FINANCIAL FILTERING

Robert J. Elliott

Department of Mathematical Sciences
University of Alberta
Edmonton, Alberta, Canada T6G 2G1
relliott@ualberta.ca

Filtering discusses the optimal recursive estimation of a noisy signal given a sequence of observations:

Questions: Do financial data, price processes, exchange rates, interest rates, contain information about latent variables? If so, how might they and their dynamics be estimated?

1. Change of Measure

A key concept is that of changing the probability measure.

Consider how to change a fair die into an unfair die, (or vice versa), or more generally, how to model changing weights on a die. The experiment of rolling a die has its set of outcomes

$$\Omega_0 = \{1, 2, 3, 4, 5, 6\}.$$

The σ-field $\mathcal{F}_B$ for this experiment is the set of all subsets of Ω, including Ω itself and the empty set $\emptyset$.

If the die is fair, there is a probability P on $\mathcal{F}_B$ such that $P(\{i\}) = \frac{1}{6}$, $1 \le i \le 6$. We consider a more general probability P such that $P(\{i\}) = p_i \ge 0$, with $\sum_{i=1}^{6} p_i = 1$.

On Ω we can also consider the σ-field $\mathcal{F}_E$ generated by the events 'ODD' $= \{1, 3, 5\}$ and 'EVEN' $= \{2, 4, 6\}$.

Then: $$\mathcal{F}_E = \big\{\Omega, \phi, \{1, 3, 5\}, \{2, 4, 6\}\big\}$$

and $$P(\{1, 3, 5\}) = p_1 + p_3 + p_5$$

$$P(\{2, 4, 6\}) = p_2 + p_4 + p_6\,.$$

Suppose $\phi : \Omega \to R$ is a (real) random variable defined on

Ω. The mean, or expected, value of ϕ is then

$$E[\phi] = \sum_{i=1}^{6} \phi(i) p_i.$$

The conditional expected value of ϕ given $\mathcal{F}_E$, that is, given 'ODD' or 'EVEN' is then a function which is constant on the smallest, non-empty subsets (the atoms) of $\mathcal{F}_E$. That is,

$$E[\phi|\mathcal{F}_E](i) = \frac{\big(\phi(1)p_1 + \phi(3)p_3 + \phi(5)p_5\big)}{(p_1 + p_3 + p_5)}, \quad i \in \{1,3,5\}$$

and

$$E[\phi|\mathcal{F}_E](i) = \frac{\big(\phi(2)p_2 + \phi(4)p_4 + \phi(6)p_6\big)}{p_2 + p_4 + p_6}, \quad i \in \{2,4,6\}.$$

Then $E[\phi|\mathcal{F}_E](i)$ can be considered as a function on $(\Omega, \mathcal{F}_B)$ and

$$E\big[E[\phi|\mathcal{F}_E]\big] = E[\phi].$$

Suppose now the die is rebalanced by introducing weights $\Lambda(i) \geq 0$ on the different faces, $1 \leq i \leq 6$. Λ can be considered as a random variable.

Under the new, weighted probability $\overline{P}$ we have

$$\overline{P}(\{i\}) = \overline{p}_i = \Lambda(i)p_i \geq 0, \quad 1 \leq i \leq 6.$$

Then

$$E[\Lambda] = \Lambda(1)p_1 + \Lambda(2)p_2 + \cdots + \Lambda(6)p_6$$

$$= \overline{p}_1 + \overline{p}_2 + \cdots + \overline{p}_6$$

and, because $\overline{P}$ is to be a probability measure,

$$E[\Lambda] = 1.$$

Now:

$$E[\Lambda\phi|\mathcal{F}_E](i) = \frac{\phi(1)\Lambda(1)p_1 + \phi(3)\Lambda(3)p_3 + \phi(5)\Lambda(5)p_5}{p_1 + p_3 + p_5}$$

for $i \in \{1, 3, 5\}$

$$E[\Lambda\phi|\mathcal{F}_E](i) = \frac{\phi(2)\Lambda(2)p_2 + \phi(4)\Lambda(4)p_4 + \phi(6)\Lambda(6)p_6}{p_2 + p_4 + p_6}$$

for $i \in \{2, 4, 6\}$.

