Blackwell Prediction

Hans Rudolf Lerche

Department of Mathematical Stochastics

University of Freiburg

SOS, Petrozavodsk

September 13, 2010

Classical Blackwell Prediction

Let $x_1, x_2, x_3, \ldots$ be a infinite 0-1 sequence, not necessarily stationary or even random.

We wish to sequentially predict the sequence:

$$\text{Guess } x_{n+1}, \text{ knowing } x_1, x_2, \ldots, x_n.$$

Of interest are algorithms which predict well for **all** 0-1 sequences.

One of them is the Blackwell algorithm.

A prediction algorithm $y_1, y_2, y_3, \ldots$ is a random 0-1 sequence with y_{n+1} being the predicted value of x_{n+1}. y_{n+1} may depend on $x_1, x_2, x_3, \ldots, x_n$ and on some other random variables.

Some further notation:

$$\overline{x}_n = \frac{1}{n}\sum_{i=1}^{n} x_i,$$ the relative frequency of "1" in the sequence $x_1, x_2, x_3, \ldots, x_n$,

$$\gamma_i = \mathbb{1}_{\{y_i = x_i\}},$$ the success indicator,

$$\overline{\gamma}_n = \frac{1}{n}\sum_{i=1}^{n} \gamma_i,$$ the relative frequency of correct prediction.

A plausible deterministic prediction scheme:

$$y_{n+1}^0 = \begin{cases} 1 & \text{if } \overline{x}_n > \frac{1}{2} \\ 0 & \text{if } \overline{x}_n \le \frac{1}{2} \end{cases} \quad \text{for } n \ge 1,$$

$$y_1^0 = 1.$$

Its strength: Let $0 \le p \le 1$.

If $x_1, x_2, x_3, \ldots$ are independent Bernoulli (p), then for $(y_n^0; n \ge 1)$

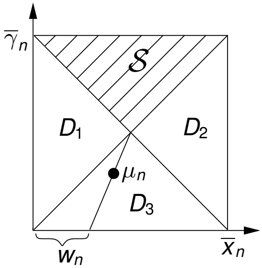
$$\overline{\gamma}_n \to \max(p, 1-p) \text{ for } n \to \infty$$

by the law of large numbers. If p is known beforehand, one cannot do better asymptotically.

Its Weakness: For $1, 0, 1, 0, 1, 0, \ldots \quad \overline{\gamma}_n = \frac{1}{n}$ for all $n \ge 1$.

Blackwell algorithm: Let $\mu_n = (\overline{x}_n, \overline{\gamma}_n) \in [0,1]^2$ and

$\mathcal{S} = \{(x, y) \in [0,1]^2 \mid y \geq \max(x, 1-x)\}$.



y_{n+1} is chosen on the basis of μ_n according to the conditional probabilities

$$P(y_{n+1} = 1) = \begin{cases} 0 & \text{if } \mu_n \in D_1 \\ 1 & \text{if } \mu_n \in D_2 \\ w_n & \text{if } \mu_n \in D_3. \end{cases}$$

When μ_n is in the interior of $\mathcal{S}$, y_{n+1} can be chosen arbitrarily. Let $y_1 = 0$.

d denotes the Euclidean distance in $\mathbb{R}^2$ and $d(x, A)$ the distance from point x to the set A.

Theorem 1

For the Blackwell-algorithm applied to any infinite 0-1 sequence $x_1, x_2, x_3, \ldots$ the sequence $(\mu_n; n \geq 1)$ converges almost surely to $\mathcal{S}$, i.e. $d(\mu_n, \mathcal{S}) \to 0$ as $n \to \infty$ almost surely.

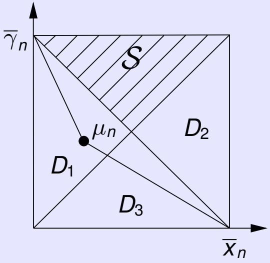
Remark

The theorem has minimax character. For every 0-1 sequence the Blackwell-algorithm is at least as successful as for *iid* Bernoulli-variables. But for those it does the best possible.

Proof

Let $d_n = d(\mu_n, \mathcal{S})$.

Case 1: $\mu_n \in D_1$



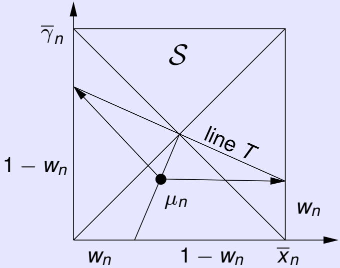
Then $d_{n+1} = \frac{n}{n+1} d_n$.

Case 2: $\mu_n \in D_2$ Then $d_n = \frac{n}{n+1} d_n$.

Case 3: $\mu_n \in D_3$

We have $\mu_{n+1} = \frac{n}{n+1}\mu_n + \frac{1}{n+1}(x_{n+1}, \gamma_{n+1})$ and

$$E(\gamma_{n+1} \mid x_{n+1} \text{ and past until } n) = \begin{cases} 1 - w_n & \text{if } x_{n+1} = 0 \\ w_n & \text{if } x_{n+1} = 1. \end{cases}$$



The conditional expectation of μ_{n+1} is closer to T than μ_n.

