Insuring against loss of evidence and capital: review

Vladimir Vovk

Department of Computer Science
Royal Holloway, University of London

Tokyo
12 November, 2012

Co-authors

Joint works with:

A. Philip Dawid, Steven de Rooij, Peter Grünwald, Wouter M. Koolen, Glenn Shafer, Alexander Shen, and Nikolai Vereshchagin

Statistical testing in game-theoretic probability

It is usually framed as a perfect-information game between two players, Forecaster and Sceptic. On each round:

- Forecaster sets prices for various gambles;
- Sceptic chooses which gambles to make.

If Sceptic multiplies by a large factor the capital he puts at risk, he has evidence against Forecaster's ability; his capital at the end of each round is a measure of his evidence against Forecaster so far. However, his capital can go up and then back down. This is counterintuitive: evidence is supposed to be monotonic; evidence that can disappear is not really evidence.

This talk 1

We will see that evidence found by Sceptic against Forecaster is monotonic, at least to some degree. Yes, reporting the maximum so far instead of the current value exaggerates the evidence against Forecaster. However, the exaggeration is not gross: Sceptic's strategy can be modified in such a way that actual evidence becomes almost as strong as the exaggerated evidence.

This gives a method for insuring against loss of evidence.

This talk 2

Our results will have two main interpretations:

- Statistical: Suppose you have a lot of evidence against Forecaster, or against a null hypothesis. How can you avoid losing all of it?
- Financial: Suppose your current capital is large. Should you continue trading (risking losing all your money) or should you stop (preventing your capital from growing further)? Can we compromise? What trade-offs are open to us?

The same mathematics.

Outline

Trading protocol

We are trading in one security X in a financial market. Normalize the initial price X_0 to 1 and the investor's initial capital $\mathcal{K}_0$ to 1. Trading protocol:

$X_0 := 1$ and $\mathcal{K}_0 := 1$
FOR $t = 1, 2, \dots$:
  Investor announces $p_t \in \mathbb{R}$
  Market announces $X_t \in [0, \infty)$
  $\mathcal{K}_t := \mathcal{K}_{t-1} + p_t(X_t - X_{t-1})$
END FOR

$\mathcal{K}_t$: Investor's capital. Trading strategy is a strategy for Investor in this protocol.

Can be interpreted in terms of evidence.

What we want to achieve

Set

$$X_t^* := \max_{s \le t} X_s.$$

We would like to have a trading strategy that guarantees

$$\mathcal{K}_t \ge F(X_t^*)$$

for all t, where, as $y \to \infty$, $F(y) \to \infty$ almost as fast as y. If this inequality can be guaranteed, F is a simple lookback adjuster (SLA).

We are interested only in nonnegative SLAs. Later in the talk: characterization of SLAs.

Admissibility

The set of SLAs is too big. More manageable subset: admissible SLAs (ASLAs).

An SLA G dominates an SLA F if $G \geq F$.

F is admissible if it is not dominated by any G different from F.

Basic result

Theorem

- *Any SLA is dominated by an ASLA.*
- *A function* $F : [1, \infty) \to [0, \infty)$ *is an ASLA if and only if it is increasing, right-continuous, and satisfies*

$$\int_1^\infty \frac{F(y)}{y^2} dy = 1.$$

$F(y) = y$ is impossible (it would mean guaranteeing $\mathcal{K}_t \geq X_t^*$ for all t), but we want to come as close to this as possible.

Examples

Let $\alpha \in (0, 1)$.

- There exists a trading strategy guaranteeing

$$\mathcal{K}_t \geq \alpha \left(X_t^*\right)^{1-\alpha}$$

  for all t.

- There exists a trading strategy guaranteeing

$$\mathcal{K}_t \geq \alpha(1+\alpha)^\alpha \frac{X_t^*}{\ln^{1+\alpha} X_t^*}$$

  whenever $X_t^* \geq e^{1+\alpha}$.