Similarly,

$$E[\Lambda|\mathcal{F}_E](i) = \frac{\Lambda(1)p_1 + \Lambda(3)p_3 + \Lambda(5)p_5}{p_1 + p_3 + p_5}\,, \quad \text{for} \quad i \in \{1,3,5\}$$

$$E[\Lambda|\mathcal{F}_E](i) = \frac{\Lambda(2)p_2 + \Lambda(4)p_4 + \Lambda(6)p_6}{p_2 + p_4 + p_6}\,, \quad \text{for} \quad i \in \{2,4,6\}.$$

If $\overline{E}$ denotes expectation with respect to $\overline{P}$, the new, weighted probability,

$$\overline{E}[\phi|\mathcal{F}_E](i) = \frac{\phi(1)\overline{p}_1 + \phi(3)\overline{p}_3 + \phi(5)\overline{p}_5}{\overline{p}_1 + \overline{p}_3 + \overline{p}_5} \quad \text{for} \quad i \in \{1,3,5\}$$

$$\overline{E}[\phi|\mathcal{F}_E](i) = \frac{\phi(2)\overline{p}_2 + \phi(4)\overline{p}_4 + \phi(6)\overline{p}_6}{\overline{p}_2 + \overline{p}_4 + \overline{p}_6} \quad \text{for} \quad i \in \{2,4,6\}\,.$$

We, therefore, observe that

$$\overline{E}[\phi|\mathcal{F}_E] = \frac{E[\Lambda\phi|\mathcal{F}_E]}{E[\Lambda|\mathcal{F}_E]}\,.$$

This is a form of Bayes' theorem. We shall state the general version and give a proof later.

Theorem. *Suppose* $(\Omega, \mathcal{F}, P)$ *is a probability space and* $\mathcal{G} \subset \mathcal{F}$ *is a sub-*σ*-field.*

Suppose $\overline{P}$ *is a second probability measure, absolutely continuous with respect to* P *and with Radon-Nikodym derivative* Λ.

(That is, for $A \in \mathcal{F}$ $\quad \overline{P}(A) = \int_A \Lambda dP$.*)*

Then if ϕ *is any* $\overline{P}$ *integrable random variable*

$$\overline{E}[\phi|\mathcal{G}] = \psi$$

where

$$\psi = \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]} \quad \textit{if} \quad E[\Lambda|\mathcal{G}] > 0$$

$$\psi = 0 \quad \textit{otherwise.}$$

Proof. See HMM page 23.

Discrete Time Markov Chains

We shall consider a process X with time parameter set $\{0, 1, 2, \dots\}$ defined on $(\Omega, \mathcal{F}, P)$. Suppose X has a general finite state space

$$S = \{s_1, s_2, \dots, s_N\}.$$

The points s_i can be elements of any set: the reals numbers, a vector space, $\{$'good', 'bad'$\}$, (for states of the economy).

However, without loss of generality, we can assume the elements of S are identified with the standard unit vectors

$$\{e_1, e_2, \dots, e_N\}, \quad e_i = (0, \dots, 0, 1, 0, \dots, 0)' \in R^N.$$

That is, without loss of generality, we can assume

$$S = \{e_1, e_2, \dots, e_N\}.$$

At each time $k \in \{0, 1, 2, \dots\}$ X_k is, therefore, one of the unit vectors e_i, $1 \le i \le N$. Write $\mathcal{F}_k = \sigma\{X_0, \dots, X_k\}$ for the σ-field generated by $X_0, \dots, X_k$. Then

$$\mathcal{F}_0 \subset \mathcal{F}_1 \subset \cdots \subset \mathcal{F}_k$$

and $\{\mathcal{F}_k\}$ is a filtration which models all possible histories of X.

