OPTIMAL SELECTION OF INFORMATION WITH RESTRICTED STORAGE CAPACITY

L. Pronzato

Laboratoire I3S, CNRS–UPRESA 6070
Bât. 4, 250 rue A. Einstein,
Sophia Antipolis, 06560 Valbonne, France

ABSTRACT

We consider the situation where n items have to be selected among a series of N presented sequentially, the information contained in each item being random. The problem is to get a collection of n items with maximal information. We consider the case where the information is additive, and thus need to maximize the sum of n independently identically distributed random variables x_k observed sequentially in a sequence of length N. This is a stochastic dynamic–programming problem, the optimal solution of which is derived when the distribution of the x_k's is known. The asymptotic behaviour of this optimal solution (when N tends to infinity with n fixed) is considered. A (forced) certainty–equivalence policy is proposed for the case where the distribution is unknown and estimated on–line.

1. INTRODUCTION AND PROBLEM STATEMENT

We consider the situation where a sequence of items y_k is proposed, e.g. for further delayed processing, the information contained in each y_k being measured by a scalar variable x_k. The sequence $\{y_k\}$ has length N and the storage capacity is $n < N$. The problem is then to derive on–line a decision rule for the storage of the y_k's which maximizes the information contained in the items. The storage is permanent: items are selected forever, that is any item selected at time j cannot be replaced by another one at time $k > j$.

Another possible setup is when a scalar parameter θ has to be estimated from observations y_k, described by $y_k = f(\theta, z_k) + \epsilon_k$ with $\{\epsilon_k\}$ an i.i.d. sequence of random variables. N experiments, characterized by the z_k's are proposed, but only n can be performed. Their selection forms a problem of *sequential experiment design*. The criterion is Fisher information, given by $\sum_{i=1}^{n} \omega^2(\theta^0, z_{k_i})$, with $k_i \in \{1, \ldots, N\}$, $\omega(\theta, z) = \partial f(\theta, z)/\partial\theta$, and θ^0 some prior nominal value of θ.

We shall assume throughout the paper that the x_k's are independent random variables. The dependent case (e.g. Markov process), and the case where information is not additive (e.g. when $y_k = f(\boldsymbol{\theta}, z_k) + \epsilon_k$, with $\boldsymbol{\theta}$ a p-dimensional vector), deserve further studies.

Let $\{u_k\}$ be the decision sequence, with $u_k = 1$ if y_k is stored/observed at time k, and $u_k = 0$ otherwise. The problem is then to maximize

$$J_N(\boldsymbol{u}_1^N) = \sum_{k=1}^{N} u_k x_k \,, \tag{1}$$

with $\boldsymbol{u}_1^N = (u_1, \ldots, u_N)$ satisfying the constraint

$$\sum_{k=1}^{N} u_k \le n \,, \tag{2}$$

the constraint will be saturated at the optimum, see (3). For any sequence $\boldsymbol{u}_1^N$ and any time (or step) j, $1 \le j \le N$, let a_j be the number of items already stored, that is $a_j = \sum_{k=1}^{j-1} u_k$ with $a_1 = 0$, and $c(j, a_j, \boldsymbol{u}_1^N)$ be the gain–to–go,

$$c(j, a_j, \boldsymbol{u}_j^N) = \sum_{k=j}^{N} u_k x_k = x_j u_j + c(j+1, a_{j+1}, \boldsymbol{u}_{j+1}^N) \,.$$

Note that x_j is known when u_j has to be chosen. Two on–line decision policies are considered in Section 2, which aim at maximizing the gain–to–go. Their asymptotic behaviour ($N \to \infty$) is studied in Section 3. Special attention is devoted to the *optimal* (closed–loop) policy. The case where the distribution of the x_k's is unknown is considered in Section 4.