Idea of the proof, I: part "if"

For every threshold c we consider the trading strategy that buys 1 unit of X at the beginning and sells it when X_t reaches c. This strategy shows that $F(y) = F_c(y) = c\mathbf{1}_{\{y \geq c\}}$ is an SLA. This function satisfies $\int_1^\infty F(y) y^{-2} dy = 1$. Moreover: any "regular" function that satisfies $\int_1^\infty F(y) y^{-2} dy = 1$ can be represented as a mixture of F_c: $F = \int_1^\infty F_c P(dc)$ for some probability measure P on $[1, \infty)$.

Lemma

An increasing right-continuous function $F : [1, \infty) \to [0, \infty)$ *satisfies* $\int_1^\infty F(y) y^{-2} dy = 1$ *if and only if there exists a probability measure* P *on* $[1, \infty)$ *such that* $F(y) = \int_{[1,y]} c P(dc)$ *for all* $y \in [1, \infty)$.

Idea of the proof, II: part “if”

Various values of c: our “experts”. We are trying to compete against the best of them.

A standard approach: mixing the experts (trivial in this setting). An example of mixing: suppose $P(\{1\}) = P(\{2\}) = 1/2$. The mixed strategy invests half of its initial capital into the strategy that sells its share of X when X_t reaches 1 and invests the other half of its initial capital into the strategy that sells its share of X when X_t reaches 2.

Idea of the proof, III: part “only if”

Imagine that time t is continuous (this can always be approximated). If X_t is Brownian motion started from 1 and stopped when it hits 0, the distribution of X^*_∞ has density $1/y^2$ on $[1, \infty)$. (Indeed, the probability that Brownian motion started at 1 hits level $y > 1$, before hitting 0 is $1/y$; therefore, the distribution function of X^*_∞ is $1 - 1/y$, and its density is $1/y^2$.)

Since $\mathcal{K}_t$ is a martingale,

$$\mathcal{K}_t \geq F(X^*_t)$$

implies

$$\int_1^\infty \frac{F(y)}{y^2} dy \leq 1.$$

(Set $t = \infty$ and take expected values.)

Applications to hypothesis testing for the Bernoulli model

P_θ: the probability measure on $\{0,1\}^\infty$ corresponding to tossing a coin with probability of 1 equal to θ.

First we will test the hypothesis that the coin is fair, $\theta = 1/2$, against a simple alternative, $\theta = 3/4$. The likelihood ratio is

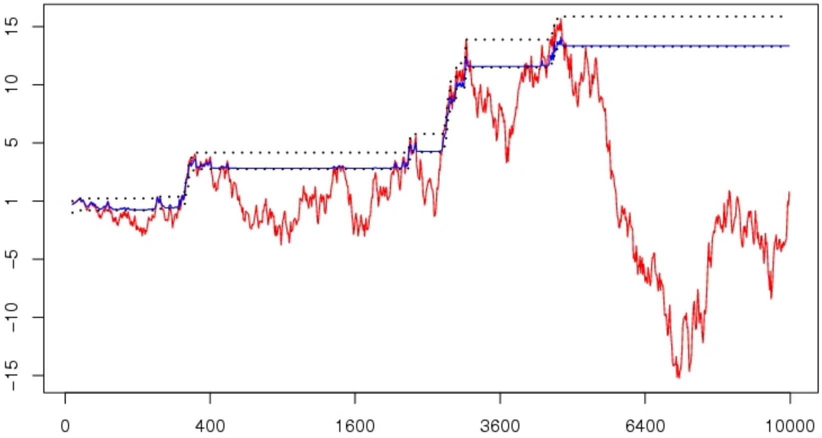
$$X_t := \frac{P_{3/4}(x_1,\dots,x_t)}{P_{1/2}(x_1,\dots,x_t)} = \frac{(3/4)^{k_t}(1/4)^{t-k_t}}{(1/2)^t} = \frac{3^{k_t}}{2^t},$$

where k_t is the number of 1s in $x_1,\dots,x_t$. It is a test martingale (nonnegative and starting from 1).

