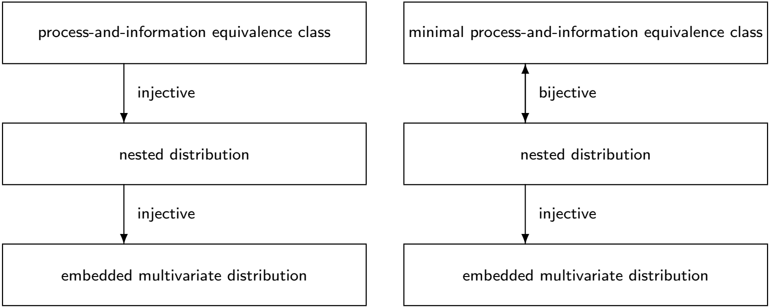
Definition. A filtration is called *minimal*, if it is generated by the process of conditional subtrees. $\mathbb{P}^{\nu_t = u}$.
If a decision is different for identical conditional trees, it may be interpreted as a random decision.

Main Results

Theorem A process-and-information pair is minimal, if and only if the tree process is equivalent to the process of conditional nested distributions.

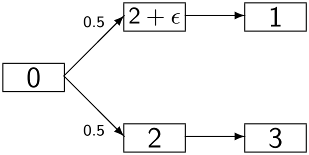
Theorem Two minimal process-and-information pairs are equivalent, if and only if they induce the same nested distribution.

Illustration

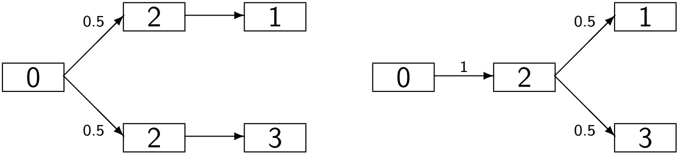


An optimal prediction problem.

(adapted from Heitsch et al. (2006).



The limit for $\epsilon \longrightarrow 0$:



In the nested distribution notation

$$\mathbb{P}_\epsilon = \left[\begin{array}{c} 1.0 \\ \hline 0 \\ \left[\begin{array}{cc} 0.5 & 0.5 \\ \hline 2+\epsilon & 2 \\ \left[\begin{array}{c} 1.0 \\ \hline 1 \end{array}\right] & \left[\begin{array}{c} 1.0 \\ \hline 3 \end{array}\right] \end{array}\right] \end{array}\right] \quad \Rightarrow \quad \mathbb{P}_0 = \left[\begin{array}{c} 1.0 \\ \hline 0 \\ \left[\begin{array}{cc} 0.5 & 0.5 \\ \hline 2 & 2 \\ \left[\begin{array}{c} 1.0 \\ \hline 1 \end{array}\right] & \left[\begin{array}{c} 1.0 \\ \hline 3 \end{array}\right] \end{array}\right] \end{array}\right].$$

First stage optimal decision

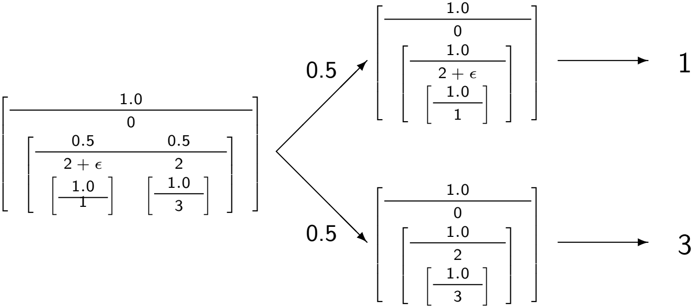
$$x_0 = 2 + \epsilon/2$$

Second stage optimal decision

$$x_1(2+\epsilon) = 1$$
$$x_1(2) = 3$$

Not equi-continuous in ϵ

In standard notation, the tree looks as



Now, viewing the solution as a function of the standard tree process, one gets

First stage optimal decision

$$x_0 = 2 + \epsilon/2$$

Second stage optimal decision

$$\left[\begin{array}{c} \dfrac{1.0}{0} \\ \left[\begin{array}{c} \dfrac{1.0}{2+\epsilon} \\ \left[\dfrac{1.0}{1}\right] \end{array}\right] \end{array}\right] \mapsto 1$$

$$\left[\begin{array}{c} \dfrac{1.0}{0} \\ \left[\begin{array}{c} \dfrac{1.0}{2} \\ \left[\dfrac{1.0}{3}\right] \end{array}\right] \end{array}\right] \mapsto 3$$

Equi-Lipschitz in ϵ w.r.t. d'

Randomized decisions in financial optimization

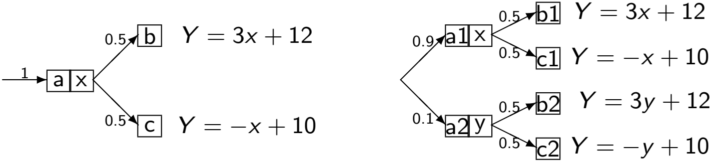
The basic question is: May randomized decisions outperform nonrandomized ones in financial optimization?