It holds $\quad (*) \quad E\left(d_{n+1}^2 \mid \text{past}(n)\right) \le \left(\frac{n}{n+1}\right)^2 d_n^2 + \frac{1}{2(n+1)^2} \quad$ for $\mu_n \in D_3$

and with $d_n = d(\mu_n, \mathcal{S})$. We have $\quad \mu_{n+1} = \frac{n}{n+1}\mu_n + \frac{1}{n+1}(x_{n+1}, \gamma_n)$.

$$\begin{aligned} d_{n+1}^2 &= d\left(\mu_{n+1}, \mathcal{S}\right)^2 \le \left\|\mu_{n+1} - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\|^2 \\ &= \left\|\tfrac{n}{n+1}\left(\mu_n - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right) + \tfrac{1}{n+1}\left[(x_{n+1}, \gamma_{n+1}) - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right]\right\|^2 \\ &= \left(\tfrac{n}{n+1}\right)^2 d_n^2 + \tfrac{1}{2(n+1)^2} + \tfrac{2n}{(n+1)^2}\left\langle \mu_n - \left(\tfrac{1}{2}, \tfrac{1}{2}\right), (x_{n+1}, \gamma_{n+1}) - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\rangle \end{aligned}$$

It holds $\quad (*) \quad E\left(d_{n+1}^2 \mid \text{past}(n)\right) \leq \left(\frac{n}{n+1}\right)^2 d_n^2 + \frac{1}{2(n+1)^2} \quad$ for $\mu_n \in D_3$

and with $d_n = d(\mu_n,\ \mathcal{S})$. We have $\quad \mu_{n+1} = \frac{n}{n+1}\mu_n + \frac{1}{n+1}(x_{n+1}, \gamma_n)$.

$$\begin{aligned} d_{n+1}^2 &= d\left(\mu_{n+1}, \mathcal{S}\right)^2 \leq \left\|\mu_{n+1} - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\|^2 \\ &= \left\|\tfrac{n}{n+1}\left(\mu_n - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right) + \tfrac{1}{n+1}\left[(x_{n+1}, \gamma_{n+1}) - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right]\right\|^2 \\ &= \left(\tfrac{n}{n+1}\right)^2 d_n^2 + \tfrac{1}{2(n+1)^2} + \tfrac{2n}{(n+1)^2}\left\langle \mu_n - \left(\tfrac{1}{2}, \tfrac{1}{2}\right), (x_{n+1}, \gamma_{n+1}) - \left(\tfrac{1}{2}, \tfrac{1}{2}\right)\right\rangle \end{aligned}$$

Taking conditional expectation $E(.\mid x_{n+1},\ \text{past}(n))$ the bracket-term vanishes because of the orthogonality of T and $\mu_n - (\frac{1}{2}, \frac{1}{2})$ and we get $(*)$.

But $(*)$ holds also for D_1, D_2 and $\mathcal{S}$.

Thus $(d_n^2; n \geq 1)$ is a nonnegative almost supermartingale with $E(d_n^2) \leq \frac{1}{2n}$.

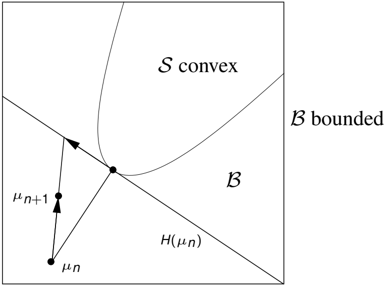
Then $Z_n = d_n^2 + \sum_{i \geq n} \frac{1}{2(i+1)^2}$ is a positive supermartingale with $EZ_n \to 0$.

The convergence theorem for supermartingales implies Theorem 1. □

Remark

For general weights see the result of F. Riedel (2008).

Approachability of a convex set by an arithmetic mean



Let $\mu_n = \frac{1}{n}\sum_{i=1}^{n} z_i; \quad z_i \in \mathcal{B}.$

If one can choose z_{n+1} such that it lies on $H(\mu_n)$, for all $n \geq 1$, then

$d(\mu_n, \mathcal{S}) \to 0.$

Sequential prediction of $d \geq 3$ categories

Let $x_1, x_2, x_3, \ldots$ be a infinite sequence with outcomes in $D = \{0, 1, \ldots, d-1\}$, not necessarily random.

$y_1, y_2, y_3, \ldots$ a sequence of predictors.

$$\overline{x}_n^{(j)} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{x_i = j\}}, \quad j \in \{0, 1, \ldots, d-1\},$$

and

$$\overline{x}_n = \left(\overline{x}_n^{(0)}, \overline{x}_n^{(1)}, \overline{x}_n^{(2)}, \ldots, \overline{x}_n^{(d-1)} \right)$$

the relative frequencies of the selected categories up to n.

$$\overline{\gamma}_n = \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{y_i = x_i\}}$$

the relative frequency of correct predictions.

$$\sum_{d-1} = \left\{ \left(q^{(0)}, q^{(1)}, q^{(2)}, \ldots, q^{(d-1)} \right) \,\middle|\, q^{(j)} \geq 0, \ \sum_{j=0}^{d-1} q^{(j)} = 1 \right\}$$

denotes the $d-1$ dimensional unit simplex in $\mathbb{R}^d$.