We will use the SLA

$$F(y) := 0.1y^{0.9}.$$

Testing a simple hypothesis vs a simple alternative



Legend

We generated $x_1, \dots, x_{10,000}$ by tossing a coin with bias $\theta = \ln 2/\ln 3 \approx 0.63$. Vertical axis: log (base 10) scale.

- The red line is traced by X_t (the current evidence against $\theta = 1/2$).
- The upper dotted line is X_t^* (the best evidence so far against $\theta = 1/2$).
- The lower dotted line $F(X_t^*) = 0.1(X_t^*)^{0.9}$ shrinks this best evidence using our SLA (calibrates the exaggerated evidence).
- The blue line is a test martingale $\mathcal{K}_t$ under the null hypothesis that satisfies $\mathcal{K}_t \geq F(X_t^*)$.

Some values

$$X_{10,000} = 2.2 \qquad X^*_{10,000} = 7.3 \times 10^{15}$$
$$F(X^*_{10,000}) = 1.9 \times 10^{13} \qquad \mathcal{K}_{10,000} = 2.2 \times 10^{13}$$

The test martingale (under the null hypothesis) $\mathcal{K}_t$ legitimately and correctly rejects the null hypothesis at time $10,000$ on the basis of X_t's high earlier values, even though the likelihood ratio $X_{10,000}$ is not high.

Testing $\theta = 3/4$ vs $\theta = 1/2$

The test martingale X_t's overwhelming values against $\theta = 1/2$ are followed, around $t = 7,000$, by overwhelming values (order of magnitude 10^{-15}) against $\theta = 3/4$. Had we been testing $\theta = 3/4$ against $\theta = 1/2$, we would have found that it can also be rejected very strongly even after calibration.

The data generating mechanism P_θ, $\theta = \ln 2 / \ln 3$, is "midway" between $P_{1/2}$ and $P_{3/4}$, and as $t \to \infty$, $\ln X_t$ oscillates between approximately $\pm 0.75\sqrt{t \ln \ln t}$.

A very uncomfortable situation for a Bayesian. If, say, his prior probabilities are $50/50$, his posterior probability that $\theta = 1/2$ will fluctuate wildly (from 10^{-15} to $1 - 10^{-15}$ in the picture).

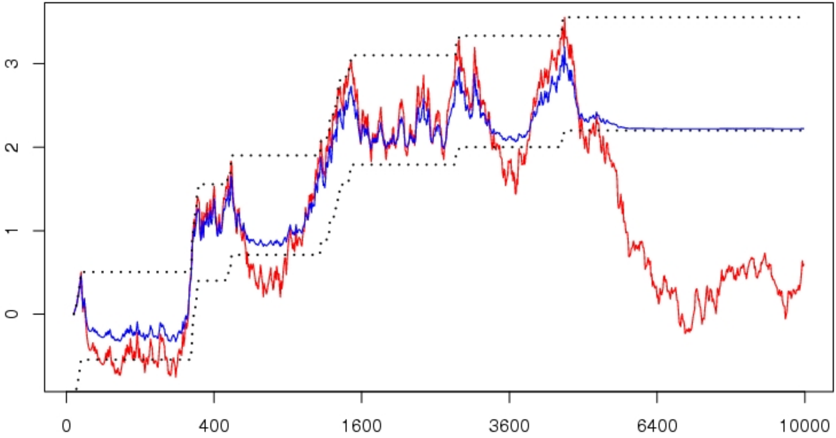
Testing a simple null hypothesis against a "composite" alternative

Retaining $\theta = 1/2$ as our null hypothesis, we now take as our alternative hypothesis the probability distribution Q obtained by averaging P_θ with respect to the uniform distribution for θ. The likelihood ratio:

$$X_t := \frac{Q(x_1, \ldots, x_t)}{P_{1/2}(x_1, \ldots, x_t)} = \frac{\int_0^1 \theta^{k_t} (1-\theta)^{t-k_t} d\theta}{(1/2)^t} = \frac{k_t!(t-k_t)!2^t}{(t+1)!}.$$

Let us generate 0s and 1s independently but with a probability for $x_t = 1$ that slowly converges to $1/2$: $\frac{1}{2} + \frac{1}{4}\sqrt{\ln t / t}$.

