LEIGH TESFATSION

A DUAL APPROACH TO BAYESIAN INFERENCE AND ADAPTIVE CONTROL

ABSTRACT. Probability updating via Bayes' rule often entails extensive informational and computational requirements. In consequence, relatively few practical applications of Bayesian adaptive control techniques have been attempted. This paper discusses an alternative approach to adaptive control, Bayesian in spirit, which shifts attention from the updating of probability distributions via transitional probability assessments to the direct updating of the criterion function, itself, via transitional utility assessments. Results are illustrated in terms of an adaptive reinvestment two-armed bandit problem.

1. INTRODUCTION

In numerous multi-stage control problems arising in the physical, biological, and social sciences, a decision maker attempting to control a system is forced to operate with necessarily incomplete information regarding the statistical and structural characteristics of his problem environment. The crux of the decision maker's problem is then to determine a tractable satisfactory procedure for using past observations to improve current and future system performance. Control problems incorporating this feature are generally referred to as adaptive.

Many currently available adaptive control techniques require the successive updating of system state probability distribution estimates via Bayes' rule. The need for state distribution estimation and updating arises from a focus on system state identification as a perceived essential intermediate step towards optimal control selection. Unfortunately, as is well documented [1, 6, 12, 13], the informational and computational requirements for probability distribution updating via Bayes' rule are often extensive. In consequence, relatively few practical applications of Bayesian adaptive control techniques have been attempted.[1]

This paper will discuss an alternative approach to adaptive control, Bayesian in spirit, which shifts attention from the updating of probability distributions via Bayes' rule, a filtering operation on transitional probability

assessments, to the direct updating of the criterion function, itself, via a filtering operation on transitional utility assessments. Analytical and computer simulation studies for the convergence and optimality properties of several specific criterion function filters are presented in previous papers [14–20]. In contrast, the present paper will focus on developing the fundamental analogy between use of criterion filters for direct criterion function updating and use of Bayes' rule for probability distribution updating. Results will be illustrated in terms of an adaptive reinvestment two-armed bandit problem.

2. THE BASIC ANALOGY

Consider, first, a simple example of the familiar parametric model underlying the Bayesian approach to statistical inference. Certain sample data, ω^*, arises as an observation of a random vector, ω. The distribution of ω is assumed to be specified by a probability model, $(\Omega, p(\omega|\theta))$, where $\Omega \subseteq R^s$ is a specified sample space, and $p(\cdot|\theta): \Omega \to R$ is a probability density function indexed by a parameter θ lying in a parameter space, Θ. The true value of θ is unknown, but the statistician has prior beliefs about this value which are expressable in the form of a prior probability density function, $p^0(\theta)$, defined over Θ. The sample ω^* increases his knowledge about θ; and the prior and sample information are combined to form an updated probability density function for θ,

$$(1) \qquad p(\theta|\omega^*) \propto p(\omega^*|\theta)p^0(\theta),$$

in accordance with Bayes' rule. If $n-1$ independent past observations $\omega_1^*, \ldots, \omega_{n-1}^*$ on ω are available to the statistician, then (1) generalizes to

$$(2) \qquad p(\theta|\omega_{n-1}^*, \ldots, \omega_1^*) \propto p(\omega_{n-1}^*|\theta)p(\omega_{n-2}^*|\theta) \ldots p(\omega_1^*|\theta)p^0(\theta).$$

Consider, now, the following simple adaptive control problem. A controller wishes to select a control, v, from a specified feasible control set, V, in an attempt to maximize his utility. The function $U(\omega, v)$ which measures his utility depends jointly on the selection v and the realization of a random vector ω taking values in some sample space $\Omega \subseteq R^s$. The form of the probability density function $p(\omega)$ governing ω is unknown, but the controller is able to express his prior beliefs about his expected utility in the form of a prior expected utility function, $U^0(v)$, defined over V. Certain sample data,

ω^*, now becomes available to the controller concerning a past realization for ω; and the controller combines his prior and sample information into an updated expected utility function

$$(3) \qquad U(v|\omega^*) = \frac{U(\omega^*, v) + U^0(v)}{2}, \quad v \in V.$$

A control $v^* \in V$ is then selected to maximize $U(v|\omega^*)$ over V.

Without the prior term $U^0(v)$, selection of a control to maximize (3) reduces to the following simple opportunity cost maxim: Select a control v^* which would have yielded maximum utility for the realized sample data, ω^*. Clearly this maxim is analogous to the maximum likelihood principle for selecting parameter estimates: Select the parameter value θ^* which maximizes the likelihood function $p(\omega^*|\theta)$. With the prior term averaged in, selection of a control v^* to maximize (3) is analogous to the Bayesian point estimate procedure of selecting a parameter estimate θ^* to maximize the posterior probability density function (1). (See [3, Chapter 6])

If a sequence $\omega_1^*, \ldots, \omega_{n-1}^*$ of $n-1$ independent past observations on ω is available, (3) generalizes to the nth period updated expected utility estimate for v,

$$(4) \qquad U_n^*(v) \equiv U(v|\omega_{n-1}^* \ldots, \omega_1^*) =$$
$$= \left(\frac{U(\omega_{n-1}^*, v) + \ldots + U(\omega_1^*, v) + U^0(v)}{n} \right)$$
$$= \left(\frac{U(\omega_{n-1}^*, v) + [n-1] U_{n-1}^*(v)}{n} \right).$$

If, for example, $U(\cdot, v)$ is bounded and continuous over Ω, it follows by the strong law of large numbers that

$$(5) \qquad U_n^*(v) \xrightarrow{n} \int_\Omega U(\omega, v) p(\omega) d\omega \quad \text{a.s.;}$$

i.e., $U_n^*(v)$ is a strongly consistent estimator for the true expected utility associated with the selection of v.