Question: Is there an algorithm such that

$$\overline{\gamma}_n \to \max \left(q^{(0)}, q^{(1)}, q^{(2)}, \ldots, q^{(d-1)} \right)$$

for every sequence $x_1, x_2, x_3, \ldots$ with values in D?

A plausible deterministic prediction scheme:

$$y_{n+1}^{0} = \min\left\{k \in D \,\middle|\, \overline{x}_n^{(k)} = \max\left(\overline{x}_n^{(0)}, \overline{x}_n^{(1)}, \overline{x}_n^{(2)}, \ldots, \overline{x}_n^{(d-1)}\right)\right\}$$
$$y_1^{(0)} = d - 1$$

Its strength: Let $q \in \sum_{d-1}$.

If $x_1, x_2, x_3, \ldots$ are independent multinomial (q), then for $(y_n^0; n \geq 1)$

$$\overline{\gamma}_n \to \max\left(q^{(0)}, q^{(1)}, q^{(2)}, \ldots, q^{(d-1)}\right)$$

by the law of large numbers.

If q is known beforehand, one cannot do better asymptotically.

Its weakness: For the sequence

$$d-1, d-2, \ldots, 1, 0, d-1, d-2, \ldots, 1, 0, d-1, \ldots$$

we have $\overline{\gamma}_n = \frac{1}{n}$ for all $n \geq 1$.

Open Problem: Let Σ_{d-1} denote the unit simplex in $\mathbb{R}^d$, let

$$W_d = \Sigma_{d-1} \times [0,1]$$

and

$$\mathscr{S} = \left\{ (q,\gamma) \in W_d \mid \gamma \geq \max\left(q^{(0)}, q^{(1)}, q^{(2)}, \ldots, q^{(d-1)}\right) \right\}.$$

Does there exist a generalized Blackwell algorithm such that for every sequence $x_1, x_2, x_3, \ldots$ with values in $D = \{0, 1, \ldots, d-1\}$, it holds

$$(\overline{x}_n, \overline{\gamma}_n) \to \mathscr{S} \ ?$$

But: The argument of Theorem 1 does not carry over directly since there are no right angles in $\mathscr{S}$ (Condition (C1) of Theorem 2 below is not satisfied.).

UNIVERSITY OF CALIFORNIA, BERKELEY

BERKELEY • DAVIS • IRVINE • LOS ANGELES • RIVERSIDE • SAN DIEGO • SAN FRANCISCO • SANTA BARBARA • SANTA CRUZ

TELEPHONE: (415) 642-2781
CABLE: UNIVCALB

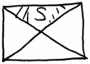
DEPARTMENT OF STATISTICS
BERKELEY, CALIFORNIA 94720

May 23, 1991

Dear Rudi,

My prediction result is Theorem 1 of the enclosed paper, applied to

$$S = \begin{cases} 0 \le x, y \le 1 \\ y \ge \max(x, 1-x) \end{cases}$$:



and

$$M = \begin{vmatrix} (0,1) & (1,0) \\ (0,0) & (1,1) \end{vmatrix}.$$

Player I picks a row of M (his prediction, 0 or 1) and Player II picks a column (what occurs, 0 or 1) and The first coordinate tells what occurs, and the second coordinate tells whether I is correct. Player I, in a series of plays, is trying to force the average point into S.

I don't know whether Theorem 1, applied in the case $k > 2$, applies to $S = v \ge \max(u_1, \dots, u_k)$ or not. Let me know what you find out.

Best regards,
David

Blackwell's Minimax Theorem in Game Theory

Two players! $M = (m_{ij})$ an $r \times s$ payoff matrix with $m_{ij} \in \mathbb{R}^d$.

$$\mathcal{P} = \left\{ p = (p_1, \ldots, p_r) \;\middle|\; p_i \geq 0, \sum_i p_i = 1 \right\}$$

the *mixed actions* of player I,

$$Q = \left\{ q = (q_1, \ldots, q_s) \;\middle|\; q_j \geq 0, \sum_j q_j = 1 \right\}$$

the *mixed actions* of player II. A *strategy* f in the repeated game of player I is a sequence $f = (f_k; k \geq 1)$ with $f_k \in \mathcal{P}$. A *strategy* g for player II is defined similarly.

Two strategies define a sequence of payoffs z_k, $k = 1, 2, \ldots$, i.e.:

In the k-th game i and j are chosen according to f_k and g_k. The payment to player I is then $m_{ij} \in \mathbb{R}^d$. f_k and g_k may depend on earlier outcomes.

Blackwell's question:

Can one control

$$\overline{z}_n = \frac{1}{n} \sum_{k=1}^{n} z_k,$$

where z_k denotes the payoff of the k-th game, with a strategy f such that $\overline{z}_n$ approaches a given set $\mathcal{S}$ independently of what player II does?

