A BAYESIAN APPROACH TO PROBLEMS IN
STOCHASTIC ESTIMATION AND CONTROL

by

Yu-Chi Ho and Robert C. K. Lee*

June 9, 1964

The research reported in this document was supported by the U. S. Army Research Office, the U. S. Air Force Office of Scientific Research, and the U. S. Office of Naval Research under the Joint Services Electronics Program by Contract Nonr-1866(16), and by the Aeronautical Division of the Minneapolis-Honeywell Regulator Company, Boston, Massachusetts. Reproduction in whole or in part is permitted for any purpose of the United States Government.

*Minneapolis-Honeywell Regulator Co., Boston, Massachusetts

Technical Report No. 448

Cruft Laboratory

Division of Engineering and Applied Physics

Harvard University

Cambridge, Massachusetts
A BAYESIAN APPROACH TO PROBLEMS IN
STOCHASTIC ESTIMATION AND CONTROL*
Yu-Chi Ho**
Harvard University, Cambridge 38, Massachusetts
and
Robert C. K. Lee
Minneapolis-Honeywell Regulator Company
Boston, Massachusetts

SUMMARY
In this paper, a general class of stochastic estimation
and control problems is formulated from the Bayesian De-
cision-Theoretic viewpoint. A discussion as to how these
problems can be solved step-by-step in principle and prac-
tice from this approach is presented. As a specific example,
the closed form Wiener-Kalman solution for linear estima-
tion in gaussian noise is derived. The purpose of the paper
is to show that the Bayesian approach provides: (I) a general
unifying framework within which to pursue further research-
es in stochastic estimation and control problems, (II) the
necessary computations and difficulties that must be over-
come for these problems. An example of nonlinear, non-
gaussian estimation problem is also solved.

SINGLE STAGE ESTIMATION PROBLEM
For purpose of illustrating the concepts involved, the
single stage estimation problem will be discussed first.
Once this is accomplished, the multistage problem can be
treated straightforwardly.

Problem Statement
The following information is assumed given -
(i) A set of measurements z₁, z₂, ..., zₖ which are
denoted by the vector z.
(ii) The physical relationship between the state of
nature which is to be estimated and the measure-
ments. This is given by

\[ z = g(x, v) \]  

where z is the measurement vector (k x 1)
 x is the state (signal) vector (n x 1)
v is the noise vector (q x 1)
(iii) The joint density function p(x, v):
 From this one readily obtains the respective mar-
ginal density functions, p(x) and p(v).

It is assumed that information for (iii) is available in
analytical form or can be approximated by analytical dis-
tributions. Item (ii) can be either in closed form or merely
computable. The problem is to obtain an estimate \( \hat{x} \) of x
and which base upon the measurements is best in some
sense to be defined later.

The Bayesian Solution
The Bayesian solution to the above problem now pro-
ceds via the following steps:

(i) Evaluate p(z) - This can be done analytically, at
least in principle, or experimentally by Monte
Carlo methods since \( z = g(x, v) \) and p(x, v) are
given. In the latter case, we assume it is possible
to fit the experimental distribution again by a mem-
ber of a family of distributions.

(ii) At this point, two alternatives are possible, one
may be superior to the other dependent on the
nature of the problem.

a) Evaluate p(x, z) - This is possible analytically if
v is of the same dimension as z and one can ob-
tain the functional relationship \( v = g^*(x, z) \) from
(I) above. Then using p(x, v) and the
theory of derived distributions, one obtains

\[ p(x, z) = p(x, v = g^*(x, z)) \mid J \]

where

\[ J = \det \left( \frac{\partial g^*(x, z)}{\partial z} \right) \]

b) Evaluate p(z/x). This conditional density func-
tion can always be obtained either analytically
whenever possible or experimentally from the
z = g(x, v) and p(x, v).

Note that (iia) may be difficult to obtain in general
since \( g^* \) may not exist either because of the non-
linear nature of g or that z, v are of different dim-
ensions. Nevertheless, (lib) can always be carried
out. This fact will be demonstrated in the nonlinear
example in the sequel.