How might the prior expected utility function $U^0(v)$ be specified? If the sample space Ω is known to the controller, one plausible specification for $U^0(v)$ might be the barycentric prior

(6) $U^0(v) \equiv \int_\Omega U(\omega, v) p^0(\omega) d\omega, \quad v \in V,$

where $p^0(\omega)$ is a prior probability density function for ω. Alternatively, if information concerning the relevant sample space Ω is sketchy, $U^0(v)$ might be specified directly, without explicit consideration of probabilities for ω.[2] Prior ignorance might be represented by a prior expected utility function of the form

(7) $U^0(v) \equiv$ constant.

Notice that the problem of 'improper priors' does not arise for prior expected utility functions. Moreover, because of the additive form of (4), prior expected utility specifications are inevitably reduced to insignificance as the number of sample observations on ω increases. There is no need, as in standard Bayesian analysis, to impose positivity and continuity restrictions on prior beliefs to ensure the eventual dominance of the sample data.

3. FURTHER COMPARATIVE PROPERTIES OF CRITERION FILTERS AND POSTERIOR DISTRIBUTIONS

In complete analogy to the form of the Kalman-Bucy filter for sequential state estimation (see [15]), the linear filtering operation described by (4) transforms the vector $(U(\omega^*_{n-1}, v), \ldots, U(\omega^*_1, v), U^0(v))$ of prior and transitional utility assessments into an updated estimate for expected utility. For brevity, any such filtering operation will be referred to as a *criterion filter*. As demonstrated in [14–20], consistent criterion filters can be designed for adaptive control problems with intertemporal objectives and with random elements dependent on time as well as past control and system state realizations. The analogy with probability density function updating via Bayes' rule still holds, but in less transparent form. For conceptual clarity, we will continue in this section to develop the analogy with Bayes' rule probability updating in terms of the simple expressions (2) and (4) derived in Section 2.

A number of properties which facilitate the analysis of updated probability density functions such as (2) are equally useful for the analysis of updated expected utility estimates such as (4). Primary among these are the existence of conjugacy classes and the existence of sufficient statistics.

Recall that a prior probability density function $p^0(\theta)$ is said to be *conjugate* to the transitional probability density function $p(\omega|\theta)$ if the

posterior probability density function defined by (2) has the same general
functional form as the prior. Conjugate classes of prior and transitional densi-
ties for (2) become conjugate classes of prior and transitional utility functions
for the updated expected utility function estimate (4) under a simple logar-
ithmic transformation. For example, in complete analogy to the well-known
conjugacy property of the normal distribution, it is readily established that
quadratic specifications

(8a) $U(\omega, v) = a - b[\omega - v]^2 = a - b\omega^2 + 2b\omega v - bv^2$;

(8b) $U^0(v) = \int_\Omega U(\omega, v)p^0(\omega)\mathrm{d}\omega$,

for the prior and transitional utility assessments result in a quadratic
specification

(9) $U_n^*(v) = a - bs + 2bmv - bv^2$

for the corresponding expected utility estimate (4), where

$$(10a) \quad m \equiv \left(\frac{\sum\limits_{i=1}^{n-1} \omega_i^* + \int \omega p^0(\omega)\mathrm{d}\omega}{n} \right);$$

$$(10b) \quad s \equiv \left(\frac{\sum\limits_{i=1}^{n=1} \omega_i^{*2} + \int \omega^2 p^0(\omega)\mathrm{d}\omega}{n} \right).$$

As in probability theory, the usefulness of the conjugacy property for
prior and transitional utility functions is that it allows one to measure, ex-
plicitly, the relative contributions of the prior information and the sample
data information to the combined information provided by the updated ex-
pected utility estimate $U_n^*(v)$. For example, given the quadratic specifications
(8), the identities (10a) and (10b) reveal that each sample datum ω_i^* is
allocated the exact same importance as the prior information in forming the
updated expected utility estimate, a direct reflection of the simple nature of
the criterion filter presently being used.

Also as in probability theory, the advantage of nontrivial sufficient statis-
tics for criterion filter estimates such as $U_n^*(v)$ is that all relevant information

is summarized in expressions having lower dimension than the number of sample observations, which facilitates computation. Of particular interest is the existence of sufficient statistics of fixed dimension, independent of sample size.

