Obtaining the Best Value for Money in Adaptive Sequential Estimation

Janne V. Kujala*

September 25, 2010

Abstract

In [Kujala, J. V., Richardson, U., & Lyttinen, H. (2010). A Bayesian-optimal principle for learner-friendly adaptation in learning games. Journal of Mathematical Psychology, 54, 247–255], we considered an extension of the conventional Bayesian adaptive estimation framework to situations where each observable variable is associated with a certain random cost of observation. We proposed an algorithm that chooses each placement by maximizing the expected gain in utility divided by the expected cost. In this paper, we formally justify this placement rule as an asymptotically optimal solution to the problem of maximizing the expected utility of an experiment that terminates when the total cost overruns a given budget. For example, the cost could be defined as the random time taken by each trial in an experiment, and one might wish to maximize the expected total information gain over as many trials as can be completed in 15 minutes. A simple, analytically tractable example is considered.

Keywords: Bayesian adaptive estimation; active learning; optimal design; decision theory; cost of observation; psychophysics.

1 Introduction

The topic of this paper is Bayesian estimation of an unobservable random variable $\Theta$ based on a sequence $y_{x_1}, \ldots, y_{x_T}$ of independent (given $\Theta$) realizations from some conditional densities $p(y_{x_t} \mid \Theta)$ indexed by trial placements $x_t$, each of which can be adaptively chosen from some set $X_t \subset X$ based on the outcomes $y := (y_{x_1}, \ldots, y_{x_{t-1}})$ of the earlier observations. We assume that the goal of choosing the placements is to maximize the expected value $E(U_T)$ of some utility function (DeGroot, 1970) $U_t := u(Y_{x_1}, \ldots, Y_{x_t})$ of the knowledge about $\Theta$ under some constraints such as a given total number $T$ of trials. The same decision theoretic framework applies to other problems as well, such as model discrimination (Cavagnaro et al., 2010; Myung and Pitt, 2009), where the utility function is based on the discriminability of two or more models based on the observed data.

In (Kujala et al., 2010), we extended this framework by considering situations where the observation of $Y_{x_t}$ is associated with some random cost $C_{x_t}$. We proposed an algorithm that chooses each placement by maximizing the expected gain in utility divided by the expected cost. In this paper, we formally justify this placement rule as an asymptotically optimal solution to the problem of maximizing the expected utility $E(U_T)$ of an experiment that terminates when the total cost overruns a given budget.

Costs of observation have been previously considered in the context of decision theory by DeGroot (1970) in several examples and there have been practical applications for such models especially in clinical trial design (e.g., Müller et al., 2006). Also, Paninski (2005) mentions a model where a certain price $C(x_{t-1}, x_t)$ has to be paid for each change of the state of the “observational apparatus” from $x_{t-1}$ to $x_t$. What is common to all these examples is that the proposed objective function is based on some difference of the expected gain and the cost. However, with such an objective, one has to equate the units of gain with the units of cost. Our approach avoids this problem by using a different kind of placement strategy based on a ratio of gain and cost instead.

2 Preliminaries

In Bayesian adaptive estimation, the loss function (negative of a utility function) can be defined, e.g., as the posterior variance of a single unknown variable
of any vector $\Theta$ of unknown variables (Kujala and Lukka, 2006; MacKay, 1992). The entropy loss function (1) is often desirable as it has certain very useful properties. In particular, the expected difference between the prior entropy and the posterior entropy given $Y_x$ is parametrization invariant (Lindley, 1956). Hence, even though the value $U_t$ of the utility function defined as the (negative) entropy as well as its change $U_t - U_{t-1}$ do depend on the parametrization chosen for $\theta$, the expected change $E(U_t - U_{t-1} \mid y, X_t = x)$ is parametrization invariant; it corresponds to the information-theoretic mutual information

$$I(\Theta; Y_x \mid y) = H(\Theta \mid y) - E[H(\Theta \mid y, Y_x) \mid y] \quad (2)$$

of $\Theta$ and $Y_x$ (given $y$).\(^3\)