(iii) Evaluate p(x/z) using the following relationships,

a) Following (iia)

\[ p(x/z) = \frac{p(x, z)}{p(z)} \]  

b) Following (iib), use the Bayes' rule

\[ p(x/z) = \frac{p(z/x) p(x)}{p(z)} \]

Depending on the class of distributions one has
assumed or obtained for p(x, v), p(z), p(x/z) this
key step may be easy or difficult to carry out.
Several classes of distribution which have nice
properties for this purpose can be found in1. The
density function p(x/z) is known as the aposteriori
density function of x. It is the knowledge about the
state of nature after the measurements z. By def-
inition, it contains all the information necessary
for estimation.

(iv) Depending on the criterion function for estimation
one can compute estimate \( \hat{x} \) from p(x/z).
Some typical examples are:

a) Criterion: Maximize the Probability (\( \hat{x} = x \))

Solution : \( \hat{x} = \text{Mode of } p(x/z) \)  

This is defined as the Most Probable Estimate

---

*The work reported in this paper is supported in part by
NONR Contract (1868116) at Harvard University and by
Minneapolis-Honeywell Regulator Company, Aeronautical
Division, Boston, Massachusetts.

**Consultant at Minneapolis-Honeywell Regulator Company.
When the apriori density function \( p(x) \) is uniform, this estimate is identical to the classical maximum likelihood estimate.

b) Criterion: Minimize \( \int |x - \hat{x}|^2 \ p(x/z) \ dx \)

Solution : \( \hat{x} = E(x/z)^* \)  

This is the conditional mean estimate.

c) Criterion: Minimize Maximum \( |x - \hat{x}| \)

Solution : \( \hat{x} = \text{Medium of } p(x/z) \)

This can be defined as the minimax estimate.

Pictorially, the three estimates are shown in Fig. 1 for a general \( p(x/z) \) for a scalar case.

![Fig. 1 Estimates based on posteriori density](image)

- \( x_a \) - Most Probable Estimate
- \( x_b \) - Conditional mean estimate
- \( x_c \) - Minimax estimate

Fig. 1 Estimates based on posteriori density

Clearly, other estimates, as well as confidence intervals can be derived from \( p(x/z) \) directly.

**Special Case of the Wiener-Kalman Filter (single stage)**

Now a special case of the above estimation problem will be considered. Let there be given

(i) A set of measurements \( z = (z_1, z_2, \ldots, z_k) \)

(ii) The physical relationship

\[ z = H x + v \]  

(iii) The independent noise and state density functions

\( p(x, v) = p(x) p(v) \)

\[ E(x) = \mathbf{X} \]

\[ \text{Cov}(x) = \mathbf{P} \]

\( p(v) \) be gaussian with

\[ E(v) = 0 \]

\[ \text{Cov}(v) = \mathbf{R} \]

Now following the steps for the Bayesian solution one has

(i) Evaluate \( p(z) \).

Since \( z = H x + v \) and \( x, v \) is gaussian and independent, one immediately gets

\( p(z) \) is gaussian

\[ E(z) = H \mathbf{X} \]

\[ \text{Cov}(z) = H \mathbf{P} \mathbf{H}^T + \mathbf{R} \]

(ii) Evaluate \( p(x, z) \).

Since \( \frac{\partial}{\partial x} \) - Identify matrix, it follows

\[ p(x, z) = p(x) p_v(z - Hx) \]

(iii) Evaluate \( p(x/z) \)

\[ p(z) = p(x, z) / p(x) \]

(iv) Now since \( p(x/z) \) is gaussian, the most probable, conditional mean, and minimax estimate all coincide and is given by \( \hat{x} \).

This is the derivation of the single stage Wiener-Kalman filter. The pair \( (\mathbf{P}, \mathbf{K}) \) is called a sufficient statistic for the problem in the sense that \( p(x/z) = p(x|z, \hat{x}) \).