Testing a simple hypothesis vs a composite alternative

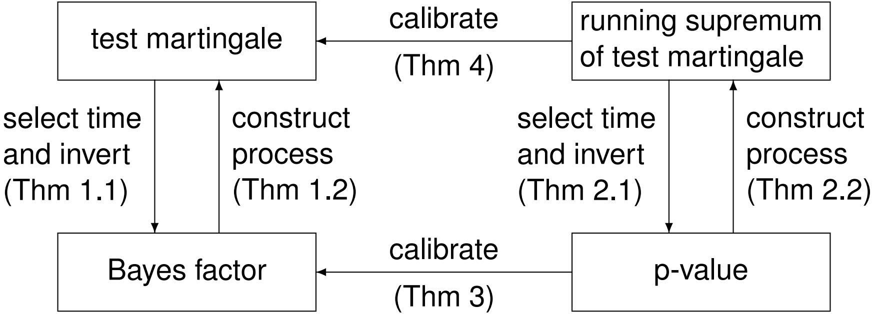


Final values

$$X_{10{,}000} = 3.5 \qquad X^*_{10{,}000} = 3{,}599$$
$$F(X^*_{10{,}000}) = 159 \qquad \mathcal{K}_{10{,}000} = 166$$

In this case, the evidence against $\theta = 1/2$ is very substantial but not overwhelming. (When testing composite vs composite, such as $\theta \le 1/2$ against $\theta > 1/2$, it is hardly possible to find any evidence.)

Advertisement



Details: in the *Statistical Science* paper (reference at the end of the talk).

Connection with p-values from the point of view of game-theoretic probability

There is a direct connection simply by the definition of game-theoretic probability.

Fix a game-theoretic martingale (i.e., Sceptic's capital process when he follows a fixed strategy). Take the highest value reached by the martingale (=Sceptic's capital) as the test statistic. By definition, the p-value will be at most $1/X^*_\infty$.

Outline

1. Simple lookback adjusters
2. Lookback adjusters
   - Definition and characterization of LAs
   - Connections
   - Option pricing: Simple American lookbacks
   - Option pricing: American lookbacks
3. Down-up adjusters
4. References

What we want to achieve in this section

More generally, we can ask when Investor can guarantee

$$\mathcal{K}_t \geq F(X_t^*, X_t), \quad \forall t.$$

Such F: lookback adjusters (LAs). They ensure against loss of evidence in a stronger sense than SLAs.

We are interested only in nonnegative SLAs and LAs F.

Admissibility

The same notion of admissibility as for SLAs.

An LA G dominates an LA F if $G \geq F$.

F is admissible if it is not dominated by any G different from F. "Admissible LA" is abbreviated to "ALA".

Lookback adjusters

Theorem

Every LA is dominated by an ALA. A positive function $F(X^*, X)$ with domain $X^* \in [1, \infty)$ and $X \in [0, X^*]$ is an ALA if and only if the following two conditions are satisfied:

- *the function*

$$F^{=}(X^*) := F(X^*, X^*), \quad X^* \in [1, \infty),$$

  is increasing, concave, and satisfies $F^{=}(1) = 1$ and $F^{=}_{\mathrm{r}}(1) \leq 1$;
- *for each $X^* \in [1, \infty)$, the function $F(X^*, X)$ is linear in X and its slope is equal to $F^{=}_{\mathrm{r}}(X^*)$.*

Notation and terminology

f_{r} means the right derivative of f.

The function $F^{=} : [1, \infty) \to [0, \infty)$ defined by

$$F^{=}(X^*) := F(X^*, X^*), \quad X^* \in [1, \infty),$$

will be called the spine of an ALA $F(X^*, X)$.

Trivial part of the theorem

Lemma

If a positive function $F(X^*, X)$, $X^* \in [1, \infty)$, $X \in [0, X^*]$, satisfies the two conditions in the theorem, it is an LA.

Proof.