Any adaptive (or non-adaptive) placement strategy defines a decision function

$$d : Y_d \rightarrow X \cup \{\lambda\} : (y_{x_1}, \ldots, y_{x_{T-1}}) \mapsto x_t,$$

whose domain $Y_d$ is the set of possible sequences of trial results (including the empty sequence) and whose value is the next placement or the special value $\lambda$ which flags the end of the experiment. Now we can define the random variable

$$Y_d := (Y_{X_1}, \ldots, Y_{X_T}) \in Y_d$$

denoting the outcome of the whole adaptive experiment governed by the decision function $d$, where $T$ is the possibly random time index of the trial that ends the experiment.

The fact that the whole adaptive experiment can be seen as just one observation $Y_d$ implies that the parametrization-invariance of the expected change of the entropy applies over the whole experiment as well, for any strategy $d$, regardless of whether the termination rule is adaptive or not, as long as the experiment eventually terminates (and even that is not strictly necessary if the value of the utility function converges with probability one). One could even allow randomized decision functions mapping to distributions over $X \cup \{\lambda\}$ instead of deterministic values. However, as randomized decisions generally gain nothing over deterministic ones (DeGroot, 1970), we shall only consider deterministic decision functions except for one reference to a random termination rule in Sec. 3 (where the randomness is outside the experimenter’s control).

Most commonly adaptive estimation algorithms use a so called greedy strategy, which chooses each placement so as to optimize the expected immediate gain in utility after the next observation:

$$x_t := \arg \max_{x \in X_t} E(U_t - U_{t-1} \mid y, X_t = x).$$

This strategy can be applied in any model for any utility function, but it is generally not the globally optimal strategy except when the experiment consists of exactly one trial.\(^4\) However, for the entropy loss function, under certain mild regularity conditions, it can be shown that the greedy strategy is asymptotically more efficient than any non-adaptive strategy (Paninski, 2005).

The globally optimal strategy is obviously to maximize the expected utility after the observation of the whole experiment result $Y_d$ w.r.t. the decision function $d$. For example, in the case of dichotomous results and constant experiment length $T$, the decision function defines a binary decision tree with $2^T - 1$ nodes, each denoting the next placement after a certain sequence of observations. Thus, the optimal strategy can in theory be found by optimizing over these $2^T - 1$ parameters. In practice, the exponentially growing number of parameters makes the optimal strategy generally intractable very soon as $T$ grows (although there are exceptions, see, e.g., Brockwell and Kadane, 2003).

3 Obtaining the best value for money

In this section, we consider the extended framework where the observation of $Y_x$ is associated with some random cost $C_x$, which given the value of $Y_x$, is independent

\(^2\)We use the same notation $\int f(\theta)p(\theta)d\theta$ for both the continuous case and the discrete case, in which it corresponds to a sum. This is measure-theoretically justified as ““d\theta” can be considered as the counting measure in the discrete case. Thus, following Lindley (1956), even though we use the familiar notation, we are in fact working in full measure-theoretic generality, allowing the density $d\theta$ to be w.r.t. any measure “d\theta”.

\(^3\)In our notation $H(A \mid \ldots)$ always denotes the conditional entropy of $A$ given the (possibly random) conditioning values $\ldots$. Thus, $H(A \mid \ldots)$ will be a random variable if any of its conditioning values are random variables unlike in the unfortunate standard notation where an expectation is implicitly taken over the conditioning values. Also, $I(\Theta; Y_x \mid y)$ denotes the mutual information of the random variables $\Theta \mid y$ and $Y_x \mid y$, that is, both $\Theta$ and $Y_x$ are conditioned on $y$. This is standard notation.