Approachability of a set in $\mathbb{R}^d$:

$\mathcal{S} \subset \mathbb{R}^d$ is approachable for player I if there exists a strategy f^* for which $d(\bar{z}_n, \mathcal{S}) \to 0$ with probability 1.

$$\text{For } p \in \mathcal{P} \text{ let } \mathcal{R}(p) = \text{conv}\left(\sum_{i=1}^{r} p_i\, m_{ij},\ j = 1, \ldots, s\right).$$

Theorem 2 (Blackwell's Approachability Result (1956))

$\mathcal{S}$ a closed convex subset of $\mathbb{R}^d$. For every $z \notin \mathcal{S}$ let y denote the closest point in $\mathcal{S}$ to z.

(C1) *For every $z \notin \mathcal{S}$ there exists a $p(z) \in \mathcal{P}$ such that the hyperplane through y which is perpendicular to the line segment $\overline{zy}$ contains $\mathcal{R}(p)$.*

If (C1) holds, then $\mathcal{S}$ is approachable.

Proof

Let $d_n = d(\mu_n, \mathcal{S})$. Let y_n denote the closest point in $\mathcal{S}$ to μ_n. Then

$$
\begin{aligned}
d_{n+1}^2 &\leq \|\mu_{n+1} - y_n\|^2 \\
&= \left\| \tfrac{n}{n+1}(\mu_n - y_n) + \tfrac{1}{n+1}(z_{n+1} - y_n) \right\|^2 \\
&= \left(\tfrac{n}{n+1}\right)^2 d_n^2 + \tfrac{K^2}{(n+1)^2} + \tfrac{2n}{(n+1)^2} \langle \mu_n - y_n, z_{n+1} - y_n \rangle .
\end{aligned}
$$

Taking conditional expectations and using orthogonality by (C1) leads to

$$
E\left(d_{n+1}^2 \;\middle|\; \text{past}(n)\right) \leq \left(\frac{n}{n+1}\right)^2 d_n^2 + \frac{K^2}{(n+1)^2}.
$$

Then further as above.

Examples:

1) **Special case:**

Let $m_{ij} \in \mathbb{R}$. Then there exists a $v \in \mathbb{R}$ such that

$$v = \min_{q \in Q} \max_{p \in \mathcal{P}} p' M q = \max_{p \in \mathcal{P}} \min_{q \in Q} p' M q.$$

(Minimax Theorem of von Neumann)

Then one can find a strategy f^* such that in a sequence of independent games $\bar{z}_n \to \mathcal{S}$, where $\mathcal{S} = [v, \infty)$.

This strategy is given by $f^* = (p^*, p^*, \ldots)$, where p^* is such that

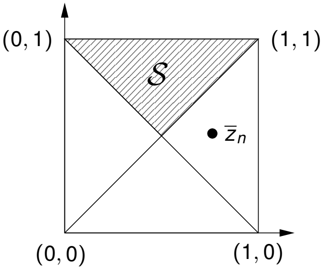
$$\sum_{i=1}^{r} p_i^{*'} m_{ij} \geq v \quad \text{for } j = 1, \ldots, s$$

holds.

2) Let

$$M = \begin{pmatrix} (0,1) & (1,0) \\ (0,0) & (1,1) \end{pmatrix}.$$

Blackwell's Interpretation: Nature chooses the column, corresponding to 0 or 1. The statistician has to predict. He chooses the row. The first component states what nature chooses, the second component whether the statistician is correct or not.



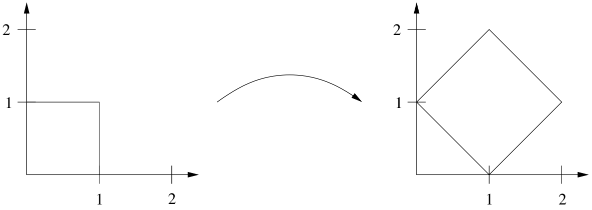
$$\bar{z}_n = (\bar{x}_n, \bar{\gamma}_n)$$

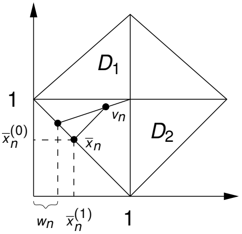
$$\bar{x}_n = \frac{1}{n}\sum_{k=1}^{n} x_k,$$ the relative frequency of "1" in the sequence $x_1, x_2, x_3, \ldots, x_n$,

$$\gamma_k = \mathbb{1}_{\{y_k = x_k\}},$$ the success indicator,

$$\bar{\gamma}_n = \frac{1}{n}\sum_{k=1}^{n} \gamma_k,$$ the relative frequency of correct prediction.