The following trading strategy witnesses that F is an LA: at any time t, take the position $p_t := F_{\mathrm{r}}^{=}(X^*_{t-1})$. □

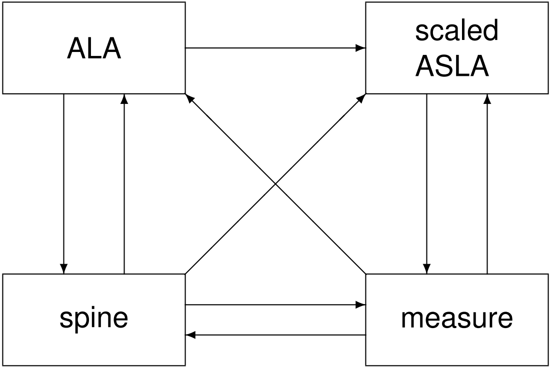
Connections, I

What is the connection between ASLAs and ALAs? Their definitions (concave vs possibly discontinuous with an integral condition) appear to have nothing in common.

The next slide: the scheme of connections between related notions. An arrow means that there is a simple expression of one via another (sometimes two simple expressions).

Scaled ASLA (more fully, scaled down ASLA) is an ASLA multiplied by $c \in [0, 1]$.

Visual frame: 3 bijections and 3 shortcuts



Connections, II: scaled ASLAs from measures

We already know how to get an ASLA from a probability measure P on $[0,\infty)$:

$$F(y) := \int_{[1,y]} uP(du).$$

But in our picture, "measure" means a probability measure on $[0,\infty]$, so the formula gives only a scaled ASLA (scaled down by $P([0,\infty)))$.

Why does it make sense to consider probability measures on $[0,\infty]$ (leading to inadmissible ASLAs)? They give

$$\mathcal{K}_t \geq F(X_t^*) + cX_t,$$

where $c := P(\{\infty\})$. Double insurance: against performing much worse than X_t (strong insurance) and against performing much worse than X_t^* (weaker insurance).

Connections, III: ALAs from spines

Remember: $F^{=}(X) := F(X, X)$.

A spine = an ALA without the linear bits (informationally redundant). The full ALA can be restored by:

$$F(X^*, X) = F^{=}(X^*) + F^{=}_{\mathrm{r}}(X^*)(X - X^*).$$

Connections, IV: spines and measures

Given a spine $F^{=}$, the corresponding measure is determined by

$$P((X,\infty]) = F^{=}_{\mathrm{r}}(X), \quad X \in [1,\infty).$$

The spine can be restored by

$$F^{=}(X) = \int_{[1,X]} uP(du) + XP((X,\infty])$$

(the capital of the experts who cashed in their share plus the capital of the experts still holding it). More generally, the corresponding ALA is

$$F(X^*,X) = \int_{[1,X^*]} uP(du) + XP((X^*,\infty]).$$

Connections, V: ALAs and scaled ASLAs

A function $F'(X^*)$ is a scaled ASLA if and only if it can be represented in the form $F(X^*, 0)$ for some ALA F. Details:

Proposition

- *Suppose* $F(X^*, X)$ *is an ALA. Then* $F'(X^*) := F(X^*, 0)$ *is a scaled ASLA. If, furthermore,* $F^{=}_{\mathrm{r}}(\infty) = 0$, $F(X^*)$ *is an ASLA.*
- *Suppose* $F'(X^*)$ *is a scaled ASLA. Then there exists a unique ALA* $F(X^*, X)$ *such that* $F'(X^*) = F(X^*, 0)$ *for all* X^**. If, furthermore,* $F'(X^*)$ *is an ASLA, this ALA* $F(X^*, X)$ *will satisfy* $F^{=}_{\mathrm{r}}(\infty) = 0$.

How to use this kind of results for option pricing

We would like to price the following exotic option O_G (a kind of perpetual American lookback): at the time t of her choice, the option's owner is entitled to $G(X_t^*)$, where G is a given nonnegative increasing function. The result about ASLAs can be restated as: the upper price of this option at time 0 (after learning X_0) is $\int_1^\infty G(y)y^{-2}dy$.