\(^4\)Strictly speaking it is possible that no maximum of the expected utility exists, in which case one should generally choose such a placement that yields an expected gain sufficiently close to the supremum (DeGroot, 1970).
of $\Theta$ and the results and costs of any other observations:

$$
\Theta \\
\sqrt{\rightarrow} \downarrow \searrow \\
Y_x \quad Y_{x'} \quad \ldots \\
\downarrow \downarrow \\
C_x \quad C_{x'} \quad \ldots
$$

The technical requirement that $C_x$ depends on $\Theta$ only through $Y_x$ is satisfied in particular if $C_x$ is a component of $Y_x$. Thus, it leads to no loss of generality if the incurred costs are observable.

Intuitively, we would like to make measurement more cost-effective by taking into account the fact that different trial placements may be associated with different costs. We consider two operational definitions of this goal:

1. maximizing the expected amount of information given by an experiment that terminates when the total cost overruns a certain predetermined budget (for example, the cost could be defined as the random time taken by each trial in an experiment, and one might wish to maximize the expected total information gain over as many trials as can be completed in 15 minutes);

2. minimizing the expected cost of an experiment that terminates when a predetermined amount of information has been obtained (for example, one might wish to minimize the expected time required to measure an unobservable variable to a certain predetermined level of accuracy).

Both definitions are reasonable, but the first one turns out to be more elegant in that it corresponds to the plain expected information maximization goal with an adaptive termination rule as discussed in Sec. 2. Hence, the optimal strategy under that goal is insensitive to the parametrization used to define the differential entropy measure of information. In contrast, simple counterexamples show that the second definition does not have this desirable property.

**Remark 1.** In the statement of the problem, we do not require the cost $C_x$ to be observable. However, for the experiment to actually terminate when the budget is overrun, either the actual costs must be observable, or alternatively, any further trials could simply fail after the budget is overrun. In the latter case, the adaptive termination rule would be random. In either case, the actual costs are irrelevant to the final Bayesian estimates of $\Theta$ as they are assumed to depend on $\Theta$ only through the fully observable results $Y_x$.

**Remark 2.** Obviously exact maximization of the expected information gain under a given budget is generally intractable for the same reasons that the usual constant number of trials case is. However, even if the information gains and costs associated with each $Y_x$ were known time-invariant constants, the problem of fitting the best value in a given constant budget would still be intractable — it is equivalent to the knapsack problem which is NP hard (see, Garey and Johnson, 1979). The heuristic we use is in fact similar to the heuristics used to find approximate solutions to the knapsack problem although we have the additional complications of randomness and the generally intractable sequential changes.

### 3.1 Heuristics

Let us define random variables denoting the gain and cost of the $t$-th observation:

$$
G_t := U_t - U_{t-1} = u(Y_{X_1}, \ldots, Y_{X_t}) - u(Y_{X_1}, \ldots, Y_{X_{t-1}}),
$$

$$
C_t := C_{X_t}.
$$

Assuming for a moment that the cost $C_x$ is defined as the time taken by the observation of $Y_x$, one might think that choosing $x$ so as to maximize the expected rate of information gain

$$
E \left( \frac{G_t}{C_t} \mid y, \ X_t = x \right)
$$

over the duration $C_t$ of the next observation would be a good heuristic. However, even though the expected rate of gain is optimal over the next trial, that is in general no longer true over repeated trials. Over a large constant unit of time, repeated i.i.d. observations of $Y_x$ are expected to result in each outcome $(y_x, c_x)$ being observed for a total cumulative duration proportional to $p(y_x, c_x \mid y)c_x$. Thus, to estimate the average rate of gain obtainable from i.i.d. observations of $Y_x$, the expectation should be taken over the distribution

$$
\int \int p(y_x, c_x \mid y)c_x \ dy_x \ dx_c
$$

instead. This leads to the objective function

$$
\int \int \left[ u(y, y_x) - u(y) \right] c_x \ dy_x \ dx_c \int \int p(y_x, c_x \mid y)c_x \ dy_x \ dx_c
$$