A transformation of the prediction square





$$\overline{x_n} = \left(\overline{x}_n^{(0)}, \overline{x}_n^{(1)}\right)$$

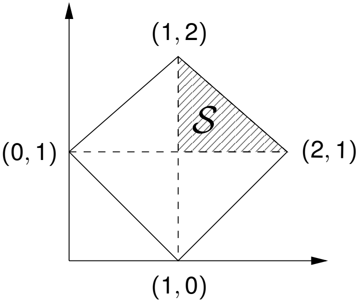
with $\overline{x}_n^{(j)} = \frac{1}{n}\sum_{i=1}^{n} \mathbb{1}_{\{x_i=j\}}, \; j = 0, 1$

and $v_n = \overline{x}_n + \overline{\gamma}_n \, \mathbb{1}_2$.

A basis for generalisations to more than two categories.

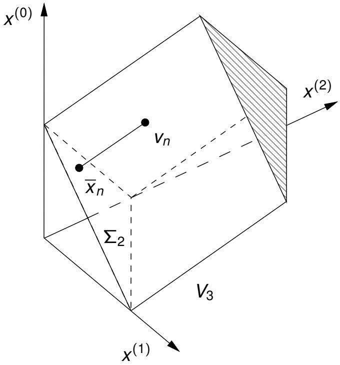
The corresponding matrix

$$M = \begin{pmatrix} (1,2) & (1,0) \\ (0,1) & (2,1) \end{pmatrix}$$



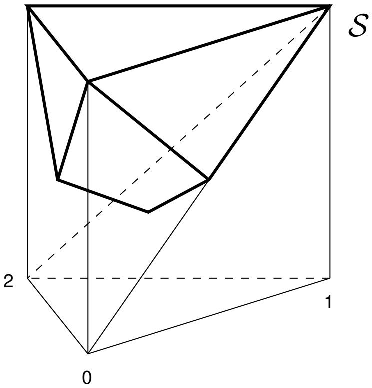
Prediction for $d = 3$

A natural prediction space for three categories. Instead of considering $(\overline{x_n}, \overline{\gamma}_n)$ as pair, we use $v_n = \overline{x}_n + \overline{\gamma}_n \mathbb{1}_3$ with $\mathbb{1}_3 = (1, 1, 1)$.



$$V_3 = \{x + \gamma \mathbb{1}_3 \mid x \in \Sigma_2,\ \gamma \in [0, 1]\}$$

Note $V_3 \subset \mathbb{R}^3$.



$$\mathcal{S} = \left\{ x + \gamma \mathbb{1}_3 \in V \mid \gamma \geq \max \left\{ x^{(0)}, x^{(1)}, x^{(2)} \right\} \right\}$$

Does $v_n \to \mathcal{S}$?

The relation between V_3 and W_3

$$W_3 = \{(x, \gamma) | x \in \Sigma_2,\ \gamma \in [0,1]\} \subset \mathbb{R}^4$$

$$V_3 = \{(x + \gamma \cdot \mathbb{1}_3) | x \in \Sigma_2,\ \gamma \in [0,1]\} \subset \mathbb{R}^3$$

W_3 and V_3 are isometric isomorph, i.e. there exists a $\varphi : W_3 \to V_3$ bijective such that for $z, \tilde{z} \in W_3$ it holds:

$$d_1(z, \tilde{z}) = d(\varphi(z), \varphi(\tilde{z})), \quad \text{where } d_1^2(z, \tilde{z}) = \sum_{i=1}^{3} (z_i - \tilde{z}_i)^2 + 3(z_4 - \tilde{z}_4)^2.$$

Since $\varphi(\mathscr{S}) = \mathcal{S}$, convergence of $(\overline{x}_n, \overline{\gamma}_n) \to \mathscr{S}$ is equivalent to that of $v_n \to \mathcal{S}$, where $v_n = \overline{x}_n + \overline{\gamma}_n \cdot \mathbb{1}_3$.

The geometric structure of V_3

We cut V_3 from each of its upper vertices down to the two lower vertices.

This yields 8 pieces of 4 different types. $\mathcal{S}$ is the piece on the top.

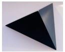


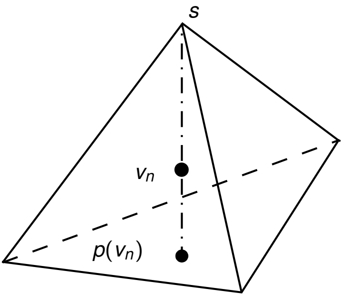


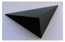
The cutting planes have $s = (\frac{2}{3}, \frac{2}{3}, \frac{2}{3})$ as joint point and are perpendicular to each other.

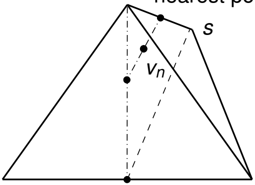
How does the algorithm randomize in the different pieces?