2. DECISION POLICIES

The x_k's are assumed to be independent non–negative random variables with probability measure $\mu(\cdot)$. We shall denote $F(\cdot)$ the c.d.f. of x, $E\{.\}$ the expectation

w.r.t. x, and $\hat{p}(s) = 1 - F(s)$, $\hat{x}(s) = \int_s^\infty x\mu(dx)$. The measure $\mu(\cdot)$ is assumed to be such that $E\{x\} < \infty$. At step j the state of the decision process is summarized in $\mathcal{I}_j = (j, a_j)$. Only *adaptive* (or feedback) deterministic decision policies will be considered, i.e. policies such that, at step j, u_j is a deterministic function of $\mathcal{I}_j$. The two policies to be considered share some properties. Assume first that one reached a situation where $a_j + N - j + 1 \leq n$, which means that retaining every item till the end would make a total of no more than n items selected. It is then optimal to select all items, that is

$$a_j + N - j + 1 \leq n \Rightarrow u_k = 1,\ k = j, \ldots, N\,, \quad (3)$$

which gives equality in (2). Assume now that n items have already been selected. One obviously has

$$a_j = n \Rightarrow u_k = 0,\ k = j, \ldots, N\,. \quad (4)$$

Only the case $n - N + j - 1 < a_j < n$ thus remains to be treated. A first *suboptimal* approach is to choose the decision sequence in open–loop, which ignores the fact that future decisions u_k, $k > j$ could make use of information $\mathcal{I}_k$, see, e.g., [1].

2.1. Open–loop feedback

At step j, with $n - N + j - 1 < a_j < n$, u_j is chosen so as to maximize the expected gain–to–go, that is

$$\hat{u}_j = \arg \max_{u \in \{0,1\}} ux_j + E\{c(j+1, a_{j+1}, \boldsymbol{u}_{j+1}^N)\}\,,$$

with $\boldsymbol{u}_{j+1}^N$ a fixed decision policy for future steps, that is function of $\mathcal{I}_j$ only. This restriction yields

$$E\{c(j+1, a_{j+1}, \boldsymbol{u}_{j+1}^N)\} = (n - a_{j+1})E\{x\}\,,$$

with $a_{j+1} = a_j + 1$ if $u = 1$ and $a_{j+1} = a_j$ otherwise. The optimal policy is thus as follows:

$$\hat{u}_j(\mathcal{I}_j) = \begin{cases} 0 & \text{if } x_j \leq E\{x\} \text{ or } a_j \geq n\,, \\ 1 & \text{otherwise.} \end{cases}$$

The choice $\hat{u}_j(\mathcal{I}_j) = 0$ when $x_j = E\{x\}$ is arbitrary. The open–loop feedback–optimal decision rule thus relies on the comparison of x_j with a fixed threshold, here the expected value of x. We can also consider more general decision rules using a fixed threshold s:

$$\hat{u}_j^s(\mathcal{I}_j) = \begin{cases} 0 & \text{if } x_j \leq s \text{ or } a_j \geq n\,, \\ 1 & \text{otherwise.} \end{cases} \quad (5)$$

The expected gain–to–go obtained with such a policy will be denoted by $E\{\hat{c}^s(j, a_j)\}$. It satisfies the recurrence equation

$$\begin{aligned} E\{\hat{c}^s(j,a_j)\} &= \hat{x}(s) + \hat{p}(s)E\{\hat{c}^s(j+1, a_j+1)\} \\ &\quad + [1 - \hat{p}(s)]E\{\hat{c}^s(j+1, a_j)\}\,. \end{aligned} \quad (6)$$

The two cases (3) and (4) respectively give $E\{\hat{c}^s(j, n + j - N - 1)\} = (N - j + 1)E\{x\}$ and $\hat{c}^s(j, n) = 0$ for all x_j. The expected gains $E\{\hat{c}^s(j, n - i)\}$ can then be computed from the recurrence above.

2.2. Closed–loop optimal decisions

We use a dynamic programming approach, see Section 4 and [2] for more details. Denote the optimal expected gain–to–go at step j by $\tilde{c}(j, a_j)$. When $a_j < n$ it satisfies the recurrence equation

$$\tilde{c}(j, a_j) = \max_{u \in \{0,1\}} ux_j + E\{\tilde{c}(j+1, a_{j+1})\}\,,$$

with $a_{j+1} = a_j + 1$ if $u = 1$ and $a_{j+1} = a_j$ otherwise. The optimal decision is thus

$$\tilde{u}_j(\mathcal{I}_j) = \begin{cases} 0 & \text{if } x_j \leq E\{\tilde{c}(j+1, a_j)\} \\ & \quad -E\{\tilde{c}(j+1, a_j+1)\} \text{ or } a_j \geq n\,, \\ 1 & \text{otherwise,} \end{cases} \quad (7)$$

which gives

$$\tilde{c}(j, a_j) = \max[x_j + E\{\tilde{c}(j+1, a_j+1)\}, E\{\tilde{c}(j+1, a_j)\}]\,.$$