$$
= \frac{E(G_t \mid y, \ X_t = x)}{E(C_t \mid y, \ X_t = x)}
$$

(4)

proposed in (Kujala et al., 2010). With the entropy loss function (1), this can be written in the convenient form

$$
\frac{I(Y_x; \Theta \mid y)}{E(C_x \mid y)}.
$$

(5)
Unlike the situation in the pure information maximization case, maximization of (4) does not correspond to maximization of the immediate expected utility. Thus, it is not the prototypical greedy algorithm, but it is still myopic in the sense that it expects the future sets of possible expected gains and costs to be similar to those of the current trial. In our view, this algorithm is the most natural generalization of the ubiquitous one-step greedy strategy to the situation where the costs of observation vary.

Remark 3. For the problem to be well-defined, the expected cost of a trial should always be positive. However, in some practically interesting situations there may be a positive probability that the actual cost is zero. (This is another reason why (3) is an inappropriate formulation—it would always be infinite in this case). In fact, as long as the expectation is positive, there is no reason why we should not allow negative costs, too.

3.2 Simple conditions for asymptotic optimality

Now we are in a position to formally prove the technical statements about asymptotic optimality that were announced in (Kujala et al., 2010). The following proposition implies that maximization of (4) is asymptotically optimal in the sense that it gives the asymptotically best gain-to-cost ratio if the set of the pairs of marginal distributions of $(G_t, C_t)$ over all possible values of $x$ do not change as each new outcome is added to the data $y$. That is, the same marginal distributions of gain and cost are allowed to be associated with different $x$ at different times as long as these $x$’s at different times are in a one-to-one correspondence.

Proposition 1. Suppose that the random variables $G$ and $C$ have finite expectations $E(G)$ and $E(C) \neq 0$. Then, almost surely (i.e., with probability 1)

$$\lim_{n \to \infty} \frac{G_1 + \cdots + G_n}{C_1 + \cdots + C_n} = \frac{E(G)}{E(C)},$$

where $G_k$ and $C_k$ denote i.i.d. copies of $G$ and $C$, respectively.

Under certain side conditions, optimality of the strategy can also be shown under the weaker assumption that the objective function always has the same maximum value $\alpha$ regardless of the past data:

Theorem 2. Suppose that there exists a constant $\alpha > 0$ such that

$$\max_{x \in \mathcal{X}} \frac{E(G_t \mid y, X_t = x)}{E(C_t \mid y, X_t = x)} = \alpha \quad (6)$$

for all possible sets $y$ of past observations. If $X_t$ is defined as the maximizer of (6) and if for some $\sigma^2 < \infty$ and $\varepsilon > 0$,

$$\begin{align*}
\text{Var}(G_t \mid Y_{X_t}, \ldots, Y_{X_{t-1}}) &\leq \sigma^2, \\
\text{Var}(C_t \mid Y_{X_t}, \ldots, Y_{X_{t-1}}) &\leq \sigma^2, \\
E(C_t \mid Y_{X_t}, \ldots, Y_{X_{t-1}}) &\geq \varepsilon
\end{align*} \quad (7)$$

for all $t$, then the gain-to-cost ratio satisfies

$$\lim_{t \to \infty} \frac{G_1 + \cdots + G_t}{C_1 + \cdots + C_t} \text{ a.s.} = \alpha.$$

This is asymptotically optimal in the sense that for any other strategy that satisfies (7), we have

$$\limsup_{t \to \infty} \frac{G_1 + \cdots + G_t}{C_1 + \cdots + C_t} \text{ a.s.} \leq \alpha.$$

Of course, these conditions for asymptotic optimality are simple (in contrast with the theory of Paninski, 2005) and cannot be expected to hold in all practical models. We present them here simply to formalize the intuitive justification for the method that in the long run, only the expected values matter and the random variations in the gains and costs get averaged out. In the next section, we give a simple example where these conditions do apply.