Specifically, a function $T_n(\omega_{n-1}, \ldots, \omega_1)$ taking Ω^{n-1} into a space of dimension less than or equal to $n-1$ is appropriately termed a *sufficient statistic* for $U_n^*(\cdot)$ if

$$(11) \quad U_n^*(v) = G_n(T_n(\omega_{n-1}^*, \ldots, \omega_1^*), v), \quad v \in V,$$

for some function $G_n(\cdot)$. Using the simple equal-weight criterion filter depicted in (4), it is obvious that a minimal sufficient statistic for $U_n^*(v)$ is providied by the fixed-dimension vector $(\omega_{n-1}^*, U_{n-1}^*(v))$, regardless of the specification for the utility function. For more complicated adaptive control problems involving, e.g., time-dependent observations ω_j^*, this simple data reduction is no longer possible. However, conditions guaranteeing the existence of sufficient statistics for posterior probability density functions such as (2) are easily transformed into conditions guaranteeing the existence of sufficient statistics for criterion filter estimates such as $U_n^*(v)$, even for more complicated filter-weight specifications. For example, if the utility function has the form

$$(12) \quad U(\omega, v) = a + \sum_{k=1}^{r} b_k(v) h_k(\omega) + c(\omega) + \mathrm{d}(v),$$

it is readily verified that $U_n^*(\cdot)$ has form (11), where

$$(13) \quad T_n(\omega_{n-1}, \ldots, \omega_1) \equiv$$
$$\equiv \left(\sum_{i=1}^{n=1} h_1(\omega_i), \ldots, \sum_{i=1}^{n-1} h_r(\omega_i), \sum_{i=1}^{n-1} c(\omega_i) \right)$$

is a sufficient statistic of fixed dimension $r + 1$. (The quadratic utility function (8a) is a special case of (12) with $r = 1$.) Moreover, a sufficient statistic of form (13) continues to exist even when the filter weights are changed to reflect the possible dependence of observations on time, control selections, and past state values. (Cf. the time-dependent filter specification (29), below.)

4. AN ILLUSTRATIVE COMPARISON WITH BAYESIAN ADAPTIVE CONTROL TECHNIQUES

In this section the criterion filtering adaptive control technique will be compared to the well-known Bayesian-dynamic programming adaptive control technique in the context of the following adaptive reinvestment problem: In each period $n \in \{1, \ldots, N-1\}$, an investor must decide how to allocate his initial capital x_n between two investment opportunities A and B, the first which yields a positive or negative net return rate $\pm s(0 < s \leqq 1)$ with unknown probabilities $p_n^{\smallfrown}$ and $1 - p_n^{\smallfrown}$, respectively, and the second which yields a net return rate r_n $(0 \leqq r_n < s_n)$ with known probability 1. The objective of the investor is to maximize the expected value of the logarithm of his initial capital x_N for period N. The investor's initial capital endowment x_1^* for period 1 is positive.

Consider, first, the simple case previously treated in Refs. [2, 4, 10] in which $p_n^{\smallfrown} \equiv p^{\smallfrown}$ and $r_n \equiv 0$ for each $n \in \{1, \ldots, N-1\}$. Thus, a risky investment opportunity A with stationary unknown net return rate distribution must be compared against a safe neutral investment opportunity B in each period n. For notational convenience, assume $s \equiv 1$. (Extension to the more general case is strightforward.) The investor's initial capital x_{n+1} in period $n+1$ is then a simple function of his initial capital x_n in period n, the net return rate $\omega_n \in \{1, -1\}$ observed for investment opportunity A in period n, and the amount $v_n \in [0, x_n]$ of capital which the investor allocated to A in period n; namely $x_{n+1} = x_n + \omega_n v_n$.

Let $G(p)$ denote the investor's prior probability density function for $p^{\smallfrown}$, and assume the investor updates $G(p)$ in accordance with Bayes' rule as follows: If k positive net return rates and m negative net return rates have been observed for A, $G(p)$ is transformed into the posterior probability density function

$$(14) \qquad G_{km}(p) \equiv p^k [1 - p]^m G(p) / \int_0^1 p^k [1 - p]^m G(p) \mathrm{d}p.$$

Finally, let

$$(15) \qquad p_{km} \equiv \int_0^1 p G_{km}(p) \mathrm{d}p.$$

The Bayesian-dynamic programming solution $(v_1^B(x_1), \ldots, v_{N-1}^B(x_{N-1}))$

for the adaptive reinvestment problem may now be found as follows. For each n, x, k, and m, let $g_n(x, k, m)$ denote the maximum attainable expected utility $E \log(x_N)$ starting with initial capital x with $N - n$ stages to go, and using the posterior probability density function (14). Then [9, Thm. 14.4, p. 101, Thm. 15.2, p. 104, and Thm. 17.6, p. 111]

$$(16a) \quad g_{N-1}(x, k, m) = \max_{0 \leq v \leq x} (p_{km} \log(x + v) + [1 - p_{km}] \log(x - v));$$

$$(16b) \quad g_n(x, k, m) = \max_{0 \leq v \leq x} (p_{km} g_{n+1}(x + v, k + 1, m) +$$

