The Blackwell Prediction for $0 - 1$ Sequences and a Generalization

Hans Rudolf Lerche

Department of Mathematical Stochastics

University of Freiburg

DAG Stat 2013, Freiburg

March 19, 2013

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 sequence $y_1, y_2, y_3, \ldots$ is a random 0-1 sequence with y_{n+1} being the predicted value of x_{n+1}.

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 for the i-th outcome,

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

A plausible deterministic prediction scheme:

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

$$y_1^{det} = 1.$$

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

If $x_1, x_2, x_3, \ldots$ are independent Bernoulli (p), then for $(y_n^{det}; n \geq 1)$

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

by the law of large numbers. Bernoulli (1713).

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

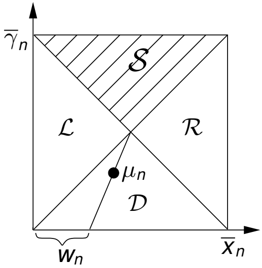
Question: Does there exist a prediction procedure with

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

for all infinite $0 - 1$ sequence ?

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

$$y_{n+1} = \begin{cases} 0 & \text{if } \mu_n \in \mathcal{L} \\ 1 & \text{if } \mu_n \in \mathcal{R} \\ 1 & \text{with probability } w_n \text{ if } \mu_n \in \mathcal{D} \end{cases}$$

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

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. $d(\mu_n, \mathcal{S}) \to 0$ almost surely as $n \to \infty$.

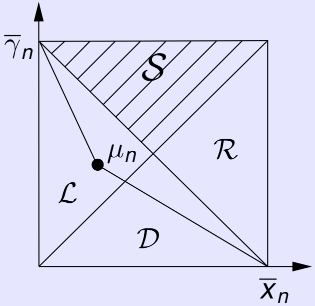
Remark

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: $\mu_n \in \mathcal{L}$



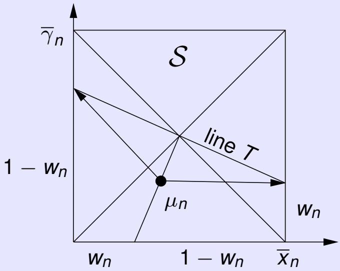
Then $d_{n+1} = \frac{n}{n+1} d_n$.

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

Case 3: $\mu_n \in \mathcal{D}$

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

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



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

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

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

It holds (*) $E\left(d_{n+1}^2 \mid \text{past}(n)\right) \leq \left(\frac{n}{n+1}\right)^2 d_n^2 + \frac{1}{2(n+1)^2}$ for $\mu_n \in \mathcal{D}$ with $d_n = d(\mu_n, \mathcal{S})$.

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

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

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 $\mu_n \in \mathcal{D}$

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

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

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

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 $\mathcal{L}$, $\mathcal{R}$ 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. □

Classical
Blackwell
Prediction

Prediction
for $d \geq 3$

References

Riedel's Result on nonequal Weights

Let $(g_n, n \geq 1)$ be a sequence of positive numbers and let $G_n = \sum_{i=1}^{n} g_i$.

Let $\widetilde{x}_n = \frac{1}{G_n} \sum_{i=1}^{n} g_i x_i$ and γ_n as above. Let $\mu_n = (\widetilde{x}_n, \gamma_n)$.

Theorem 2

Assume (i) $\sum_{n \geq 1} \left(\frac{g_n}{G_n} \right)^2 < \infty$ *and* (ii) $\sum_{n \geq 1} \left(\frac{g_n}{G_n} \right) = \infty$.

Then $d(\mu_n, \mathcal{S}) \to 0$ *almost surely as* $n \to \infty$.

Examples:

1) $g_n = n^\gamma$ for some $\gamma > 0$. Then $\frac{g_n}{G_n} = O\left(\frac{1}{n}\right)$.

2) $g_n = e^{\lambda n^\alpha}$ for some $\lambda > 0$. Then $\frac{g_n}{G_n} = O\left(n^{\alpha - 1}\right)$.

Thus: convergence for $0 < \alpha < \frac{1}{2}$.

3) $g_n = e^{\lambda n}$. No convergence !

Sequential prediction of $d \geq 3$ categories

Let $x_1, x_2, x_3, \ldots$ be a infinite sequence with outcomes in $D = \{0, 1, \ldots, d-1\}$.

Let $\overline{x}_n^{(j)}$ denote the relative frequency of the j-th outcome up to n and

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

Let $y_1, y_2, y_3, \ldots$ be a sequence of predictors with values in D and γ_n the relative frequency of correct predictions.