The Markov property states that

$$P(X_{k+1} = e_j | \mathcal{F}_k) = P(X_{k+1} = e_j | X_k).$$

Write

$$a_{ji} = P(X_{k+1} = e_j | X_k = e_i)$$

$$A = (a_{ji}) \in R^{N \times N}.$$

A is called the transition matrix of the Markov chain X. Then

$$E[X_{k+1}|X_k] = \sum_{i=1}^{N} E[X_{k+1}|X_k = e_i] \langle X_k, e_i \rangle$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} E[\langle X_{k+1}, e_j \rangle | X_k = e_i] \langle X_k, e_i \rangle e_j$$

$$= \sum_{i=1}^{N} \sum_{j=1}^{N} a_{ji} \langle X_k, e_i \rangle e_j = AX_k.$$

Define

$$V_{k+1} := X_{k+1} - AX_k \in R^N.$$

Then $$X_{k+1} = AX_k + V_{k+1}.$$

This is the semimartingale representation of the Markov chain because V is a martingale increment. That is,

$$\begin{aligned} E[V_{k+1}|\mathcal{F}_k] &= E[X_{k+1} - AX_k|X_k] \\ &= E[X_{k+1}|X_k] - AX_k = 0 \in R^N. \end{aligned}$$

The Markov chain is a simple kind of stochastic process on S. However, a more simple process would be one in which X is independently and uniformly distributed over its state space S at each time k.

This is modeled by supposing there is a probability $\overline{P}$ on $(\Omega, \mathcal{F})$ such that at each time k :

$$\overline{P}(X_{k+1} = e_j|\mathcal{F}_k) = \overline{P}(X_{k+1} = e_j) = 1/N.$$

Given such a simple process, and its probability $\overline{P}$, we shall construct a new probability P so that under P X is a Markov chain with transition matrix A.

Note that, if $A = (a_{ji})$ is a transition matrix, then $a_{ji} = P(X_{k+1} = e_j | X_k = e_i) \geq 0$ and

$$\sum_{j=1}^{N} a_{ji} = \sum_{j=1}^{N} P(X_{k+1} = e_j | X_k = e_i) = 1.$$

Suppose $A = (a_{ji})$ is any transition matrix. Suppose (X_k), $k \in \{0, 1, 2, \dots\}$ is a process on the finite state space $\{e_1, \dots, e_N\}$ such that, under a probability $\overline{P}$,

$$\overline{P}(X_{k+1} = e_j | \mathcal{F}_k) = \overline{P}(X_{k+1} = e_j) = 1/N.$$

That is, the probability distribution of X is independent, and uniformly distributed at each time.

Lemma 1. *Define* $\overline{\lambda}_\ell = N \sum_{j=1}^{N} (\langle AX_{\ell-1}, e_j \rangle \langle X_\ell, e_j \rangle)$ *and* $\overline{\Lambda}_k = \prod_{\ell=1}^{k} \overline{\lambda}_\ell$.

A new probability measure P *is defined by putting*

$$\frac{dP}{d\overline{P}}\Big|_{\mathcal{F}_k} = \overline{\Lambda}_k .$$

Then under P X *is a Markov chain with transition matrix* A.

Proof. Note first that

$$\overline{E}[\,\overline{\lambda}_\ell|\mathcal{F}_{\ell-1}] = N\overline{E}\Big[\sum_{j=1}^{N}(\langle\, AX_{\ell-1}, e_j \,\rangle \,\langle\, X_\ell, e_j \,\rangle)|\mathcal{F}_{\ell-1}\Big]$$

$$= N\,\frac{1}{N}\,\sum_{j=1}^{N}\langle\, AX_{\ell-1}, e_j \,\rangle$$

$$= \sum_{i=1}^{N}\,\sum_{j=1}^{N}\langle\, X_{\ell-1}, e_i \,\rangle\, a_{ji} = 1.$$

Then, using the Bayes' theorem

$$P(X_{k+1} = e_j|\mathcal{F}_k) = E[\,\langle\, X_{k+1}, e_j \,\rangle\,|\mathcal{F}_k]$$

$$= \frac{\overline{E}[\overline{\Lambda}_{k+1}\langle\, X_{k+1}, e_j \,\rangle\,|\mathcal{F}_k]}{\overline{E}[\overline{\Lambda}_{k+1}|\mathcal{F}_k]}\ .$$