This yields the following backward recurrence equation for $E\{\tilde{c}(j, a_j)\}$:

$$\begin{aligned} &E\{\tilde{c}(j,a_j)\} = \\ &\quad \hat{x}[E\{\tilde{c}(j+1,a_j)\} - E\{\tilde{c}(j+1,a_j+1)\}] \\ &\quad + \hat{p}[E\{\tilde{c}(j+1,a_j)\} - E\{\tilde{c}(j+1,a_j+1)\}] \\ &\quad \times E\{\tilde{c}(j+1,a_j+1)\}) \\ &\quad + [1 - \hat{p}(E\{\tilde{c}(j+1,a_j)\} - E\{\tilde{c}(j+1,a_j+1)\}] \\ &\quad \times E\{\tilde{c}(j+1,a_j)\}\,. \end{aligned} \quad (8)$$

The two cases (3–4) still give $\tilde{c}(j, j + n - N - 1) = (N - j + 1)E\{x\}$ and $\tilde{c}(j, n) = 0$ for all x_j. The optimal decision (7) at step j is obtained by comparing x_j to the threshold $E\{\tilde{c}(j+1, a_j)\} - E\{\tilde{c}(j+1, a_j+1)\}$, which is a function of j and a_j. The optimal thresholds $E\{\tilde{c}(j+1, n-i)\} - E\{\tilde{c}(j+1, n-i+1)\}$ can thus be computed from the recurrence above.

3. PERFORMANCE CHARACTERISTICS

3.1. Fixed threshold

For any step j, characterized by $\mathcal{I}_j = (j, a_j)$, denote the expected gain–to–go $E\{\hat{c}^s(j, a_j)\}$ by $\hat{\lambda}_m^k$, with $k = n - a_j$ the current storage capacity and $m = N + 1 - j - n + a_j$ the number of steps–to–go before reaching the situation (3). One thus gets with this new notation:

$$\begin{aligned} \hat{\lambda}_0^k &= kE\{x\} \quad \forall k \geq 0\,, \\ \hat{\lambda}_m^0 &= 0 \quad \forall m \geq 0\,, \end{aligned} \quad (9)$$

and to the backward recurrence (6) corresponds the forward recurrence:

$$\hat{\lambda}_m^k = \hat{x}(s) + \hat{p}(s)\hat{\lambda}_m^{k-1} + [1-\hat{p}(s)]\hat{\lambda}_{m-1}^k \,. \qquad (10)$$

The analytic expression of $\hat{\lambda}_m^k$ is then as follows [5].

Theorem 1 *For any $m, k \geq 0$,*

$$\begin{aligned}\hat{\lambda}_m^k &= E\{x\}[1-\hat{p}(s)]^m \sum_{j=0}^{k-1} \hat{p}^j(s)(k-j)\mathtt{C}_{m+j-1}^j \\ &\quad + \hat{x}(s) \sum_{j=0}^{k-1} \hat{p}^j(s) \sum_{l=0}^{m-1} [1-\hat{p}(s)]^l \mathtt{C}_{l+j}^j \,. \qquad (11)\end{aligned}$$

The limiting behaviour of $\hat{\lambda}_m^k$ when m tends to infinity is given by the following theorem, see [5].

Theorem 2 *For any fixed $k \geq 0$ and any s such that $\hat{p}(s) > 0$ the decision policy (5) is such that*

$$\lim_{m\to\infty} \hat{\lambda}_m^k = k\hat{x}(s)/\hat{p}(s) \,. \qquad (12)$$

3.2. Optimal decisions

We use notations similar to previous section, and denote the optimal expected gain–to–go $E\{\tilde{c}^s(j, a_j)\}$ by $\tilde{\lambda}_m^k$, with $k = n - a_j$ and $m = N + 1 - j - n + a_j$. The two situations (3) and (4) now give

$$\begin{aligned}\tilde{\lambda}_0^k &= kE\{x\} \quad \forall k \geq 0 \,. \\ \tilde{\lambda}_m^0 &= 0 \quad \forall m \geq 0 \,. \qquad (13)\end{aligned}$$

Next theorem gives the limiting performances of the optimal decision rule when the support of the probability measure $\mu(\cdot)$ is bounded from above [5].