$$+ [1 - p_{km}] g_{n+1}(x - v, k, m + 1)),$$

$1 \leq n \leq N - 2$, $x \in (0, \infty)$, and k and $m \in \{0, 1, \ldots\}$. Proceedings in real time from period 1 to period $N - 1$, generate capital allocations $v_n^B \equiv v_n^B(x_n) \in [0, x_n]$ which satisfy the recurrence relations

$$(17a) \quad g_n(x_n, k_n, m_n) = p_{k_n m_n} g_{n+1}(x_n + v_n^B, k_n + 1, m_n) +$$

$$+ [1 - p_{k_n m_n}] g_{n+1}(x_n - v_n^B, k_n, m_n + 1), \quad 1 \leq n \leq N - 2;$$

$$(17b) \quad g_{N-1}(x_{N-1}, k_{N-1}, m_{N-1}) = p_{k_{N-1} m_{N-1}} \log(x_{N-1} + v_{N-1}^B) +$$

$$+ [1 - p_{k_{N-1} m_{N-1}}] \log(x_{N-1} - v_{N-1}^B),$$

where k_n and m_n denote the number of positive and negative net return rates observed for A prior to period n.

The structure of the Bayesian-dynamic programming solution is extremely simple; namely,

$$(18) \quad v_n^B(x_n) = \begin{cases} [2p_{k_n m_n} - 1] x_n & \text{if} \quad 2p_{k_n m_n} > 1; \\ 0 & \text{if} \quad 2p_{k_n m_n} \leq 1. \end{cases}$$

Consider, now, the criterion filtering approach to the adaptive reinvestment problem with $\hat{p}_n \equiv \hat{p}$, $r_n \equiv 0$, and $s \equiv 1$. Let 'nth period utility' be defined by

$$(19) \quad U(\omega_n, v_n, x_n) \equiv \log(x_n + \omega_n v_n) = \log(x_{n+1}).$$

Given definition (19), it can be shown [21, p. 153] that period-by-period utilities exhibit positive linear correlation in the sense that maximum expected utility in each period $n + 1$ is a positive linear affine function of the

utility realized in period n; hence [21, Thm. 3.3] the original control problem,

$$(20) \qquad \max_{v_1(x_1), \ldots, v_{N-1}(x_{N-1})} \int \ldots \int \log (x_N) p(\omega_{N-1})$$

$$\ldots p(\omega_1) d\omega_{N-1} \ldots d\omega_1$$

and the myopic sequential control problem,

$$(21) \qquad \max_{v_n(x_n) \in [0, x_n]} \int U(\omega_n, v_n(x_n), x_n) p(\omega_n) d\omega_n, \quad 1 \le n \le N - 1,$$

yield identical optimal feasible allocations $(v_1^{\text{opt}}(x_1), \ldots, v_{N-1}^{\text{opt}}(x_{n-1}))$.

The criterion filter adaptive control technique suggested for (21) is as follows:

Period 1. Specificy, for each possible capital state $x \ge 0$, a prior criterion function $U^0(\cdot, x): [0, x] \to R$ for preobservation state-conditioned expected utility evaluation of the feasible capital allocations $v \in [0, x]$. Select an allocation $v_1^* \equiv v_1^*(x_1^*) \in [0, x_1^*]$ to maximize $U^0(\cdot, x_1^*)$, where x_1^* denotes the initial capital endowment for period 1. Finally, observe a net return rate $\omega_1^* \in \{1, -1\}$ and record the new capital state $x_2 = x_1^* + \omega_1^* v_1^*$ for period 2.

Period n(n $\ge$ 2). For each feasible capital allocation $v \in [0, x_n]$, estimate nth period expected utility by

$$(22) \qquad U_n^*(v, x_n) \equiv \left(\frac{\sum_{j=1}^{n-1} U(\omega_j^*, v, x_n) + U^0(v, x_n)}{n} \right).$$

Select an allocation $v_n^* \equiv v_n^*(x_n^*) \in [0, x_n]$ to maximize $U_n^*(\cdot, x_n)$. Finally, observe a net return rate $\omega_n^* \in \{1, -1\}$ and record the new capital state $x_{n+1} = x_n + \omega_n^* v_n^*$ for period $n + 1$.

Thus, for each period $n \ge 1$ and each $v \in [0, x_n]$, the expected utility estimator $U_n^*(v, x_n)$ is obtained by means of a linear filtering operation on a vector $(U(\omega_{n-1}^*, v, x_n), \ldots, U(\omega_1^*, v, x_n), U^0(v, x_n))$ of utility assessments associated with the past observations $\omega_1^*, \ldots, \omega_{n-1}^*$ and the current capital state x_n. As in Section 2, the particular weighting scheme used in (22) results in a criterion filter which is directly analogous to Bayes' rule for updating probability priors.