Formally, $\int_1^\infty G(y)y^{-2}dy$ is the smallest initial capital c such that there exists a trading strategy starting with c and guaranteeing $\mathcal{K}_t \geq G(X_t^*)$ for all t (intuitively, the seller can always meet his obligation).

Risk-neutral probability, I

The formula $\int_1^\infty G(y)y^{-2}dy$ assumes $X_0 = 1$. Without this assumption, the upper price at time 0 is $\int_1^\infty G(X_0 y)y^{-2}dy$ (by re-scaling).

Notice: the upper price of O_G can be written as $X_0 \int_{X_0}^\infty G(y)y^{-2}dy$: the expected value of G with respect to the probability measure P on $[X_0, \infty)$ with density X_0/y^2. It plays the role of risk-neutral probability (but, unusually, emerges in a heavily incomplete market).

Risk-neutral probability, II

The basic interpretation of risk-neutral probability: P is the distribution of the maximum of Brownian motion started at X_0 and stopped when it hits 0.

But the argument for the density X_0/y^2 of X^*_∞ depends only on X_t being a martingale without awkward (i.e., upward) jumps that converges to $X_\infty \notin (0, \infty)$. In particular, we can assume the standard "Black–Scholes model"

$$\frac{dX_t}{X_t} = \sigma dW_t,$$

where W_t is Brownian motion. The worst-case upper price = the expected value w.r. to the most standard model.

Pricing O_G at time t

This is easy: replace X_0 by X_t and $G(y)$ by $G(X_t^* \vee y)$.

So the upper price of O_G at time t is

$$X_t \int_{X_t}^{\infty} G\left(X_t^* \vee y\right) y^{-2} dy.$$

General remark: we will get the same expressions for pricing European options with the given payoff at the time ∞.

American lookbacks, I

Suppose the payoff of our American lookback is $F(X_t^*, X_t)$ (allowed to depend on X_t).

For each concave increasing $G : [X_0, \infty) \to \mathbb{R}$, define

$$\overline{G}(X^*, X) := G(X^*) + G_{\mathrm{r}}(X^*)(X - X^*), \qquad X^* \in [X_0, \infty), \quad X \in [0, X^*].$$

American lookbacks, II

The upper price $\overline{\mathbb{E}}(F \mid X_0)$ of the American option paying $F(X_t^*, X_t)$ can be determined in two steps:

- Let $H : [X_0, \infty) \to [0, \infty)$ be the smallest concave increasing function such that $\overline{H} \geq F$. Lemma: It exists (but maybe $H := \infty$).
- The function H determines $\overline{\mathbb{E}}(F \mid X_0)$ via

$$\overline{\mathbb{E}}(F \mid X_0) = H(X_0).$$

Given the initial capital $H(X_0)$, the option's seller can meet his obligation by holding $p_t := H_{\mathrm{r}}(X_{t-1}^*)$ units of X at time t.

Outline

Buy low, sell high 1

The adjusters we considered so far: “sell high”. Now we want to implement both parts of the traders’ adage “buy low, sell high”.

If we know in advance an interval (a, b) $(0 \le a \le 1 \le b)$ such that the price upcrosses (a, b) (there are $s < t$ such that $X_s \le a$ and $b \le X_t$), we can increase our initial capital b/a-fold. In practice we do not know which upcrossings are going to occur…

Buy low, sell high 2

We are looking for strategies whose payoff scales nicely with the extremity of the upcrossing present. Suppose a financial advisor claims to have achieved this with a secret strategy and publishes a function G ($G(a, b) \approx b/a$ in some sense) promising that her strategy will

> *keep our capital above $G(a, b)$ for each upcrossing (a, b).*

We would like to answer the following questions:

1. Should we believe her? Is it actually possible to guarantee G? (If so, we call G a down-up adjuster.)
2. Is she ambitious enough? Or can one guarantee strictly more than G? (Is G admissible?)