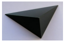
We put $P(Y_{n+1} = j) = p^{(j)}(v_n)$ for $j = 0, 1, 2$ and define $p(v_n) := (p^{(0)}(v_n), p^{(1)}(v_n), p^{(2)}(v_n))$ as follows:

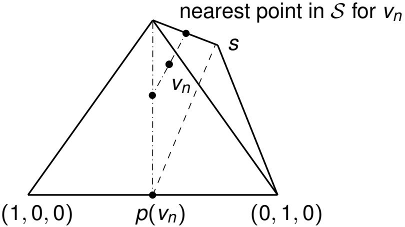
Type 1: $v_n \in$ 

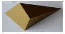


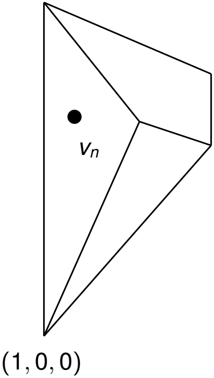
Type 2: $v_n \in$ 

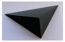


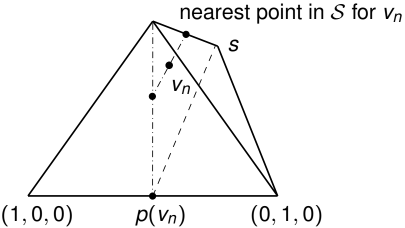
Type 2: $v_n \in$ 

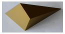


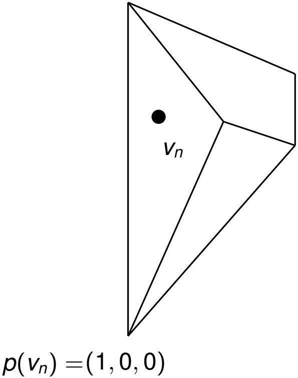
Type 3: $v_n \in$ 



Type 2: $v_n \in$ 



Type 3: $v_n \in$ 



The corresponding matrix

Let

$$M = \begin{pmatrix} (2,1,1) & (0,1,0) & (0,0,1) \\ (1,0,0) & (1,2,1) & (0,0,1) \\ (1,0,0) & (0,1,0) & (1,1,2) \end{pmatrix}$$

Prediction for three categories with payments e_i, $i = 0, 1, 2$ (the unit vectors) or $e_i + \mathbb{1}_3$, $i = 0, 1, 2$.

$$\begin{aligned} V_3 &= \operatorname{conv}\{e_i,\ e_i + \mathbb{1}_3;\ i = 0, 1, 2\}. \\ \mathcal{S} &= \left\{ x + \gamma \mathbb{1}_3 \in V_3 \,\middle|\, \gamma \geq \max\left\{x^{(0)}, x^{(1)}, x^{(2)}\right\} \right\} \\ \overline{z}_n &= \overline{x}_n + \overline{\gamma}_n \mathbb{1}_3 \end{aligned}$$

$d(\overline{z}_n, \mathcal{S}) \to 0$ if (C1) is satisfied in $V_3 \setminus \mathcal{S}$.

The result for $d \in \mathbb{N}$

Let

$$\Sigma_{d-1} = \left\{ \left(q^{(0)}, \ldots, q^{(d-1)}\right) \;\middle|\; q^{(i)} \geq 0, \sum_{i=1}^{d-1} q^{(i)} = 1 \right\}$$

$$V_d = \Sigma_{d-1} + [0,1] \cdot \mathbb{1}_d$$

$$\mathcal{S}_d = \left\{ x + \gamma \mathbb{1}_d \in V_d \;\middle|\; \gamma \geq \max\left\{x^{(0)}, \ldots, x^{(d-1)}\right\} \right\}$$

$$v_n = \overline{x}_n + \overline{\gamma}_n \mathbb{1}_d$$

Theorem 3

Let $d \geq 2$. There exists a generalized Blackwell algorithm such that for every sequence $x_1, x_2, x_3, \ldots$ with values in $D = \{0, 1, 2, \ldots, d-1\}$, it holds that $v_n \to \mathcal{S}$ as $n \to \infty$ almost surely.

We apply Theorem 2 to

$$M = \begin{pmatrix} (2,1,\dots,1) & \dots & (0,0,\dots,1) \\ & \ddots & \\ (1,0,\dots,0) & \dots & (1,1,\dots,2) \end{pmatrix}.$$

How to randomize?

Let e_i, $i = 0, \dots, d-1$ denote the standard unit vectors and $\mathbb{1}_d = (1, \dots, 1)$.

Let E_i denote the affine spaces

$$E_i = A(e_0, \dots, e_{i-1}, e_i + \mathbb{1}_d, e_{i+1}, \dots, e_{d-1}),\ i = 0, 1, \dots, d-1.$$

They have $n_i = \frac{2}{d}\mathbb{1}_d - e_i$, $i = 0, 1, \dots, d-1$ as normal vectors and intersect all in $s = (\frac{2}{d}, \dots, \frac{2}{d})$.

The E_i are pairwise perpendicular to each other and devide V_d in 2^d pieces. $\overline{v}_n$ lies in one of these pieces.

Then we have

$$\mathcal{S}_d = \{z_\gamma = x + \gamma\mathbb{1}_d \in V_d \mid \langle z_\gamma - n_i, n_i\rangle \geq 0,\ \forall i \in D\}$$

with $D = \{0, 1, 2, \dots, d-1\}$.