Figure 1: A simplified version of the psychometric model typically assumed in psychophysical experiments, with an infinite slope at the threshold $\theta$. 
4 Example: A simplified psychometric model

Suppose that for all \( x \in \mathbb{R}, Y_x \in \{0, 1\} \) is a dichotomous random variable defined by
\[
\Pr\{Y_x = 1 \mid \theta\} = \begin{cases} a, & x < \theta \\ b, & x \geq \theta \end{cases}
\]
for some \( 0 < a < b < 1 \), see Fig. 1. While this model is simple, it does have the important feature that there is uncertainty of the results due to both the fact that \( \Theta \) is unknown and the fact that for a given \( \theta \), the result \( Y_x \mid \theta \) is random. This model differs from the typical psychometric model only in that, due to the infinite slope at the threshold, the obtainable information gains do not decrease over time as the scale of the uncertainty reduces. However, this detail is not very important during the first few trials, and therefore this example serves to illustrate the relative efficiencies of different placement strategies that can be expected in a typical psychophysical experiment and explain the success of the greedy strategy in psychophysical measurement.

4.1 Success of the greedy strategy

Denoting by \( h(p) := -p \log p - (1 - p) \log(1 - p) \) the entropy of a binary distribution with probabilities \( p \) and \( 1 - p \),\(^5\) the expected information gain of observing \( Y_x \) given any prior data \( y = (y_{x_1}, \ldots, y_{x_{t-1}}) \) can be written as
\[
I(Y_x; \Theta \mid y) = H(Y_x \mid y) - E[H(Y_x \mid \Theta) \mid y] = h(a + (b - a) Pr\{\Theta \leq x \mid y\}) - (h(a) + (h(b) - h(a)) Pr\{\Theta \leq x \mid y\}).
\]
This expression depends on the placement \( x \) and the prior data \( y \) only through \( z := Pr\{\Theta \leq x \mid y\} \). Assuming that the prior on \( \Theta \) is absolutely continuous w.r.t. the Lebesgue measure, the same will be true for the posterior given \( y \), and one can always choose \( x \) so as to attain any value of \( z \in [0, 1] \). As (8) is continuous on the compact interval \( z \in [0, 1] \), it attains a maximum value (which is independent of any prior data). It follows that the greedy strategy yields the maximum expected total information gain over any given constant number of trials.

4.2 Optimal strategy under varying costs

We shall now extend the model with random costs \( C_x \) associated with observing \( Y_x \). For simplicity, let us assume that the distribution of \( C_x \) is fully determined by the value of \( Y_x \), i.e., it does not depend directly on \( x \). Then,
\[
E(C_x \mid \theta) = \begin{cases} c_a, & x < \theta \\ c_b, & x \geq \theta \end{cases}
\]
for some \( c_a \) and \( c_b \), which we assume to be positive to avoid pathological cases.

As before, the expected information gain is given by (8), and so the objective function is
\[
\frac{I(Y_x; \Theta \mid y)}{E(C_x \mid y)} = \frac{h(a + (b - a)z) - (h(a) + (h(b) - h(a))z)}{c_a + (c_b - c_a)z},
\]
where we denote \( z := Pr\{\Theta \leq x \mid y\} \). As before, this expression depends on the prior data \( y \) and the placement \( x \) only through \( z \) which can always attain any value in \([0, 1]\). Thus, placing the next trial so as to maximize (9) is by Theorem 2 the asymptotically optimal strategy as it yields the same maximum value of the objective function at every trial regardless of the prior data (and it is easy to show that the other assumptions of the theorem hold, too). Proposition 1 does apply as well, since the distributions of gain and cost at different trials are identical for placements that correspond to the same \( z \) value.