For direct comparison with the Bayesian-dynamic programming approach, assume the investor specifies barycentric prior criterion functions of the form

$$(23) \qquad U^0(v, x) = p^0 U(1, v, x) + [1 - p^0] U(-1, v, x),$$

where $p^0 \equiv \int_0^1 pG(p)\mathrm{d}p$ is the prior estimator for $p\hat{\ }$ generated by the prior probability density function $G(p)$. Let $p_n^* \equiv [k_n + p^0]/n$, where k_n denotes the number of positive return rates observed for A prior to period n. The nth period expected utility estimator (22) can then be equivalently expressed as

$$(24) \qquad U_n^*(v, x_n) = p_n^* U(1, v, x_n) + [1 - p_n^*] U(-1, v, x_n).$$

The capital allocation which maximizes (24) is easily verified to be

$$(25) \qquad v_n^*(x_n) = \begin{cases} [2p_n^* - 1]x_n & \text{if } 2p_n^* > 1; \\ 0 & \text{if } 2p_n^* \leq 1. \end{cases}$$

Comparing (18) and (25), the form of the feedback control law $(v_n^*(x_n))$ generated by the criterion filtering method is identical to the form of feedback control law $(v_n^B(x_n))$ generated by Bayesian-dynamic programming methods. Suppose, for example, that the prior probability density function $G(p)$ is given by the Beta distribution density function

$$(26) \qquad G(p) = p^{a-1}(1-p)^{b-1}/B(a, b),$$

where $B(a, b)$ is the Beta function with parameters $a > 0$ and $b > 0$, a specification which permits great flexibility in the form of $G(p)$. Then $p^0 = a/[a + b]$ and, letting $\mathrm{d}_n \equiv [n - 1 + a + b]$,

$$(27) \qquad p_{k_n m_n} = ([k_n + a]/\mathrm{d}_n) = (np_n^*/\mathrm{d}_n) + ([a - p^0]/\mathrm{d}_n) \sim p_n^*.$$

Clearly, if $a + b = 1$, the feedback control laws (18) and (25) actually coincide.

In any case, by a strong law argument, $p_n^* \to p\hat{\ }$ a.s. If the Bayesian posterior mean $p_{k_n m_n}$ also converges to $p\hat{\ }$, then $v_n^*(x) \sim v_n^B(x) \sim v_n^{\mathrm{opt}}(x)$ a.s., where $(v_n^{\mathrm{opt}}(x_n))$ is the optimal feedback control law. For this latter result to hold, it is not essential that the prior criterion functions $U^0(v, x)$ have the barycentric form (23). By a strong law argument, $k_n/n \to p\hat{\ }$ a.s.; thus $v_n^*(x) \sim v_n^{\mathrm{opt}}(x)$ a.s. as long as there exist positive constants K' and K'' such that $K' \log(x - v) \leq U^0(v, x) \leq K'' \log(x + v)$ for all $x \in (0, \infty)$ and all $v \in (0, x)$.

Consider, now, the much more difficult problem in which the net return rate probabilities $p_n^\wedge$ are nonstationary. If the probabilities $p_n^\wedge$ are entirely unrelated, then no updating of prior beliefs in any form is beneficial. How can the investor determine whether or not this is, indeed, the case?

To apply standard Bayesian techniques to this inference problem, it is necessary to specify posterior probability density functions

$$(28) \qquad G_n(p \,|\, \omega_{n-1}, \ldots, \omega_1) \propto G_n(\omega_{n-1}, \ldots, \omega_1 \,|\, p) G_n(p),$$

where now the transitional density $G_n(\cdot \,|\, p)$ and prior density $G_n(p)$ are functions of time. There no longer exists a simple sufficient statistic such as (k_n, m_n) which transforms the available sample data in each period n into a vector of fixed dimension. Specification of the posterior densities (28) would thus seem to entail a prohibitive amount of computation. In addition, the specifications $G_n(\cdot \,|\, p)$ and $G_n(p)$ for the statistical model must be regarded as true by the investor; if he is unsure of their form, then in principle he should introduce an additional parameter to index a class of possible statistical models, and he should introduce an additional prior distribution over this parameter.

Even if one accepts, as this author does, the conceptual correctness of the Bayesian approach based on coherency principles, it seems clear that alternate approaches must be explored if practical adaptive control techniques are to be developed. Criterion filtering represents one possible approach to adaptive control which decreases computational complexity while retaining the essence of the Bayesian message: prior and sample information are to be combined to form updated expected utility evaluations over the set of feasible actions. How might criterion filtering be applied to the problem at hand?

The simple criterion filter (22) is no longer adequate. Consider the more general criterion filter

$$(29) \qquad U_n^*(v, x_n) = \left(\sum_{j=1}^{n-1} H_{nj} U(\omega_j^*, v, x_n) + H_{n0} U^0(v, x_n) \right) K_n,$$

where the normalization factor K_n is given by

$$(30) \qquad K_n^{-1} \equiv \left(\sum_{j=1}^{n-1} H_{nj} \right) + H_{n0}.$$

Intuitively, the weight H_{nj} in (29) measures the appropriateness of treating the pseudo utility observation $U(\omega_j^*, v, x_n)$ as if it were an actual observation on the currently relevant utility $U(\omega_n, v, x_n)$. If the probability $\hat{p}_j$ governing the past observation ω_j^* is close to $\hat{p}_n$, then H_{nj} should be close to 1.0; if $\hat{p}_j$ and $\hat{p}_n$ are very dissimilar, then H_{nj} should be close to 0.0.