The answer is yes.

Example. We want to maximize the following functional

$$\mathcal{A}(Y) = \begin{cases} \mathbb{E}(Y), & \text{if } \mathbb{V}@\mathsf{R}_{0.1}(Y) \geq 10 \\ -\infty, & \text{otherwise.} \end{cases}$$

Here $\mathbb{V}@\mathsf{R}_{0.1}(Y)$ is the 10%-quantile of Y.



Left: The optimal value is $\mathcal{A}(Y) = 11$. Right: The randomized decision gives a better optimal value of $\mathcal{A}(Y) = 11.1$.

An Example with concave objective

Example. Let $\mathcal{A}(Y) = -\mathbb{E}|Y - \mathbb{E}Y|$. $\mathcal{A}$ is concave, but not compound convex. Let $\xi \sim U[0,1]$ and let

$$H(x_1, x_2, \xi) = \begin{cases} x_1 + \mathbb{1}_{\xi \le x_2} & \text{if } 0 \le x_1 \le 1; 0.1 \le x_2 \le 0.9 \\ -\infty & \text{otherwise} \end{cases}$$

The decision which maximizes $\mathcal{A}[H(x_1, x_2, \xi)]$ is: x_1 arbitrary and $x_2 = 0.1$ or 0.9. The objective value is -0.18. If randomization is allowed, one may choose $x_1 = 0$ and $x_2 = 0.9$ with probability $1/2$ and $x_1 = 1$ and $x_2 = 0.1$ also with probability $1/2$. This gives the better value of -0.1 for the objective function.

We consider a multi-stage stochastic optimization problem of the form

$$\max\{\mathcal{A}[H(x_0, Y_1, \ldots, x_{T-1}, Y_T)] : x \lhd \boldsymbol{\mathcal{F}}\},$$

where $Y = (Y_1, \ldots, Y_T)$ is a random scenario process,
$x = (x_0, \ldots, x_{T-1})$ is the sequence of decisions,
$H(x_0, Y_1, \ldots, x_{T-1}, Y_T)$ is the profit function and
$\mathcal{A}$ is a version-independent acceptability functional.

Recall that a version independent $\mathcal{A}(\cdot|\mathcal{F}_1)$ is called compound convex , if for all $Y \in dom\mathcal{A}$

$$\mathcal{A}(Y) \leq \mathbb{E}\left[\mathcal{A}(Y|\mathcal{F}_1)\right].$$

Theorem. Suppose that the probability functional $\mathcal{A}$ is compound convex. Then the solution of the above problem can always be chosen as nonrandom, i.e. randomized decisions cannot outperform nonrandomized ones.

Distances between nested distributions

It is equivalent to speak about nested distributions or about equivalence classes of minimal process-and-information pairs. Since a nested distribution is a distribution on the Polish space $\Xi_{1:T}$, the notion of Wasserstein distance makes sense. If $\mathbb{P}$ and $\tilde{\mathbb{P}}$ are two nested distributions on $\Xi_{1:T}$, then the distance $d(\tilde{\mathbb{P}}, \mathbb{P})$ is well defined. This distance makes sense, even if one process is discrete and the other is not.

Proposition. Let $\mathbb{P}$ resp. $\tilde{\mathbb{P}}$ be two nested distributions and let P resp. $\tilde{P}$ their multivariate projections. Then $d(P, \tilde{P}) \le d(\mathbb{P}, \tilde{\mathbb{P}})$.

Proposition. The multi-period $\mathbb{A}$V@R

$$m\mathbb{A}\mathrm{V@R}_\alpha(\xi_1, \ldots, \xi_T; \mathcal{F}_0, \ldots, \mathcal{F}_{T_1}) = \sum_{t=1}^{T} \mathbb{E}[\mathbb{A}\mathrm{V@R}_\alpha(\xi_t | \mathcal{F}_{t-1})]$$

is a version-independent and the mapping $\mathbb{P} \mapsto m\mathbb{A}\mathrm{V@R}_\alpha(\mathbb{P})$ is Lipschitz, more precisely,

$$|m\mathbb{A}\mathrm{V@R}_\alpha(\mathbb{P}) - m\mathbb{A}\mathrm{V@R}_\alpha(\tilde{\mathbb{P}})| \le (1/\alpha) d(\mathbb{P}, \tilde{\mathbb{P}}).$$