The definition of $p(v_n) \in \Sigma_{d-1}$

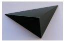
a) Let $v_n \notin \mathcal{S}_d$. Let $\{i_0, \ldots, i_j\}$ be a subset of $\{1, \ldots, n\}$ such that:
$\langle \overline{v}_n - n_l, n_l \rangle < 0$ for $l = i_0, \ldots, i_j$ with some $0 < j \leq d - 1$ and
$\langle \overline{v}_n - n_l, n_l \rangle \geq 0$ for all other l.
Let $A_1 = A\left(\frac{2}{d}\mathbb{1}_d, \overline{v}_n, e_{i_{j+1}}, \ldots, e_{i_{d-1}}\right)$ and $A_2 = A(e_{i_0}, e_{i_1}, \ldots, e_{i_j})$.
Let $A_1 \cap A_2 = \{p_0\}$. We put $p(v_n) = p_0$.

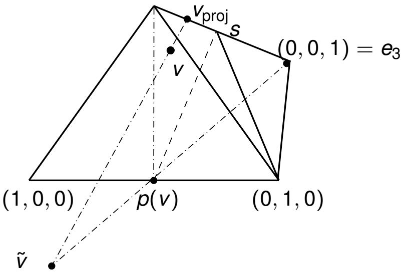
b) If $\overline{v}_n \in \partial\mathcal{S}_d$, let $\nu = \#\{i \in D \mid \overline{v}_n \in E_i\}$.
Then put
$$p^{(i)}(\overline{v}_n) := \begin{cases} \frac{1}{\nu} & \text{if } \overline{v}_n \in E_i \\ 0 & \text{if } \overline{v}_n \notin E_i \end{cases} \qquad \text{for } i = 0, \ldots, d-1.$$

c) If $\overline{v}_n \in \mathcal{S}_d \setminus \partial\mathcal{S}_d$, then put $p^{(i)}(\overline{v}_n) = \frac{1}{d}$ for $i = 0, \ldots, d-1$.

With this definition, condition (C1) of Theorem 2 ist satisfied.

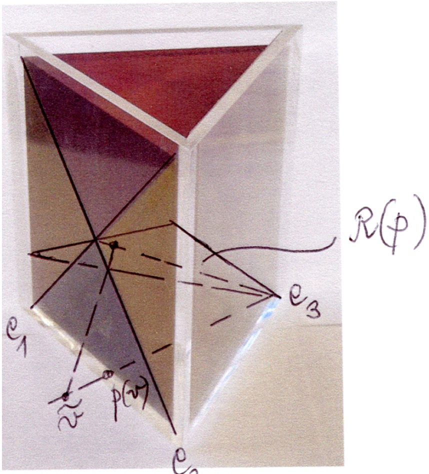
Why does (C1) hold? (for $d = 3$)

Consider **Type 2:** $v \in$ 



Construct $\tilde{v}$ as $\{\tilde{v}\} = A' \cap A''$ with $A' = A(v, v_{\text{proj}})$, $A'' = A(p(v), e_3)$.

Then $\tilde{v}_{\text{proj}} = v_{\text{proj}}$ and $v - v_{\text{proj}} \perp A(\mathcal{R}(p(v))) \Leftrightarrow \tilde{v} - \tilde{v}_{\text{proj}} \perp A(\mathcal{R}(p(v)))$.



Further we have: $\tilde{v} - \tilde{v}_{\text{proj}} \perp e_3 - \tilde{v}_{\text{proj}}$

$$\tilde{v} - \tilde{v}_{\text{proj}} \perp e_i + p^{(i)} \mathbb{1}_3 - \tilde{v}_{\text{proj}}, \quad i = 1, 2$$

Thus $A(\mathcal{R}(p(v))) = \tilde{v}_{\text{proj}} + \text{span}\left(e_3 - \tilde{v}_{\text{proj}}, e_i + \mathbb{1}_3 p^{(i)} - \tilde{v}_{\text{proj}}, \quad i = 1, 2\right)$

and $A(\mathcal{R}(p(v))) \perp \tilde{v} - \tilde{v}_{\text{proj}}$

Universal portfolios (due to T. Cover)

Let $\mathbf{x} = (x_1, \ldots, x_m) \geq 0$ be a market vector for one investment period.

x_i is the number of units returned from an investment of 1 unit in the i-th stock.

$\mathbf{b} = (b_1, \ldots, b_m) \geq 0$, $\sum_{i=1}^{m} b_i = 1$, a portfolio vector.

$S = \mathbf{b}'\mathbf{x}$ is the factor of capital increase using portfolio $\mathbf{b}$ for one period.

After n investment periods (starting with $S_0 = 1$) the capital is

$$S_n = \prod_{j=1}^{n} \mathbf{b}'\mathbf{x}_j = \exp\left(\int \ln(\mathbf{b}'\mathbf{x}) dp_n(\mathbf{x}) \right)$$

with $p_n := \frac{1}{n} \sum_{i=1}^{n} \mathbb{1}_{\{\mathbf{x}_i = \mathbf{x}\}}$ where $\mathbf{x} \in \{a_1, \ldots, a_M\} \subset \mathbb{R}^m$.