Question: Is there an algorithm such that

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

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

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

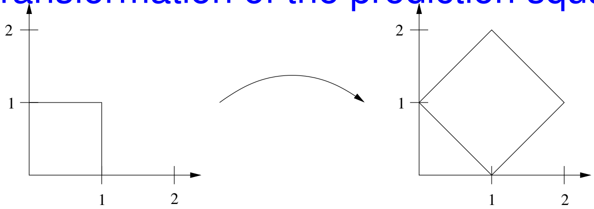
$$(\overline{x}_n, \overline{\gamma}_n) \to \mathscr{S} \ ?$$

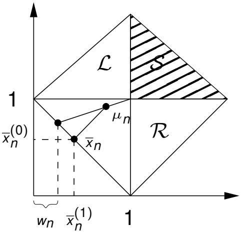
Problem: The argument of Theorem 1 does not carry over directly since there are no right angles in $\mathscr{S}$. See $d = 3$.

Exercise: By which factor has one have to stretch the $[0,1]$-axis, to get right angles of the cutting planes in the stretched prism?

Back to the case with two outcomes:

A transformation of the prediction square





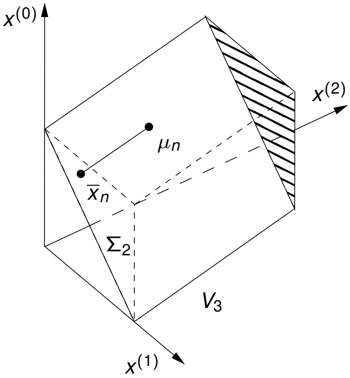
$$\mu_n = \left(\overline{x}_n^{(0)}, \overline{x}_n^{(1)}\right) + \overline{\gamma}_n(1, 1)$$

A basis for generalisations to more than two categories.

Prediction for $d = 3$

A natural generalization:

Instead $(\overline{x_n}, \overline{\gamma}_n)$ we use $\mu_n = \overline{x}_n + \overline{\gamma}_n \, \mathbb{1}_3$ with $\mathbb{1}_3 = (1, 1, 1)$.



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

The geometric structure of V_3

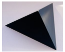
We cut V_3 from each of its upper vertices down to the two lower vertices. This yields 8 pieces of 4 different types. $\mathcal{S}$ is the piece on the top.

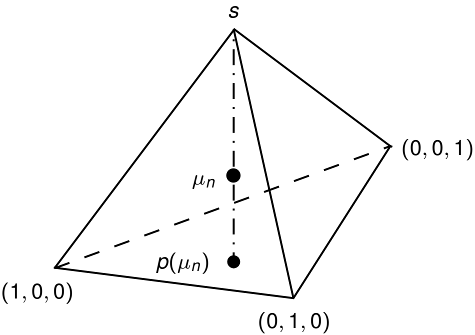
The cutting planes have $s = (\frac{2}{3}, \frac{2}{3}, \frac{2}{3})$ as joint point and are perpendicular to each other.

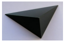
How does the algorithm randomize in the different pieces?

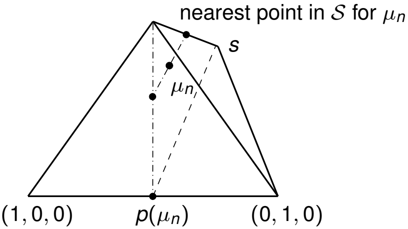
Geometrical interpretation of the randomisation probability

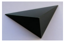
$p(\mu_n) := (p^{(0)}(\mu_n), p^{(1)}(\mu_n), p^{(2)}(\mu_n))$ as follows:

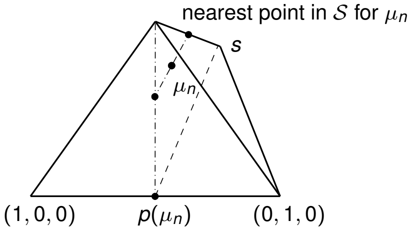
Type 1: $\mu_n \in$ 

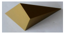


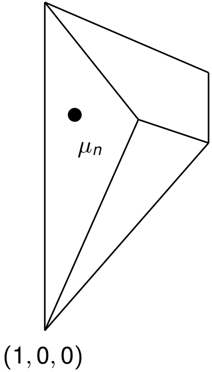
Type 2: $\mu_n \in$ 

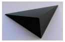


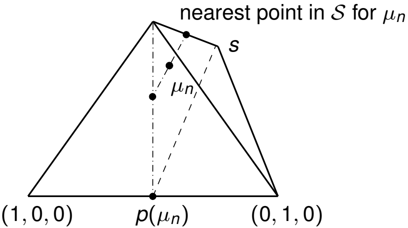
Type 2: $\mu_n \in$ 

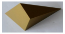


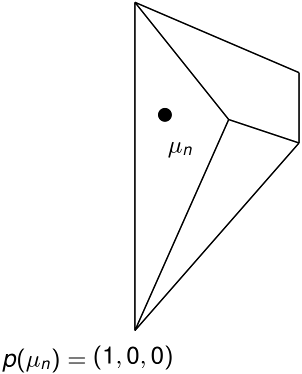
Type 3: $\mu_n \in$ 



Type 2: $\mu_n \in$ 



Type 3: $\mu_n \in$ 



The Result for $d = 3$

$$\Sigma_2 = \left\{ \left(q^{(0)}, q^{(1)}, q^{(2)}\right) \;\middle|\; q^{(i)} \geq 0, \sum_{i=0}^{2} q^{(i)} = 1 \right\}$$