Theorem 3 *Assume that $\bar{M} = \min\{x | F(x) = 1\} < \infty$. Then for any fixed $k \geq 0$*

$$\lim_{m\to\infty} \tilde{\lambda}_m^k = k\bar{M} \,. \qquad (14)$$

From the theorems above, the performances of the optimal policy can be far superior to those of the open–loop decisions, e.g. when the probability measure $\mu(\cdot)$ has a density with thin tail. Consider now the case of a measure with density $\varphi(\cdot)$ having an infinite support. One can show that $\lim_{m\to\infty} \tilde{\lambda}_m^k = \infty$ for any $k > 0$, so that in this case the optimal policy (7) will outperform any open–loop policy (5). Analytic results can be obtained in the case where the tail of $\varphi(\cdot)$ is exponentially decreasing [5].

Theorem 4 *Assume that the measure $\mu(\cdot)$ has a density $\varphi(\cdot)$ which satisfies: $\exists x_0$ such that $\forall x > x_0$,*

$$\varphi(x) = a\exp(-bx) \,. \qquad (15)$$

Then for any fixed $k > 0$

$$\tilde{\lambda}_m^k = \frac{k\log m}{b} + \frac{1}{b}\log(\frac{a^k}{b^k k!}) + o(1) \,, \; m \to \infty \,.$$

One can show, moreover, that exponentially decreasing tails are the only ones such that $s_m^2 - s_m^1$ tends to a constant $c \neq 0$ when m tends to infinity [5], where $s_m^k = \tilde{\lambda}_m^k - \tilde{\lambda}_{m+1}^{k-1}$ corresponds to the optimal threshold in the decision rule (7). When the tail of $\varphi(\cdot)$ is not exponentially decreasing, numerical integration can be used to compute $\hat{x}(s)$ and $\hat{p}(s)$. Heavier the tail of $\varphi(\cdot)$, faster the increase of $\tilde{\lambda}_m^k$ as m grows with k fixed.

4. ON–LINE ESTIMATION OF THE DISTRIBUTION

In previous sections, optimal decisions were derived from the knowledge of $\mu(\cdot)$. We assume now that $\mu(\cdot)$ is unknown and estimated on–line, with $\hat{\mu}_j$ the measure estimated at time j (after x_j has been observed). We denote $\mathcal{J}_j = (j, a_j, \boldsymbol{x}_1^j)$ the information used at time j for both estimation and decision. An important feature of the problem is that decisions are *neutral* with respect to estimation (see [3] for a definition of neutrality in control problems): the sequence $\{x_k\}$ is observed whatever the decision sequence $\{u_k\}$, and decisions have no effect on the accuracy of the estimation of the distribution. The problem to be solved at step j can be written as

$$\begin{aligned}\max_{u_j\in\mathcal{U}_j} [u_j x_j + \hat{E}_j\{ &\max_{u_{j+1}\in\mathcal{U}_{j+1}} [u_{j+1}x_{j+1} + \cdots \\ &\hat{E}_{N-2}\{\max_{u_{N-1}\in\mathcal{U}_{N-1}} [u_{N-1}x_{N-1} \\ &+ \hat{E}_{N-1}\{\max_{u_N\in\mathcal{U}_N} [u_N x_N]\}]\}\cdots]\}] \,,\end{aligned}$$

where $\hat{E}_k\{.\}$ denotes $E\{.|\mathcal{J}_k\}$ and

$$\mathcal{U}_k = \begin{cases} \{0\} & \text{if } a_k \geq n \,, \\ \{0,1\} & \text{otherwise.} \end{cases}$$

It seems particularly difficult to derive the optimal policy for $N > 3$, $n > 1$. We thus only consider decisions based on forced *certainty equivalence*: at step j, $\tilde{u}_j^{CE}(\mathcal{J}_j)$ is determined as in Section 2.2, with $\mu(\cdot)$ replaced by $\hat{\mu}_j(\cdot)$, that is $\tilde{u}_j^{CE}(\mathcal{J}_j) = \tilde{u}(\mathcal{I}_j | \mu = \hat{\mu}_j)$. We show in [5] that, in spite of the neutrality property,

this forced certainty equivalence policy is suboptimal, which contradicts a conjecture by Patchell and Jacobs [4] (see also [1]). We use for $\hat{\mu}_j$ the empirical distribution of the x_k's, initialized by M random samples. Each expected value $\hat{E}_k\{.\}$ is thus based on $k+M$ samples.