**MULTI-STAGE ESTIMATION PROBLEM**

The problem formulation and the solution in this case is basically similar to the single stage problem. The only additional complication is that now the state is changing from stage to stage according to some dynamic relationship and that the posteriori density function is to be computed recursively.

**Problem Statement**

It is assumed that at any stage \( k = 1 \), the following data is given as a result of previous computation or as part of the problem statement:

*Note: step (iiib) is redundant.

**\( p_v(z - Hx) \) means substituting \( z - Hx \) for \( v \) in \( p(v) \).
The system equations governing the evaluation of the state.

\[ x_{k+1} = f(x_k, w_k) \]
\[ z_{k+1} = h(x_{k+1}, v_{k+1}) \]

where \( x_{k+1} \) is the state vector at \( k+1 \).

\( v_{k+1} \) is the measurement noise at \( k+1 \).

\( z_{k+1} \) is the additional measurement available at \( k+1 \).

\( w_k \) is the disturbance vector at \( k \).

The complete set of measurements \( Z_{k+1} \) is

\[ ( z_1, \ldots, z_{k+1} ) \]

The density functions

\[ p(x_{k+1} / z_k, z_{k+1}) = p(x_{k+1} / z_k) \]

\[ p(w_k, v_{k+1} / x_k) \] statistics of a vector random sequence with components \( w_k \) and \( v_{k+1} \) which depends on \( x_k \).

Now it is required to estimate \( x_{k+1} \) based on measurements \( z_1, \ldots, z_{k+1} \).

The Bayesian Solution

The procedure is analogous to the single stage case.

(i) Evaluate \( p(x_{k+1} / x_k) \). This can be accomplished either experimentally or analytically from knowledge of \( p(w_k, v_{k+1} / x_k), p(x_k / z_k) \) and (21).

(ii) Evaluate \( p(z_{k+1} / x_k, x_{k+1}) \). This is derived from \( p(w_k, v_{k+1} / x_k) \) and (21).

(iii) Evaluate

\[ p(x_{k+1}, z_{k+1} / z_k) = \int p(x_{k+1}, z_{k+1} / z_k) \]
\[ p(x_{k+1} / z_k) \] \( dx_k \)

From this the marginal density functions \( p(x_{k+1} / z_k) \) and \( p(z_{k+1} / z_k) \) can be directly evaluated.

(iv) Evaluate

\[ p(x_{k+1} / z_{k+1}) = \frac{p(x_{k+1}, z_{k+1} / z_k)}{p(z_{k+1} / z_k)} \]

from (22)

\[ \int p(x_{k+1} / z_{k+1}) p(x_{k+1}, z_{k+1} / z_k) dx_k = 1 \]

\[ \int p(x_{k+1} / z_{k+1}) p(x_{k+1}, z_{k+1} / z_k) dx_k \] \( dx_{k+1} \)

Eqn. (24) is a functional integral difference equation governing the evaluation of the a-posteriori density function of the state of (21).

(v) Estimates for \( x_{k+1} \) can now be obtained from \( p(x_{k+1} / z_{k+1}) \) exactly as in the single stage case.

Special Case of the Wiener-Kalman Filter

The given data at \( k+1 \) is specified as follows:

The physical model is given by

\[ x_{k+1} = f(x_k, w_k) \]
\[ z_k = H x_k + v_k \]

where \( w \) and \( v \) are independent, white, gaussian random sequences with

\[ p(x_k / z_k) \] is Gaussian

\[ E(z_k) = H x_k \]
\[ Cov(z_k) = P_k \]

\[ p(w_k, v_{k+1} / x_k, z_k) = p(w_k, v_{k+1}) = p(w_k) p(v_{k+1}) \]
\[ E(v_k) = E(v_{k+1}) = 0 \]
\[ Cov(w_k) = Q; Cov(v_{k+1}) = R \]

Since in this case, the noise \( w_k, v_{k+1} \) is not dependent on the state, Eqn. (24) simplifies to

\[ p(x_{k+1} / z_{k+1}) = \frac{p(x_{k+1} / z_{k+1})}{p(z_{k+1} / z_k)} \]