New phenomenon

In the case of “up adjusters” (SLAs), we could just combine the basic guarantees $F(y) := c\,\mathbf{1}_{\{y \ge c\}}$ achievable by trading strategies of selling when the price reaches a given threshold. Now combinations of the basic guarantees (exploiting a given upcrossing (a, b)) are typically not admissible! For example, consider the mixture

$$G(a,b) := \mathbf{1}_{\{a \le 1 \text{ and } b \ge 2\}} + \mathbf{1}_{\{a \le \frac{1}{2} \text{ and } b \ge 1\}}.$$

Direct calculation shows: this can be guaranteed from the initial capital $\frac{11}{12} < 1$. So G in not achievable: it is dominated by the adjuster $\frac{12}{11} G$.

Why?

Consider the price path $(1, 2, \frac{1}{2}, 1, 0, \ldots)$. Starting from 1 our mixture strategy achieves 2 whereas 1 would have been sufficient ($[\frac{1}{2}, 2]$ is never upcrossed).

This suggests that the strategy wastes part of the initial capital.

Characterization of down-up adjusters

A function $G : (0,1] \times (0,\infty) \to [0,\infty)$ a candidate guarantee if it is upper semi-continuous, decreasing in its first argument, and increasing in its second argument. The second-argument upper inverse of G is

$$G^{-1}(a,h) := \inf\{b \geq a : G(a,b) \geq h\}.$$

Theorem (characterisation)

A candidate guarantee G is an adjuster if and only if

$$\int_0^\infty 1 - \exp\left(\int_0^1 \frac{da}{a - G^{-1}(a,h)}\right) dh \leq 1.$$

Moreover, G is admissible if and only if this holds with equality.

Example adjusters 1

- Considering $G(a, b)$ that do not depend on a we recover the results about SLAs.
- Candidate guarantees of the form $G(a, b) = F(b - a)$ have upper price

$$\int_0^\infty F(y) \frac{e^{-1/y}}{y^2} dy.$$

The distribution with density $\frac{e^{-1/y}}{y^2} dy$ is the risk-neutral measure of the largest upcrossed length. This is analogous to $y^{-2} dy$ being the risk-neutral measure of the maximum price.

Example adjusters 2

- Guarantees of the form $G(a, b) = F(b/a)$ for increasing and unbounded F have infinite upper price.
- An admissible down-up adjuster that approaches the ideal $G(a, b) \approx b/a$ is

$$G(a, b) := \frac{(b-a)^p}{a^q} \frac{\left(\frac{p-q}{p}\right)^p}{\Gamma(1-p)},$$

for any $0 \leq q < p < 1$.

Outline

Proofs and related results, I

Statistical interpretation:

Glenn Shafer, Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk.
Test martingales, Bayes factors and p-values.
Statistical Science **26**, 84–101 (2011).

A. Philip Dawid, Steven de Rooij, Glenn Shafer, Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk.
Insuring against loss of evidence in game-theoretic probability.
Statistics and Probability Letters **81**, 157–162 (2011).

Proofs and related results, II

Financial interpretation:

A. Philip Dawid, Steven de Rooij, Peter Grünwald, Wouter Koolen, Glenn Shafer, Alexander Shen, Nikolai Vereshchagin, and Vladimir Vovk.
Probability-free pricing of adjusted American lookbacks.
arXiv:1108.4113 [q-fin.PR], August 2011.

This is the most detailed exposition.

Proofs and related results, III

This paper shows that a continuous price path of a financial security behaves (for the purpose of option pricing) like a time-changed Brownian motion (and so explains the appearance of the density $1/y^2$):

Vladimir Vovk.
Continuous-time trading and the emergence of probability.
Finance and Stochastics,
16, 561–609 (2012).

All these papers are available from `http://probabilityandfinance.com` (Working Papers 28, 33, 34, and 37).

Proofs and related results, IV

Down-up adjusters:

Wouter M. Koolen and Vladimir Vovk.
Buy low, sell high.
In: *Proceedings of the Twenty Third International Conference on Algorithmic Learning Theory* (ed. by N. Bshouty, G. Stoltz, N. Vayatis and T. Zeugmann). *Lecture Notes in Artificial Intelligence*. Berlin: Springer, 2012.

Thank you for your attention!