4.3 Specific examples

After some tedious algebra, it can be shown that the objective function (9) has a unique maximizer corresponding to a zero of the expression
\[
c_a h(a) - c_a h(b) + (c_a a - c_b b) \log \omega + c_b (1 - a) - c_a (1 - b)) \log(1 - \omega)
\]
where \( \omega := a + (b - a)z \). As before, this expression depends on the prior data \( y \) and the placement \( x \) only through \( z \) which can always attain any value in \([0, 1]\).

Assuming first the trivial case \( C_x = 1 \), i.e., constant cost, we have \( c_a = c_b = 1 \) and the optimality condition becomes \( h(a) - h(b) + (a - b) \log \omega + (b - a) \log(1 - \omega) = 0 \), which yields the optimizer
\[
\omega^* = \frac{1}{1 + \exp\left(\frac{h(b) - h(a)}{b - a}\right)}.
\]

To give an example of a nontrivial definition of cost, in (Kujala et al., 2010), it was assumed that each failure of the observer costs one unit due to potential loss of motivation. Applying the same definition here, we have \( C_x = |Y_x - 0| \), i.e., each 0-result costs one unit. It follows \( c_a = 1 - a \) and \( c_b = 1 - b \) and the optimality condition becomes \( (1 - b)h(a) - (1 - a)h(b) + (a - b) \log \omega = 0 \), which yields the optimizer
\[
\omega^* = \exp\left(-\frac{(1 - a)h(b) - (1 - b)h(a)}{b - a}\right).
\]
Figure 2: (Left) An illustration of the optimal placements for the simplified psychometric model shown in Fig. 1 for different values of $a$ and $b$. These are the optimal placements assuming a uniform prior for $\Theta$ on $[0,1]$. For any other (absolutely continuous) prior, the optimal placement is obtained by interpreting the value on the $y$-axis as the fractile $z = \Pr(\Theta \leq x \mid y)$. (Right) Optimal placements under a random cost of observation defined as $C_x = 1 - I_x$ (i.e., each 0-result costs one unit).

Table 1: Cross-comparison of the values (in bits) of the two objective functions (2) and (5) under the two adaptive strategies (max-info and cost-aware), each of which maximizes one of the two objective functions. The cost-aware algorithm has significantly better cost-efficiency, and conversely, significantly worse per-trial efficiency.

Table 1

<table>
<thead>
<tr>
<th></th>
<th>$a = .05, b = .95$</th>
<th>$a = .05, b = .8$</th>
</tr>
</thead>
<tbody>
<tr>
<td>max-info</td>
<td>$I(Y_x;\Theta</td>
<td>y)$</td>
</tr>
<tr>
<td>cost-aware</td>
<td>$I(Y_x;\Theta</td>
<td>y)$</td>
</tr>
<tr>
<td>$E(C_x)$</td>
<td>$E(C_x)$</td>
<td>$E(C_x)$</td>
</tr>
<tr>
<td>max-info</td>
<td>0.714</td>
<td>0.481</td>
</tr>
<tr>
<td>cost-aware</td>
<td>0.394</td>
<td>0.408</td>
</tr>
<tr>
<td>$E(C_x)$</td>
<td>1.427</td>
<td>0.803</td>
</tr>
<tr>
<td>$E(C_x)$</td>
<td>2.187</td>
<td>0.925</td>
</tr>
</tbody>
</table>

Figure 2 shows the optimal placements (10) and (11) for both cases. For the case of unit cost, the optimal placement appears to be close to the median of the distribution of $[\Theta | y]$ under several values of $a$ and $b$. With two-alternative forced choice design (i.e., guessing rate $a = 1/2$), the optimal placement tends to $z = 0.6$ as $b \to 1$, and for any $a$ and $b$, the placement $z$ is within $[1/e, 1 - 1/e] \approx [0.3679, 0.6321]$. Comparing these placements to those of the cost-aware algorithm supports the intuitive characterization given in (Kujala et al., 2010), while the pure information maximization works much like binary search, roughly bisecting the uncertainty distribution at each step, the cost-aware variation instead chooses the placement at a certain higher percentile closer to the easier end. The exact percentile of the optimal placement seems to depend mostly on the lapsing rate $1 - b$ and less on the guessing rate $a$. This dependence was to be expected: if there is going to be a large probability of careless mistakes anyway, then it does not pay off to make the trials very easy, and conversely, if careless mistakes are unlikely, then the easiest trials will be virtually free and the placements close to that end will yield the best value for money even though the gains over one trial are smaller.