Hence the solution only involved the evaluation of the three density functions on the r.h.s. of (24) given the data (25 - 27). This is carried out below:

From (27), it is noted that \( p(x_{k+1} / z_k) \) is gaussian and independent of \( v_{k+1} \)

\[ E(x_{k+1} / z_k) = 0 \]
\[ Cov(x_{k+1} / z_k) = P_k \]
\[ E(z_k) = H x_k \]
\[ Cov(z_k) = Q; Cov(v_{k+1}) = R \]

Similarly, \( p(z_{k+1} / z_k) \) is gaussian and

\[ E(z_{k+1} / z_k) = H x_k \]
\[ Cov(z_{k+1} / z_k) = H M_{k+1} H^T + R \]

Finally \( p(x_{k+1} / z_{k+1}) \) is also gaussian with

\[ E(x_{k+1} / z_{k+1}) = H x_k \]
\[ Cov(z_{k+1} / z_{k+1}) = R \]

Footnote added in proof.

This development of the multistage Wiener-Kalman filtering method is very similar to a paper by Drs. H. Rauch, F. Tung, and C. Striebel entitled 'On The Maximum Likelihood Estimate for Linear Dynamic Systems' presented at the SIAM Conference on System Optimization, 1964, Monterey, California. The only difference between the two developments is this: The Rauch-Tung-Striebel paper does not explicitly compute \( p(x, z) \) but simply computes its maximum and uses it as the estimate. In the author's approach, the computation of the maximum plays a secondary role. The explicit calculation of the posteriori probability is emphasized as the Bayesian viewpoint.

The authors are indebted to Prof. A. E. Bryson for bringing this reference to their attention.

*Footnote added in proof.*

The product of the two density functions yields \( p(w_k, v_{k+1} / x_k, z_k) \) by the markov property of (21). It is also assumed that if \( p(w, v / x) = p(w, v) \) then \( w, v \) is a white random sequence.
Combining (28-30) using (24), one gets

\[ p(x_{k+1}/Z_{k+1}) = \frac{|H M_{k+1} H^T + R|^{1/2}}{(2π)^{n/2}|R|^{1/2}} \exp \left\{ \frac{1}{2} \left[ (x_{k+1} - \hat{x}_{k+1})^T H M_{k+1} H^T - R^{-1} (x_{k+1} - H \hat{x}_{k+1}) \right] \right\} \]

Now completing squares in \( \{ \} \), one gets,

\[ p(x_{k+1}/Z_{k+1}) = \frac{|H M_{k+1} H^T + R|^{1/2}}{(2π)^{n/2}|R|^{1/2}} \exp \left\{ -\frac{1}{2} (x_{k+1} - \hat{x}_{k+1})^T H M_{k+1} H^T \right\} \]

where

\[ \hat{x}_{k+1} = \hat{x}_{k} + H \hat{x}_{k} \]

\[ p_{k+1} = M_{k+1} H^T \]

or equivalently

\[ p_{k+1} = M_{k+1} - M_{k+1} H^T (H M_{k+1} H^T + R)^{-1} H M_{k+1} \]

and

\[ M_{k+1} = M_{k} + H \]

Eqns. (33-36) are exactly the discrete Wiener-Kalman filter in the multistage case [3, 4].

A SIMPLE NONLINEAR NONGAUSSIAN
ESTIMATION PROBLEM

The discussions in the above sections have been carried out in terms of continuous density functions. However, it is obvious that the same process can be applied to problems involving discrete density function and discontinuous functional relationships. It is worthwhile, at this point, to carry out one such solution for a simple contrived example which nevertheless illustrates the application of the basic approach.

The problem can be visualized as an abstraction of the following physical estimation problem. An infrared detector followed by a threshold device is used in a satellite to detect hot targets on the ground. However, extraneous signals, particularly reflection from clouds, obscure the measurements. The problem is to design a multistage estimation process to estimate the presence of hot targets on the ground through measurement of the output of the threshold detector.