For example, if the investor believes that the net return rates ω_n^* for investment opportunity A are governed by a seasonally varying distribution with season duration t, i.e.,

(31) $\hat{p}_n = \hat{p}_j$ if $n \equiv j \bmod t$,

then a reasonable filter weight specification for (29) might be

$$(32) \quad H_{nj}^t = \begin{cases} 1 & \text{if} \quad n \equiv j \bmod t \\ 0 & \text{otherwise.} \end{cases}$$

Less conservatively, the investor might specify an exponential weight scheme

(33) $H_{nj}^t = \exp(-\,\mathrm{d}(n, j))$.

where $\mathrm{d}(.,.)$ is a suitable distance function on the positive integers satisfying $\mathrm{d}(n, j) = 0$ if $n \equiv j \bmod t$.

In analogy to the well-known bias/variance trade-off associated with the use of spectral windows in spectral density estimation, it can be shown that use of large-support weight schemes such as (33) in place of small-support weight schemes such as (32) tends to reduce the convergence time and increase the inconsistency of the resulting expected return estimates (29), assuming the underlying specification (31) is correct. How can the investor determine a satisfactory specification for an underlying probability model, and a satisfactory trade-off between convergence time and inconsistency?

One possible approach is as follows. Suppose in period 1 the investor is able to span the possible model structures by a finite collection $\mathscr{H} = \{H^t | t \in T\}$ of distinct filter weight schemes

(34) $H^t = \{H_{nj}^t | 1 \leqq n \leqq N - 1, \quad 0 \leqq j \leqq N - 1\}$

For example, if the investor believes that the net return rates ω_n^* are seasonally varying, with the season duration being either two, three, or four periods, he could specify a collection $\mathscr{H}$ comprising distinct filter weight schemes H^2, H^3, and H^4, each generated as in (32). Assuming such a

collection $\mathcal{H}$ has been specified and a sequence $(\omega_{n-1}, v_{n-1}, \ldots ; , \omega_1, v_1)$ of past return rates and allocations recorded, it is proposed that the investor in period n select a criterion filter weight scheme $H^{\hat{t}}$ for (29) which satisfies the following mean squared error optimization problem:

$$(35) \quad \min_{t \in T} \sum_{k=1}^{n-1} \left[\left(\begin{array}{l} A_k: \text{the actual} \\ \text{realized utility} \\ \text{in period } k \text{ for} \\ \text{allocation } v_k \end{array} \right) - \left(\begin{array}{l} B_k^t: \text{the estimated} \\ \text{expected utility in} \\ \text{period } k \text{ for allo-} \\ \text{cation } v_k \text{ using } H^t \end{array} \right) \right]^2 ,$$

where

$$(36) \quad A_k \equiv U(\omega_k, v_k, x_k)$$

and

$$(37) \quad B_k^t \equiv \left(\sum_{j=1}^{n-1} H_{kj}^t U(\omega_j, v_k, x_k) + H_{k0}^t U^0(v_k, x_k) \right) K_k^t.$$

The suggested approach for selecting and adaptively updating a criterion filter weight scheme for (29) is closely analogous to the usual Bayesian procedure for discriminating among a set of discrete alternative statistical models for $G_n(\cdot | p)$ in (28). The key distinction is that attention has been focused directly on updating the criterion function estimate. In contrast, the Bayesian procedure indirectly updates the criterion function estimate by first updating probabilities.

Finally, consider the general adaptive reinvestment problem with non-stationary net return rate probabilities $p_n^{\hat{}}$ for investment opportunity A and nonstationary deterministic net return rates r_n for investment opportunity B. The appropriate state equation is now

$$(38) \quad x_{n+1} = x_n + \omega_n v_n + r_n [x_n - v_n] \equiv f_n(\omega_n, v_n, x_n);$$

and the investor's optimization problem is no longer equivalent to a sequence of myopic optimization problems. (See [21, Example 5.2].) How might criterion filtering be applied?

Letting $F_n(x)$ denote the maximum attainable expected utility $E \log (x_N)$ beginning in period n with initial capital x, it can be shown [9] that a feasible allocation $(v_1^{\text{opt}}(x_1), \ldots, v_{N-1}^{\text{opt}} (x_{N-1}))$ is optimal if and only if the following dynamic programming optimality equations are satisfied almost surely:

(39a) $F_{N-1}(x_{N-1}) = E_{N-1} \log (f_{N-1}(\omega_{N-1}, v_{N-1}^{\text{opt}}(x_{N-1}), x_{N-1}));$

(39b) $F_n(x_n) = E_n [F_{n+1} \circ f_n(\omega_n, v_n^{\text{opt}}(x_n), x_n)],$

$1 \leqq n \leqq N - 2$, where $E_n[\,\cdot\,]$ denotes expectation with respect to $\{\hat{p_n},$ $1 - \hat{p_n}\}$. The crux of the investor's problem in each period n is to devise a satisfactory estimate $C_n^{n*}(v, x_n)$ for the relevant nth period criterion function

(40) $C_n(v, x_n) \equiv E_n [F_{n+1} \circ f_n(\omega_n, v, x_n)].$

A dynamic criterion-filtering procedure for generating these estimates, developed and tested in [18], is briefly outlined below.