Let $W(\mathbf{b}, p) := \int \ln(\mathbf{b}'\mathbf{x})dp(\mathbf{x})$ and $W^*(p) := \sup_{\mathbf{b}} W(\mathbf{b}, p)$ for a probability measure p.

A sequential choice of portfolios $\mathbf{b}_1$, $\mathbf{b}_2(\mathbf{x}_1)$, ..., $\mathbf{b}_n(\mathbf{x}_1, \ldots, \mathbf{x}_{n-1})$ results in capital

$$S_n = \prod_{k=1}^{n} \mathbf{b}_k'(\mathbf{x}_1, \ldots, \mathbf{x}_{k-1})\mathbf{x}_k$$

Let $W_n = \frac{1}{n} \ln S_n = \frac{1}{n} \sum_{k=1}^{n} \ln \mathbf{b}_k'\mathbf{x}_k$. W_n is the cumulative log-return using portfolio choice $\mathbf{b}_1, \mathbf{b}_2(\mathbf{x}_1), \ldots$ We seek to drive W_n above $W^*(p_n)$ by appropriate choices of $\mathbf{b}_k$.

Invoke Blackwell's approachability result.

Let $V = \{(p, W) \mid p \in \sum_{M-1}, |W| \leq L\} \subset \mathbb{R}^{M+1}$ and

$$\mathcal{S} = \{(p, W) \mid W \geq W^*(p), p \in \textstyle\sum_{M-1}\}.$$

Now consider the situation at time n.

Let $\mu_n = (p_n, W_n) \in \mathbb{R}^{M+1}$ denote the current empirical probability p_n and log-return W_n. Let $\mathbf{t}_n$ denote the closest point in the convex set $\mathcal{S}$ to μ_n. Denote $\mathbf{t}_n = (\widehat{p}_n, \widehat{W}_n)$. By construction we have $\widehat{W}_n = W^*(\widehat{p}_n)$. There is also the portfolio $\mathbf{b}^*(\widehat{p}_n) = \arg\max_{\mathbf{b}} W(\mathbf{b}, \widehat{p}_n)$.

$\mathbf{b}^*(\widehat{p}_n)$ generates the supporting hyperplane at $\mathbf{t}_n$:

$$H_n = \left\{(p, W) \mid p \in \Sigma_{M-1}, W = \int \ln(\mathbf{b}^{*\prime}\mathbf{x})dp(\mathbf{x})\right\}.$$

We now set $\mathbf{b}_{n+1} = \mathbf{b}^*(\widehat{p}_n)$.

Define $z_{n+1} = (\mathbb{1}(\mathbf{x}_{n+1}), \ln \mathbf{b}'_{n+1}\mathbf{x}_{n+1})$

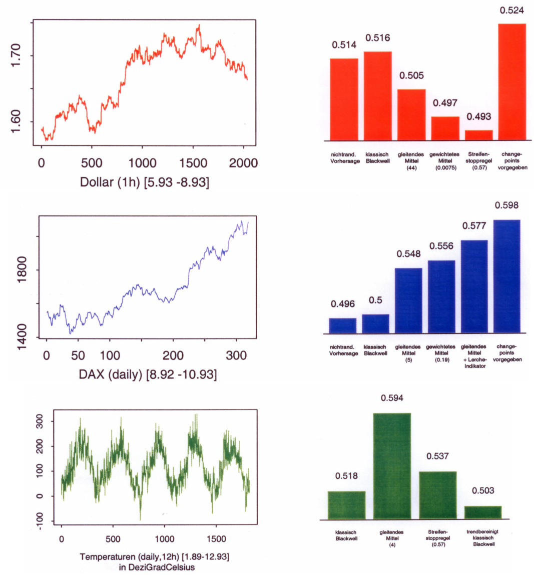
and observe that $\mu_{n+1} = \dfrac{1}{n+1}\displaystyle\sum_{k=1}^{n+1} z_k$.

Then $d(\mu_n, S) \to 0$ and as a consequence

$$W_n - W(\mathbf{b}^*(p_n)) \to 0 \text{ as } n \to \infty.$$

What is harder to predict
the US-Dollar, the DAX, or the weather?

relative frequency of success



References

- Blackwell, D. (1956): An Analog of the Minimax Theorem for Vector Payoffs. Pacific Journal of Mathematics, 6, 1–8.
- Cover,T. M.(1991): Universal Portfolios. Mathematical Finance, 1, 1–29.
- Lerche, H.R., Sakar, J. (1994): The Blackwell Prediction Algorithm for infinite 0-1 sequences and a generalization. In *Statistical Decision Theory and Related Topies V*, Ed.: S.S. Gupta, J.O. Berger, Springer Verlag, 503–511.
- Riedel, F. (2008): Blackwell's Theorem with Weighted Averages. Preprint.
- Sandvoss, R. (1994): Blackwell Vorhersageverfahren – zur Komplexität von Finanzdaten. Diplomarbeit Universität Freiburg.

E N D