Example:

Assume that the true measure $\bar{\mu}(\cdot)$ has the truncated normal density $\varphi(x) = \frac{\sqrt{2}}{\sigma\sqrt{\pi}} \exp(-\frac{x^2}{2\sigma^2})$, $x \geq 0$. We take $N = 100$, $n = 10$, and use the empirical distribution of the x_k's to evaluate expected values $\hat{E}_k\{.\}$. Five random samples $x_{-4}, \ldots, x_0$ are used for the initialization. 500 repetitions of the experiment led to the results in Table 1.

	$E\{J_N(\boldsymbol{u}_1^N)\}$	$\mathrm{std}\{J_N(\boldsymbol{u}_1^N)\}$
$u_j = \tilde{u}_j^{CE}(\mathcal{J}_j)$	19.2	1.8
$u_j = \hat{u}_j^{CE}(\mathcal{J}_j)$	13.4	2.0

Table 1: Empirical means and standard–deviation of the cumulative gains (1) for the certainty–equivalence closed–loop and open–loop feedback policies (500 repetitions).

The values of the optimal expected gain–to–go $\tilde{c}(1,0)$ and open–loop expected gain $\hat{c}(1,0)$ for $\mu = \bar{\mu}$ are respectively 19.79 and 13.66. $\diamond$

The example above shows that the decrease of performances due to the estimation of the distribution can be marginal. Note, however, that initialization may be crucial for a particular run: a bad choice for $\hat{\mu}_0$ may produce the selection of first items proposed, with small associated values of x, before the empirical distribution is corrected. It seems advisable for that reason to choose M not too small ($M = 5$ in Example 3). The measure $\mu(\cdot)$ could also be parameterized, e.g. with a density $\varphi(x) = \varphi(x, \theta)$, and θ estimated on–line (see the case $x = \omega^2(\theta, z)$ in the introduction). Again, certainty equivalence (θ replaced by $\hat{\theta}_j$ estimated from $\boldsymbol{x}_1^j$ at time j) would give a solution easy to implement but in general suboptimal.

5. FURTHER DEVELOPMENTS AND CONCLUSIONS

The optimal decision rule for the maximization of the sum of n i.i.d. variables in a sequence of length N has been derived. Its superiority over an open–loop feedback policy has been evidenced. The asymptotic behaviours of both policies have been considered in the case where n is fixed while N tends to infinity. Further considerations could concern the case where both n and N tend to infinity, with say n/N fixed.

The case where the variables are dependent deserves further studies. It would also be of special interest to study how these results could be extended to the case where the criterion to be maximized is not additive. For instance, in the multidimensional case, one may wish to maximize $\det \sum_{i=1}^n \boldsymbol{\omega}(\theta^0, z_{k_i}) \boldsymbol{\omega}^T(\theta^0, z_{k_i})$, with random z_k's, $\boldsymbol{\omega}(\theta, z) = \partial f(\boldsymbol{\theta}, z)/\partial \boldsymbol{\theta}$ and $\boldsymbol{\theta}$ p-dimensional.

Certainty equivalence has been forced in the case where the distribution of the random variables is unknown. This approach is not optimal, although the problem can be stated as neutral. Quantifying the resulting loss of optimality then seems a difficult but challenging problem.

Acknowledgements
The author wishes to thank Prof. Anatoly A. Zhigljavsky for motivating and helpful discussions.

6. REFERENCES

[1] Y. Bar-Shalom and E. Tse. Dual effect, certainty equivalence, and separation in stochastic control. *IEEE Transactions on Automatic Control*, 19(5):494–500, 1974.

[2] D. Bertsekas, editor. *Dynamic Programming. Deterministic and Stochastic Models.* Prentice-Hall, Englewood Cliffs, 1987.

[3] A. Fel'dbaum. *Optimal Control Systems.* Academic Press, New York, 1965.

[4] J. Patchell and O. Jacobs. Separability, neutrality and certainty equivalence. *International Journal of Control*, 13(2):337–342, 1971.

[5] L. Pronzato. Optimal selection of information with restricted storage capacity. Technical Report 97–01, Laboratoire I3S, CNRS-URA 1376, Sophia Antipolis, 06560 Valbonne, France, January 1997.