Period 1. For each $n \in \{1, \ldots, N - 1\}$, generate[3] a data set $\{\omega_{-M}(n), \ldots,$ $\omega_0(n)\}$ using a prior probability estimate for $\hat{p_n}$. Compute the prior optimality equation estimates

(41) $F_{N-1}^{1*}(x) \equiv \max_{v \in [0, x]} C_{N-1}^{1*}(v, x)$

$$\equiv \max_{v \in [0, x]} \left(\frac{\sum\limits_{i=0}^{M} \log (f_{N-1}(\omega_{-i}(N - 1), v, x))}{M + 1} \right);$$

$$F_n^{1*}(x) \equiv \max_{v \in [0, x]} C_n^{1*}(v, x)$$

$$\equiv \max_{v \in [0, x]} \left(\frac{\sum\limits_{i=0}^{M} [F_{n+1}^{1*} \circ f_n(\omega_{-i}(n), v, x)]}{M + 1} \right),$$

$x > 0$, $n \in \{1, \ldots, N - 2\}$. Select a control $v_1^* \equiv v_1^*(x_1^*) \in [0, x_1^*]$ for period 1 which satisfies $F_1^{1*}(x_1^*) = C_1^{1*}(v_1^*, x_1^*)$. Record the observation ω_1^* for period 1 and the new state $x_2 = f_1(\omega_1^*, v_1^*, x_1^*)$ for period 2.

Period n ($2 \leqq n \leqq N - 1$). Compute the optimality equation estimates

(42) $F_{N-1}^{n*}(x) \equiv \max_{v \in [0, x]} C_{N-1}^{n*}(v, x)$

$$\equiv \max_{v \in [0, x]} \left(H_{N-1,0} C_{N-1}^{1*}(v, x) + \right.$$

$$\left. + \sum_{j=1}^{n-1} H_{N-1,j} \log (f_{N-1}(\omega_j^*, v, x)) \right) K_{N-1}^n;$$

$$F_k^{n*}(x) \equiv \max_{v \in [0, x]} C_k^{n*}(v, x)$$

$$\equiv \max_{v \in [0, x]} \left(H_{k0} C_k^{1*}(v, x) + \sum_{j=1}^{n-1} H_{kj} F_{k+1}^{n*} {}^\circ f_k(\omega_j^*, v, x) \right) K_k^n,$$

$x > 0$, $k \in \{n, \ldots, N-2\}$. Select a control $v_n^* \equiv v_n^*(x_n) \in [0, x_n]$ for period n which satisfies $F_n^{n*}(x_n) = C_n^{n*}(v_n^*, x_n)$. Record the observation ω_n^* for period n and the new initial state $x_{n+1} = f_n(\omega_n^*, v_n^*, x_n)$ for period $n + 1$.

The filter weights H_{kj} in (42) can be adaptively selected using the general variational principle (35), as before, with return for period n defined to be $\log (x_{n+1})$.

1. DISCUSSION

One important question that must be asked concerns the computational requirements of the dynamic criterion filtering algorithm. Criterion filtering can be classified as an open-loop feedback adaptive control method, i.e., a feedback adaptive control method which directs the controller in each period n to ignore the fact that future observations will be made. (See [13].) Open-loop feedback adaptive control methods have a great computational advantage over closed-loop adaptive control methods, such as Bayesian-dynamic programming, which attempt to take into account both past and potential future observations in each decision period. The principal computational advantage of the criterion filtering method relative to many existing open-loop feedback adaptive control methods is the absence of required integrations. Although the computational burden is by no means eliminated, it is approximately reduced to the level of N deterministic dynamic programming problems.

Other important questions concern the optimality properties of the criterion filtering approach to dynamic adaptive control problems. Namely:

Does the filter-generated control law have the same general structure as the optimal control law?

Do the filter-generated optimality equation estimates have the same general structure as the true underlying optimality equations?

How do the state, control, and utility trajectories realized by use of the filter-generated control law compare to the state, control, and utility trajectories realized by use of the optimal control law?

A detailed analytical and computer simulation study of these equations is carried out in [18] for a dynamic linear-quadratic control problem with random state coefficients. An affirmative answer is provided for the first two questions; and the total utility realized under the filter control law is shown to be approximately on par with the total utility realized under the optimal control law for the tested range of time horizons, utility function coefficients, and mean and standard deviation values for the random state coefficients. Further computer programs are currently being designed to test criterion filtering methods for an adaptive multi-stage team decision problem in which the observations ω_n^* of one team are time and state dependent controls implemented by a competing team acting in accordance with an unknown feedback control law. (See [22].)