Let \( s_k \) (target) be a scalar independent Bernoulli process with,

\[ p(s_k) = (1 - q) \delta (s_k) + q \delta (1 - s_k) \]

*The notation \( \delta(x) = \{0 \, x \neq 0\} \) is used here. Also, \( p(x) \) is to be interpreted as mass functions.

For \( n_k \) (cloud noise) be a scalar Markov process with,

\[ p(n_k) = (1-a) \delta (n_k) + a \delta (1 - n_k) \]

\[ p(n_{k+1}/n_k) = (1-a - \frac{n_k}{2}) \delta (n_{k+1}) + (a + \frac{n_k}{2}) \delta (1-n_{k+1}) \]

and the scalar measurement,

\[ z_k = s_k \delta n_k \]

where \( \delta \) indicates the logical "OR" operation.

Essentially Eqns. (37-40) indicate the fact that as the detector sweeps across the field of view, cloud reflection tends to appear in groups while targets appear in isolated dots.

Now we proceed to the Bayesian solution. First, we have,

\[
\begin{array}{cccccc}
\text{Probabil-} & (1-a)(1-q) & q(1-a) & a(1-q) & aq \\
\text{ity of } z_1 & 0 & 1 & 0 & 1 \\
\end{array}
\]

or\begin{align}
p(z_1) &= (1-a)(1-q) \delta (z_1) + (a + q - aq) \delta (z_1 - 1) \quad (41) \\
\end{align}

Also,

\begin{align}
p(z_1/n_1) &= \delta (z_1 - 1) n_1 \\
\end{align}

\begin{align}
p(n_1/z_1) &= \frac{p(z_1/n_1) p(n_1)}{p(z_1)} \\
\end{align}

\begin{align}
&= (1-a'(z_1)) \delta (n_1) + a'(z_1) \delta (n_1 - 1) \\
\end{align}

where

\begin{align}
a'(z_1) &= \frac{a \delta (z_1 - 1)}{(1-a)(1-q) \delta (z_1) + (a + q - aq) \delta (z_1 - 1)} \quad (44) \\
\end{align}

Similarly,

\begin{align}
p(z_1/s_1) &= \delta (z_1 - 1) s_1 + (1-a) \delta (z_1) + a \delta (z_1 - 1) \delta (1-s_1) \\
\end{align}

and

\begin{align}
p(s_1/z_1) &= \frac{p(z_1/s_1) p(s_1)}{p(z_1)} \\
\end{align}

\begin{align}
&= (1 - q'(z_1)) \delta (s_1) + q'(z_1) \delta (s_1 - 1) \\
\end{align}

where

\begin{align}
q'(z_1) &= \frac{q \delta (z_1 - 1)}{(1-a)(1-q) \delta (z_1) + (a + q - aq) \delta (z_1 - 1)} \quad (47) \\
\end{align}

and a reasonable estimate is

\begin{align}
\hat{s}_1 &= 1 \text{ if } q'(z_1) > \epsilon \\
\hat{s}_1 &= 0 \text{ if } q'(z_1) < \epsilon \quad (48) \\
\end{align}

where \( \epsilon \) may be interpreted as an alarm.
Now consider a second measurement \( z_2 \) has been made. One has,

\[
p(n_2/z_1) = \int_{-\infty}^{\infty} p(n_2/n_1) p(n_1/z_1) \, dn_1
\]

(49)

which after straightforward but somewhat laborious manipulations becomes,

\[
a'(z_2) = (1-a - \frac{1}{2}) \delta(n_2) + \frac{a'(z_1)}{2} \delta(n_2 - 1)
\]

\[
\delta(1-a(z_1)) \delta(n_2) + a(z_1) \delta(n_2 - 1)
\]

Furthermore,

\[
p(s_2/z_1) = p(s_2) = (1-q) \delta(s_2) + q \delta(s_2 - 1)
\]

(50)

Equas. (49) and (50) now take the place of (37) and (38) and by the same process, one can get in general,

\[
p(n_k/z_k) = p(n_k/z_{k-1}, z_k-1, \ldots)
\]