Table 1 quantifies the improvement in cost-efficiency that the cost-aware algorithm provides over the plain information maximization strategy (as well as the corresponding deterioration in per-trial efficiency).

5 Discussion

We have shown that the cost-aware placement strategy is asymptotically optimal under certain idealized conditions. These conditions do not always hold in practice, but it can still be expected that the strategy works well in the same models where the greedy information maximization strategy works with constant per-trial cost. In particular, we expect that Paninski’s (2005) asymptotic effi-
ciency result for the greedy strategy mentioned in Sec. 2 can be generalized for the cost-aware strategy under the same mild regularity conditions.

In all our examples so far, we have only considered discrete cost variables. An obvious topic for future work is calculation of the objective function (5) for some response time model with the cost $C_x$ defined as the response time (including any pre-stimulus delays). In any experiments where the placement of a trial can affect its duration, this formulation can increase the efficiency per time unit over the pure information maximization greedy algorithm. In particular, in an $n$-choice task, the response times generally increase with $n$, and so a smaller value of $n$ might turn out to be optimal even though it yields a higher guessing rate. The cost-aware formulation may also be useful in adaptive memory retention experiments (see, e.g., Cavagnaro et al., 2010), where the delay after which the recall of a stimulus is tested is varied. Another topic for future work is generalization of the cost-aware strategy for a multi-step “lookahead” (see, e.g., King-Smith et al., 1994) to bring it closer to the globally optimal strategy.

A Proofs

Proof of Proposition 1. Having finite expectations, the i.i.d. sequences $G_k$ and $C_k$ satisfy the strong law of large numbers:

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} (G_k - E(G_k)) \overset{a.s.}{=} 0,$$

$$\lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} (C_k - E(C_k)) \overset{a.s.}{=} 0.$$

Thus,

$$\lim_{n \to \infty} \frac{\sum_{k=1}^{n} G_k}{\sum_{k=1}^{n} C_k} \overset{a.s.}{=} \frac{E(G)}{E(C)}.$$

Theorem 3 (A strong law of large numbers for martingales). Let $X_t = Z_1 + \cdots + Z_t$ be a martingale. If

$$\sum_{t=1}^{\infty} \frac{E(|Z_t|^2)}{t^2} < \infty,$$

then $X_t/t \overset{a.s.}{\to} 0$.

Proof. E.g., (Chow, 1967) or (Shiryaev, 1996, Theorem 4, p. 519).

Corollary 4. Suppose $Y_t$ and $Z_t$ are sequences of random variables such that $Z_t$ is independent of $Z_1, \ldots, Z_{t-1}$ given $y_1, \ldots, y_{t-1}$ and

$$\begin{cases} E(Z_t \mid y_1, \ldots, y_{t-1}) = 0, \\ \text{Var}(Z_t \mid y_1, \ldots, y_{t-1}) \leq \sigma^2, \end{cases}$$

for all values of $t$ and $y_1, \ldots, y_{t-1}$. Then,

$$\frac{Z_1 + \cdots + Z_t}{t} \overset{a.s.}{\to} 0.$$

Proof. Denoting $y = (y_1, \ldots, y_{t-1})$, the conditional independence assumption and the assumption on the conditional expectation yield

$$E(Z_t \mid Z_1, \ldots, Z_{t-1}) = \int E(Z_t \mid Z_1, \ldots, Z_{t-1}, y)p(y)dy$$