University of Southern California

NOTES

[1] Kalman filtering techniques, which are based on the sequential Bayesian updating of the conditional state mean and covariance matrix, have been successfully applied to certain problems in the physical sciences (e.g., satellite trajectory determination) for which a well-understood dynamical state equation can be satisfactorily linearized, the control objective (cost) function can be satisfactorily represented in quadratic form, controls are unconstrained by state variables, and random disturbances can be satisfactorily modeled as white noise second-order processes entering additively into the linearized state equation. (see [11].) These conditions, which imply the applicability of the well-known certainty equivalence theorem to separate the state estimation process from the control process, are often lacking in socioeconomic adaptive control problems. However, see [8] for references to certain applications of Kalman filtering techniques in the field of commercial demand forecasting. Also, see [7, Section 10] for applications of Bayesian techniques to several economic linear-quadratic adaptive control problems

with random state coefficients. Similar applications can also be found in several Special Issues on Control Theory put out by the *Annals of Economic and Social Measurement*.
[2] For example, the procedure suggested by De Finetti [5, page 87] for defining the prevision (expectation) of a random quantity $X(\omega)$ could be used to generate, directly, the prior expected utility specifications $U^0(v)$. Specifically, De Finetti defines a decision maker's prevision $\bar{x}$ for $X(\omega)$ to be the real number $\bar{x}$ he chooses when faced with the following decision: The decision maker will suffer a dollar penalty proportional to $(X(\omega) - \bar{x})^2$ if ω obtains, where he is free to choose $\bar{x}$ as he pleases. For the example at hand, $X(\omega) \equiv U(\omega, v)$.
[3] If actual observations are available for the net return rates ω_n, these may be used in place of or in conjunction with simulated observations. More generally, the prior estimates $C_n^{1*}(v, x)$ can be specified directly, without reliance on actual or simulated data.

REFERENCES

[1] Aoki, M.: 1977, 'Adaptive Control Theory: Survey and Potential Applications to Decision Processes', Invited Paper, presented at the Stochastic Control Workshop, AIDS National Meeting, Chicago.

[2] Arrow, K.: 1965, *Aspects of the Theory of Risk Bearing*, Yrjo Johnsson Foundation, Helsinki.

[3] Barnett, V.: 1973, *Comparative Statistical Inference*, New York, John Wiley & Sons.

[4] Bellman, R. and Kalaba, R.: 1957, 'On the Role of Dynamic Programming in Statistical Communication Theory', IRE *Transactions on Information Theory*, IT-3, 197-203.

[5] De Finetti, B.: 1970, *Theory of Probability*, Vol. 1, New York: John Wiley & Sons.

[6] DeGroot, M.: 1970, *Optimal Statistical Decisions*, New York, McGraw-Hill.

[7] Fienberg, S. and Zellner, A. eds.: 1975, *Studies in Bayesian Econometrics and Statistics*, Amsterdam, North-Holland Publishing Company.

[8] Harrison, P. and Stevens, C.: 1976, 'Bayesian Forecasting', *Journal of the Royal Statistical Society*, Series B, 38, 205-247.

[9] Hinderer, K.: 1970, *Foundations of Non-Stationary Dynamic Programming with Discrete Time Parameter*, New York, Springer Verlag.

[10] Kalaba, R. and Tesfatsion, L.: 1978, 'Two Solution Techniques for Adaptive Reinvestment: A Small Sample Comparison', *Journal of Cybernetics* 8, 101-111.

[11] Kalman, R.: 1978, 'A Retrospective After Twenty Years: From the Pure to the Applied', Invited Paper, Chapman Conference on Applications of the Kalman Filter, Pittsburgh, forthcoming in *Applied Mathematics and Computation*.

[12] Lindley, D.: 1971, *Bayesian Statistics: A Review*, Philadelphia: SIAM.

[13] Saridis, G.: 1977, *Self-Organizing Control of Stochastic Systems*, New York, Marcel Dekker.

[14] Tesfatsion, L.: 1976, 'Bayes' Theorem for Utility', Discussion Paper 76-65, Center for Economic Research, University of Minnesota.

[15] Testfatsion, L.: 1978, 'A New Approach to Filtering and Adaptive Control', *Journal of Optimization Theory and Applications* 25, 247-261.

[16] Tesfatsion, L.: 1978, 'A New Approach to Filtering and Adaptive Control: Stability Results', *Applied Mathematics and Computation* 4, 27–44.

[17] Tesfatsion, L.: 1977, 'A New Approach to Filtering and Adaptive Control: Optimality Results', *Journal of Cybernetics* 7, 133–146.

[18] Tesfatsion, L.: 1979, 'Direct Updating of Intertemporal Criterion Functions for a Class of Adaptive Control Problems', IEEE *Transactions on Systems, Man, and Cybernetics*, SMC-9, 143–151.

[19] Tesfatsion, L.: 1979, 'Criterion Filtering Methods for Adaptive Control', *Proceedings*, 12th Annual Asilomar Conference on Circuits, Systems, and Computers, Pacific Grove, California, IEEE Computer Society, 73–76.

[20] Tesfatsion, L.: 1980a, 'A Conditional Expected Utility Model for Myopic Decision Makers', *Theory and Decision*, 12, 185–206.

[21] Tesfatsion, L.: 1980b, 'Global and Approximate Global Optimality of Myopic Economic Decisions', *Journal of Economic Dynamics and Control* 2, 135–160.

[22] Tesfatsion, L.: 1980c, "C^3 Modeling with Symmetrical Rationality', *Applied Mathematics and Computation*, 6, 51–61.