$$V_3 = \Sigma_2 + [0,1] \cdot \mathbb{1}_3 , \qquad \mathbb{1}_3 = (1,1,1)$$

$$\mathcal{S}_3 = \left\{ x + \gamma \mathbb{1}_3 \in V_3 \;\middle|\; \gamma \geq \max\left\{x^{(0)}, x^{(1)}, x^{(2)}\right\} \right\}$$

$$\mu_n = \overline{x}_n + \overline{\gamma}_n \mathbb{1}_3 .$$

Theorem 3

Let $d = 3$. For the generalized Blackwell algorithm it holds:

For every sequence $x_1, x_2, x_3, \ldots$ with values in $\{0,1,2\}$

$$d(\mu_n, \mathcal{S}_3) \to 0 \quad \textit{almost surely} \quad \textit{as } n \to \infty.$$

The Result for $d \geq 3$

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

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

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

Theorem 4

Let $d \geq 3$. There exists a generalized Blackwell algorithm such that for every sequence $x_1, x_2, x_3, \dots$ with values in $D = \{0, 1, 2, \dots, d-1\}$, it holds that

$$d(\mu_n, \mathcal{S}_d) \to 0 \quad \textit{almost surely} \quad \textit{as } n \to \infty.$$

How to randomize?

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

Let E_i denote the affine spaces

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

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

The E_i are pairwise perpendicular to each other and devide V_d in 2^d pieces. μ_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 $D = \{0, 1, 2, \ldots, d-1\}$.

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

a) Let $\mu_n \notin \mathcal{S}_d$. Let $\{i_0, \dots, i_j\}$ be a subset of $\{1, \dots, n\}$ such that:

$\langle \mu_n - n_l, n_l \rangle < 0$ for $l = i_0, \dots, i_j$ with some $0 < j \leq d - 1$ and

$\langle \mu_n - n_l, n_l \rangle \geq 0$ for all other l.

Let $A_1 = A\left(\frac{2}{d}\mathbb{1}_d, \mu_n, e_{i_{j+1}}, \dots, e_{i_{d-1}}\right)$ and $A_2 = A(e_{i_0}, e_{i_1}, \dots, e_{i_j})$.

Let $A_1 \cap A_2 = \{p_0\}$. We put $p(\mu_n) = p_0$.

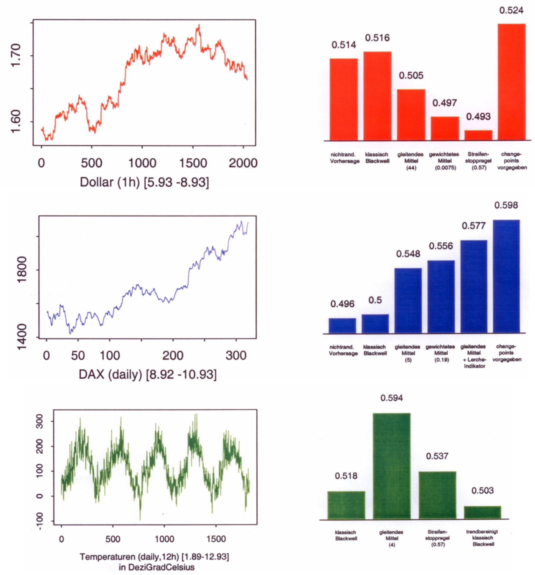
b) If $\mu_n \in \partial\mathcal{S}_d$, let $\nu = \#\{i \in D \mid \mu_n \in E_i\}$.

Then put $$p^{(i)}(\mu_n) := \begin{cases} \frac{1}{\nu} & \text{if } \mu_n \in E_i \\ 0 & \text{if } \mu_n \notin E_i \end{cases} \qquad \text{for } i = 0, \dots, d-1.$$

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

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

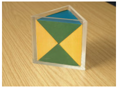
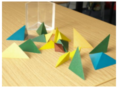
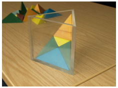
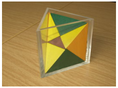
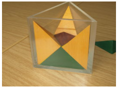
relative frequency of success



The Puzzle to the Prism

We cut V_3 not only from one vertex of above to two below, but also vice versa.

- How many pieces show up?
- How looks the central piece?

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.
- Lerche, H. R. (2012): Blackwell Prediction of Categorical Data. Game Theory and Applications, 15, 139–151 (arXiv:1102.2729).
- Riedel, F. (2008): Blackwell's Theorem with Weighted Averages. Preprint.
- Sandvoss, R. (1994): Blackwell Vorhersageverfahren – zur Komplexität von Finanzdaten. Diplomarbeit Universität Freiburg.

Thank you

for your attention !