(51)

\[
= (1-a^* (Z_k)) \delta(n_k) + a^*(Z_k) \delta(n_k - 1)
\]

\[
a^*(Z_k) \delta a'(z_k, z_k-1, \ldots)
\]

\[
a(Z_{k-1}) \delta (z_{k-1}) = (1-a(Z_{k-1})) (1-q) \delta(z_k) + q a(Z_{k-1}) q \delta(z_{k-1})
\]

(52)

\[
a(Z_{k-1}) \delta a(z_k, z_{k-2}, \ldots)
\]

\[
= a \cdot \frac{a'(Z_{k-1})}{2}
\]

(53)

\[
p(s_k/z_k) = p(s_k/z_{k-1}, z_k-1, \ldots)
\]

(54)

\[
= (1-q') (Z_k) \delta(s_k) + q' (Z_k) \delta(s_k - 1)
\]

\[
q' (Z_k) \delta q'(z_k, z_k-1, \ldots)
\]

\[
q \delta (z_k - 1)
\]

(55)

\[
p(n_{k+1}/z_k) = (1-a(Z_k)) \delta(n_{k+1} - 1) + a(Z_k) \delta(n_{k+1} - 1)
\]

(56)

\[
p(s_{k+1}/z_k) = p(s_k)
\]

(57)

Equas. (51-57) now represent the general recursion solution for the multistage estimation process.

As a check, two possible observed sequences for \( z \), namely \((0,1)\) and \((1,1)\) are considered. When \( a = 1/4 \) and \( q = 1/4 \) it is found that \( p(0/2, z_1) = 0.571 \) and \( 0.337 \) respectively. This agrees with intuition since the sequence \((1,1)\) has a higher probability of being cloud reflections. On the other hand, the numbers also showed that under the circumstances, it is very difficult to detect targets with accuracy using the system contrived here.

Often times one is actually interested in \( p(s_k/Z_k) \) with \( s > 0 \) in order to obtain the so-called "smoothed" estimate for \( s_k \). The desired density function can be computed from \( p(s_k/Z_k) \) by further manipulations. However, the calculation becomes involved and will not be done here.

**RELATIONSHIP TO GENERAL BAYESIAN STATISTICAL DECISION THEORY**

It is worthwhile to point out the relationship of the above formulation and solution of the estimation problem to and its difference from the general statistical decision problem. For simplicity, the single stage case is considered again. In the general statistical decision problem, the input data is somewhat different. One typical form is:

\[
p(x) \quad \text{a priori density of } x
\]

\[
\{e\} \quad \text{a set of choices of experiments from which we can derive measurements } z
\]

\[
p(z/x, e) \quad \text{-conditional density of } z \text{ for given } x \text{ and } e.
\]

\[
\{u\} \quad \text{a set of choices of decisions}
\]

\[J(e, z, u, x)\quad \text{a criterion function which is a possible function of } e, z, u \text{ and } x.
\]

The problem is then stated as the determination of \( e \) and \( u \) so that \( E(J) \) is optimized. The optimal \( J \) is given by

\[
J_{opt} = \max_{e} \{ \min_{u} \int p(z/x, e) \, dz \}
\]

(58)**

\[
p(x/z, e) \, dx \}
\]

\[
p(z/e) \, dz
\]

Thus, the main differences between the estimation problem and the general decision problem are as follows:

(i) In the estimation problem there is no choice of experiment. One always makes the same type of measurement \( z \) given by \( g(x, v) \). To generalize the estimation problem, one can specify:

\[
z_e = g_e(x, v); \{e\} = 1, 2, \ldots \text{ possible sets of measurements}
\]

(59)

and then require that

\[
\bar{E} = \text{Opt} \{ \bar{E} \}_{e = 1, 2, \ldots}
\]

(ii) In the general decision problem, the function \( z = g(x, v) \) is implicit in \( p(z/x, e) \). Hence step (iia) and (iib) for the estimation solution is not required. This often is a tremendous simplification.