Because $\overline{\Lambda}_{k+1} = \overline{\Lambda}_k \overline{\lambda}_{k+1}$ and $\overline{\Lambda}_k$ is $\mathcal{F}_k$ measurable this is

$$\frac{\overline{E}[\overline{\lambda}_{k+1} \langle X_{k+1}, e_j \rangle \,|\mathcal{F}_k]}{\overline{E}[\overline{\lambda}_{k+1}|\mathcal{F}_k]} = N\,\overline{E}[\,\langle\, AX_k, e_j\rangle\, \langle\, X_{k+1}, e_j\,\rangle\,|\mathcal{F}_k]$$

$$= \langle\, AX_k, e_j\,\rangle$$

and, as this depends only on X_k this equals

$$P(X_{k+1} = e_j | X_k).$$

If $X_k = e_i$ we see that $P(X_{k+1} = e_j | X_k = e_i) = a_{ji}$ and so, under P X is a Markov chain with transition matrix A.

Remarks. This result gives an explicit construction of the new measure under which an i.i.d. uniform process becomes a Markov chain. A key step is the use of the following conditional form of Bayes' Theorem:

Theorem 2. *Suppose* $(\Omega, \mathcal{F}, P)$ *is a probability space and* $\mathcal{G} \subset \mathcal{F}$ *is a sub* σ*-field of* $\mathcal{F}$*. Suppose* $\overline{P}$ *is a second* *probability measure which is absolutely continuous with respect to P and which has Radon-Nikodym derivative $\frac{d\overline{P}}{dP} = \Lambda$.*

Then if ϕ is any $\overline{P}$ integrable random variable

$$\overline{E}[\phi|\mathcal{G}] = \psi \quad \textit{where}$$

$$\psi = \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]} \quad \textit{if} \quad E[\Lambda|\mathcal{G}] > 0$$

$$\textit{and} \qquad \psi = 0 \quad \textit{otherwise.}$$

Proof. We must show that, if A is any set in $\mathcal{G}$, then

$$\int_A \overline{E}[\phi|\mathcal{G}]d\overline{P} = \int_A \psi d\overline{P}.$$

Write $G = \{w : E[\Lambda|\mathcal{G}] = 0\}$, so $G \in \mathcal{G}$. Then $\int_G E[\Lambda|\mathcal{G}]dP = 0 = \int_G \Lambda dP$. Now $\Lambda \geq 0$ a.s. so either $P(G) = 0$ or the restriction of Λ to G is 0 a.s. In either case $\Lambda = 0$ a.s. on G. Now $G^c = \{w : E[\Lambda|\mathcal{G}] > 0\}$. Suppose $A \in \mathcal{G}$. Then $A = B \cup C$ where $B = A \cap G^c$ and $C = A \cap G$.

Also,

$$\int_A \overline{E}[\phi|\mathcal{G}]d\overline{P} = \int_A \phi d\overline{P} = \int_A \phi\Lambda dP$$
$$= \int_B \phi\Lambda dP + \int_C \phi\Lambda dP.$$

Now $\Lambda = 0$ a.s. on $C \subset G$ so

$$\int_C \phi\Lambda dP = 0 = \int_C \psi d\overline{P} \tag{1}$$

by definition of ψ. Also,

$$\int_B \psi d\overline{P} = \int_B \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}\, d\overline{P}$$
$$= \overline{E}\Big[I_B \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}\Big]$$
$$= E\Big[I_B\Lambda \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}\Big]$$
$$= E\Big[E\big[I_B\Lambda \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}\, |\mathcal{G}\big]\Big]$$
$$= E\Big[E[I_B\Lambda|\mathcal{G}]\, \frac{E[\Lambda\phi|\mathcal{G}]}{E[\Lambda|\mathcal{G}]}\Big]$$
$$= E\big[I_B E[\Lambda\phi|\mathcal{G}]\big]$$

$$= E[I_B \Lambda \phi] = \int_B \Lambda \phi dP$$

$$= \int_B \phi d\overline{P}. \tag{2}$$

Adding (1) and (2)

$$\int_C \Lambda \phi dP + \int_B \Lambda \phi dP = \int_A \Lambda \phi dP$$

$$= \int_A \overline{E}[\phi|\mathcal{G}] d\overline{P} = \int_A \psi d\overline{P}. \qquad \square$$