$$= \int E(Z_t \mid y)p(y)dy = 0,$$

which implies that $Z_1 + \cdots + Z_t$ is a martingale. Furthermore, the assumptions on the conditional statistics yield

$$E(Z_t^2) = \int E(Z_t^2 \mid y)p(y)dy$$

$$= \int [\text{Var}(Z_t \mid y) + E(Z_t \mid y)^2] dy \leq \sigma^2,$$

which implies

$$\sum_{t=1}^{\infty} \frac{E(Z_t^2)}{t^2} \leq \frac{\sigma^2}{6} < \infty.$$

Hence, the assumptions of Theorem 3 are satisfied and we obtain the statement.

Proof of Theorem 2. Assuming that $X_t$ is chosen as the maximizer of (6), the sequences $Y_t := Y_X$, and $Z_t := G_t - \alpha C_t$ satisfy the assumptions of Corollary 4 with $E(G_t - \alpha C_t \mid Y_1, \ldots, Y_{t-1}) = 0$ and $\text{Var}(G_t - \alpha C_t \mid Y_1, \ldots, Y_{t-1}) \leq (1 + \alpha)^2 \sigma^2$ for all $t$, and we obtain $(Z_1 + \cdots + Z_t)/t \overset{a.s.}{\to} 0$. The sequences $W_t := C_t - E(C_t \mid Y_1, \ldots, Y_{t-1})$ and $Y_t$ also satisfy the assumptions of Corollary 4, and we obtain

$$\frac{C_1 + \cdots + C_t}{t} = \frac{\sum_{k=1}^{t} E(C_k \mid Y_1, \ldots, Y_{k-1})}{t}$$

$$= \frac{W_1 + \cdots + W_t}{t} \overset{a.s.}{\to} 0.$$

With the assumption $E(C_t \mid Y_1, \ldots, Y_{t-1}) \geq \varepsilon$, this implies $\lim_{t \to \infty} (C_1 + \cdots + C_t)/t \geq \varepsilon$, and we obtain

$$\frac{G_1 + \cdots + G_t}{C_1 + \cdots + C_t} = \frac{(G_1 - \alpha C_1) + \cdots + (G_t - \alpha C_t)}{C_1 + \cdots + C_t} + \alpha$$

$$= \frac{(Z_1 + \cdots + Z_t)/t}{(C_1 + \cdots + C_t)/t} + \alpha \overset{a.s.}{\to} \alpha.$$
To prove the latter part of the theorem, let the logic of choosing $X_t$ now be arbitrary. The sequence $Z'_t := G'_t - \alpha C_t$, where we use the upwards adjusted gains 

$$G'_t := G_t + E(\alpha C_t - G_t | Y_1, \ldots, Y_{t-1}) \geq 0 \text{ by (6)}$$

satisfies the assumptions of Corollary 4 with $E(G'_t - \alpha C_t | Y_1, \ldots, Y_{t-1}) = 0$ and $\text{Var}(G'_t - \alpha C_t | Y_1, \ldots, Y_{t-1}) \leq (1 + \alpha)^2 \sigma^2$ (the adjustment is constant given $Y_1, \ldots, Y_{t-1}$ so it does not affect this variance bound). Repeating the same steps as in the optimal case, we obtain

$$\frac{G_1 + \cdots + G_t}{C_1 + \cdots + C_t} \leq \frac{G'_1 + \cdots + G'_t}{C_1 + \cdots + C_t} \text{ a.s.} \rightarrow \alpha,$$

which implies the statement.

Acknowledgments

This research was supported by the Academy of Finland (grant number 121855) and by the European Commission’s FP6 Marie Curie Excellence Grants (MCEXT-CE-2004-014203). The author would like to thank Benja Fallenstein, Antti Penttinen, and Rauli Ruohonen for discussions and two anonymous reviewers and an anonymous action editor for many helpful suggestions.

References