(iii) In the estimation problem the criterion function \( J \) is always a simple function of \( x \) only. There is, furthermore, no choice of action (one has to make an estimate by definition). On the other hand, the general decision problem is more analogous to a combined estimation and control problem where one has a further choice of action after determining \( p(x/z) \), and like a control problem, the criteria function is generally more complex.

(iv) It is, however, to be noted that the key step is the computation of \( p(x/z) \) for both problems. The choice of action is determined only after the computation of \( p(x/z) \). Thus, a general decision problem can be decomposed into two problems, namely, determination of \( p(x/z) \) (estimation problem) and choice of action (control problem). In control-theoretic technology, this fact is called the Generalized Decomposition Axiom.

As an example, consider the single stage Wiener-Kalman problem. The added requirement that,

\[
J(e, z, u, x) = \text{J (u, x)} = E \left\| Bx + u \right\|^2
\]

(60)

\[
= \int \left\| Bx + u \right\|^2 \, p(x/z) \, dx
\]

*For other equivalent forms see [1]

**See [1]
be a minimum. Expanding (60), one gets
\[ J = E \left\| x \right\|^2_B + 2 u^T B x + u^T u \] (61)

Clearly,
\[ u_{opt} = u(x(z)) = u(z) = -B \mathbb{E}(z) \] (62)

which is one of the fundamental results of linear stochastic control. Thus, the control action \( u \) is only a function of the criterion \( J \) and the posterior density function \( p(x/z) \). In fact, in this case only \( \mathbb{E}(x) \) of \( p(x/z) \) is needed. We call \( x \) as the minimal sufficient statistic for the control problem.

In the more general multistage case, the decomposition property clearly still holds. The only difference being that \( p(x_{k+1}/z_{k+1}) \) is now dependent on \( u_k \). However, this dependence is entirely deterministic since in a given situation one always knows what \( u_k \) 's are. In fact, in the Wiener-Kalman control problem, it is known that \( u_k \) is a linear function of \( x_k \) only.

CONCLUSION

In the above sections, the problem of estimation from the Bayesian viewpoint is discussed. It is the author's thesis that this approach offers a unifying methodology, at least conceptually, to the general problems of estimation and control.

The posterior conditional density function \( p(x/z) \) is seen to be the key to the solution of the general problem. Difficulties associated with the solution of the general problem now appears more specifically as difficulties in steps leading to the computation of \( p(x/z) \). From the above discussions, it is relatively obvious that these difficulties are:

(i) Computation of \( p(x/z) \) -

In both the single stage or multistage case, this problem is complicated by the nonlinear functional relationships between \( z \) and \( x \). Except in the case when \( z \) and \( x \) are linearly related or when \( z \) and \( x \) are scalars, very little can be done in general, analytically or experimentally. As was mentioned earlier, this difficulty does not appear in the usual decision problem, since there it is assumed that \( p(x/z) \) is given as part of the problem.

(ii) Requirement that \( p(x/z) \) be in analytical form.

This is an obvious requirement if we intend to use the solution in real-time applications. It will not be feasible to compute \( p(x/z) \) after \( z \) has occurred.

(iii) Requirement that \( p(x) \), \( p(z) \), \( p(x/z) \) be conjugate distributions. *

This is simply the requirement that \( p(x) \) and \( p(x/z) \) be density functions from the same family. Note that all the examples discussed in this paper possess this desirable property. This is precisely the reason that multistage computation can be done efficiently. This imposed a further restriction on the functions \( g \), \( f \) and \( h \).

The difficulties (i - iii) listed above are formidable ones. It is not likely that they can be easily circumvented except for special classes of problems, such as those discussed. However, it is worthwhile first to pinpoint these difficulties. Researches toward their solution can then be effectively initiated. Finally, it is felt that the Bayesian approach offers a unified and intuitive viewpoint particularly adaptable to handle modern day control problems where the state and the Markov assumptions play a fundamental role.

*See [1]

REFERENCES