Summary The state space of a finite state Markov chain can, without loss of generality, be identified with the set $S = \{e_1, e_2, \ldots, e_N\}$ of unit vectors $e_i \in R^N$, $e_i = (0, \ldots, 1, \ldots, 0)'$. The semimartingale representation of the chain is

$$X_{k+1} = AX_k + V_{k+1}$$

$$\text{where} \quad E[V_{k+1}|\mathcal{F}_k] = 0 \in R^N.$$

Suppose we do not observe X directly. Rather there is a function c with values in a finite set and we observe the values in a finite set and we observe the values

$$Y_{k+1} = c(X_k, w_{k+1}), \quad k = 0, 1, 2, \ldots .$$

The $\{w_k\}$ here form a sequence of independent identically distributed random variables which are independent of X.

This provides an example of a Hidden Markov Model (HMM): the Markov chain X is not observed directly but is hidden in the 'noisy' observations Y.

Suppose the range of c consists of M points (in an arbitrary set). As before, these can be identified with the set of unit vectors $\{f_1, f_2, \ldots, f_M\}$ where $f_j = (0, \ldots, 1, \ldots, 0)' \in R^M$.

Recall $\mathcal{F}_k = \sigma\{X_0, \ldots, X_k\}$. Write

$$\mathcal{Y}_k = \sigma\{Y_1, \ldots, Y_k\}$$

$$\text{and} \quad \mathcal{G}_k = \sigma\{X_0, \ldots, X_k, Y_1, \ldots, Y_k\}.$$

Then, for example, $\mathcal{F}_0 \subset \mathcal{F}_1 \subset \mathcal{F}_2 \subset \ldots$ so $\{\mathcal{F}_k\}$, $\{\mathcal{Y}_k\}$

and $\{\mathcal{G}_k\}$ are increasing families of σ-fields. They are the FILTRATIONS representing possible histories of the state process X, the observation process Y and both processes (X, Y). Write

$$c_{ji} = P\big(Y_{k+1} = f_j | X_k = e_i\big), \quad 1 \leq j \leq M, \quad 1 \leq i \leq N.$$

Then $c_{ji} \geq 0$ and, $\sum_{j=1}^{M} c_{ji} = 1$. As before,

$$E[Y_{k+1} | X_k] = C X_k$$

$$\text{where} \quad C = (c_{ji}).$$

Define $W_{k+1} := Y_{k+1} - C X_k$. Then:

$$\begin{aligned} E[W_{k+1} | \mathcal{G}_k] &= E[Y_{k+1} - C X_k | X_k] \\ &= C X_k - C X_k = 0 \in R^M. \end{aligned}$$

Therefore, W is a martingale increment and the semimartingale representation of Y is:

$$Y_{k+1} = C X_k + W_{k+1}.$$

Note we are assuming a unit delay between X_k and its observation Y_{k+1}. Formula for the zero delay case are similar.

Similar to our construction of the Markov chain X we can construct Y by changing a probability measure.

Suppose under some probability measure $\overline{P}$ Y is a process such that

$$\overline{P}\big(Y_{k+1} = f_j | \mathcal{G}_k\big) = \overline{P}\big(Y_{k+1} = f_j\big) = 1/M.$$

Further, under $\overline{P}$ X is a Markov chain which is independent of Y, with state space

$$S = \{e_1, \ldots, e_N\}$$

and transition matrix $A = (a_{ij})$. That is,

$$X_{k+1} = AX_k + V_{k+1}$$

where $\overline{E}[V_{k+1}|\mathcal{G}_k] = \overline{E}[V_{k+1}|\mathcal{F}_k] = \overline{E}[V_{k+1}|X_k] = O \in R^N$.

Suppose $C = (c_{ji})$, $1 \leq j \leq M$, $1 \leq i \leq N$, is a matrix with $c_{ji} \geq 0$ and $\sum_{j=1}^{M} c_{ji} = 1.$

Similar calculations to those of Lemma 1 give the following result:

Lemma 3. *Define*

$$\overline{\lambda}_\ell = M \sum_{j=1}^{M} \left(\langle\, CX_{\ell-1}, f_j \,\rangle \, \langle\, Y_\ell, f_j \rangle \right)$$

$$\textit{and} \qquad \overline{\Lambda}_k = \prod_{\ell=1}^{k} \overline{\lambda}_\ell .$$

A new probability measure P *can be defined by putting* $\left.\frac{dP}{d\overline{P}}\right|_{\mathcal{G}_k} = \overline{\Lambda}_k .$

Then, under P, X *remains a Markov chain with transition matrix* A *and*

$$P\big(Y_{k+1} = f_j \big| X_k = e_i\big) = c_{ji} .$$

That is, under P

$$X_{k+1} = AX_k + V_{k+1}$$

$$\textit{and} \quad Y_{k+1} = CX_k + W_{k+1} .$$

Problem: Suppose we observe $Y_1, \dots, Y_k$. We wish to estimate $X_0, X_1, \dots, X_k$. The best (mean square) estimate of X_k given $\mathcal{Y}_k = \sigma\{Y_1, \dots, Y_k\}$ is

$$E[X_k | \mathcal{Y}_k] \in R^N.$$

However, $\overline{P}$ is a much better measure under which to work. Using the Bayes' result, Theorem 2,

$$E[X_k | \mathcal{Y}_k] = \frac{\overline{E}[\overline{\Lambda}_k X_k | \mathcal{Y}_k]}{\overline{E}[\overline{\Lambda}_k | \mathcal{Y}_k]} .$$

Write $q_k := \overline{E}[\overline{\Lambda}_k X_k | \mathcal{Y}_k] \in R^N$. Then q_k is an unnormalized conditional expectation of X_k given the observations $\mathcal{Y}_k$.

Note $X_k = e_i$ for only one i, $1 \le i \le N$.

Therefore, $\sum_{i=1}^{N} \langle X_k, e_i \rangle = 1$ and with $\mathbf{1} = (1, 1, \dots, 1)' \in$

R^N

$$\begin{aligned}\langle q_k, \mathbf{1}\rangle &= \overline{E}[\,\langle\,\overline{\Lambda}_k X_k, \mathbf{1}\,\rangle\,|\mathcal{Y}_k] \\ &= \overline{E}\,[\,\overline{\Lambda}_k\,\langle X_k, \mathbf{1}\,\rangle\,|\mathcal{Y}_k] \\ &= \overline{E}\,\Big[\,\overline{\Lambda}\Big(\sum_{i=1}^{N}\langle\, X_k, e_i\rangle\Big)\Big|\mathcal{Y}_k\Big] \\ &= \overline{E}\,[\,\overline{\Lambda}_k|\mathcal{Y}_k]\end{aligned}$$

so $$E[X_k|\mathcal{Y}_k] = \frac{q_k}{\langle q_k, \mathbf{1}\,\rangle}\,.$$

We can concentrate on the dynamics of q. The following results show how q is updated as each new observation Y_{k+1} arrives.

Theorem 4. *Write* $B(Y_{k+1})$ *for the diagonal matrix with entries*

$$M\Big(\sum_{j=1}^{M} c_{ji}\,\langle Y_{k+1}, f_j\rangle\Big).$$

Then $q_{k+1} = AB(Y_{k+1})q_k$.

Proof. $q_{k+1} = \overline{E}[\,\overline{\Lambda}_{k+1} X_{k+1} | \mathcal{Y}_{k+1}]$

$$= \overline{E}\left[\overline{\Lambda}_k \Big(M \sum_{j=1}^{N} (\,\langle C X_k, f_j\rangle \,\langle Y_{k+1}, f_j\rangle\,)\Big) (A X_k + V_{k+1}) | \mathcal{Y}_{k+1}\right]$$

Under $\overline{P}$ the Y are i.i.d., and independent of X, and X is a Markov chain with transition matrix A. Therefore, V_{k+1} is a martingale increment, independent of $\mathcal{Y}_{k+1}$, so

$$q_{k+1} = \overline{E}\left[\overline{\Lambda}_k \Big(M \sum_{j=1}^{M} (\,\langle\, C X_k, f_j\rangle \,\langle\, Y_{k+1}, f_j\,\rangle\Big) A X_k | \mathcal{Y}_k\right]$$

$$= M \sum_{i=1}^{N} \overline{E}\left[\,\langle X_k, e_i\rangle\, \overline{\Lambda}_k | \mathcal{Y}_k\right] \Big(\sum_{j=1}^{M} c_{ji} \langle\, Y_{k+1}, f_j\,\rangle\Big) A e_i$$

$$= M \sum_{i=1}^{M} \langle\, \overline{E}\,[\,\overline{\Lambda}_k X_k | \mathcal{Y}_k], e_i\,\rangle\, \Big(\sum_{j=1}^{M} c_{ji}\,\langle\, Y_{k+1}, f_j\,\rangle\,\Big) A e_i$$

$$= M \sum_{i=1}^{M} \langle\, q_k, e_i\rangle\, \Big(\sum_{j=1}^{M} c_{ji}\,\langle\, Y_{k+1}, f_j\rangle\,\Big) A e_i .$$

Write $B(Y_{k+1})$ for the diagonal matrix with entries $M(\sum_{j=1}^{M} c_{ji}\,\langle\, Y_{k+1}, f_j\rangle\,)$. Then

$$q_{k+1} = A B(Y_{k+1}) q_k .$$

This updates the estimate, given Y_{k+1}.

2. Continuous Observations of a Markov Chain

In this section we shall consider the case of continuously valued observations y of a finite state Markov chain X.

The time parameter will still be discrete.

For example X might represent state of the economy: good, bad, average. The observations could be the price of some asset.

Under the real world probability P the Markov chain X has, as in the first lecture, dynamics

$$X_{k+1} = AX_k + V_{k+1}. \tag{2.1}$$

Again we are identifying the state space of the Markov chain with the set of unit vectors $S = \{e_1, \ldots, e_N\}$,

$$e_i = (0, \ldots, 1, \ldots, 0)' \in R^N \quad \text{and} \quad A = (a_{ji})$$

$$\text{where} \quad a_{ji} = P(X_{k+1} = e_j | X_k = e_i).$$

However, the chain X is not observed directly. We suppose there is a (real valued) process y such that

$$y_{k+1} = c(X_k) + \sigma(X_k)\omega_{k+1}. \tag{2.2}$$

Note we are supposing the observation at time $k+1$ depends on the state of X at time k. This one-step delay is reasonable as y may not react immediately to X. However, the alternative zero time-delay model can be treated with minor changes to notation.

The $\omega_1, \omega_2, \dots$ are a sequence of independent, identically distributed $N(0,1)$ random variables. Therefore,

$$P(\omega_k \le \alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx.$$

As before, we shall write

$$\mathcal{F}_k = \sigma\{X_0, X_1, \dots, X_k\}$$

$$\mathcal{Y}_k = \sigma\{y_1, y_2, \dots, y_k\}$$

$$\mathcal{G}_k = \sigma\{X_0, X_1, \dots, X_k, y_1, y_2, \dots, y_k\}.$$

Note that any real function h of X is given by a vector $h = (h_1, \ldots, h_N)' \in R^N$ so that $h(X) = \langle h, X \rangle$. Consequently, there are vectors $c = (c_1, \ldots, c_N)'$ and $\sigma = (\sigma_1, \ldots, \sigma_N)'$ such that $c(X_k) = \langle c, X_k \rangle$ and $\sigma(X_k) = \langle \sigma, X_k \rangle$. We assume that $\sigma_i > 0$ for $1 \leq i \leq N$.

On the underlying space $(\Omega, \mathcal{F})$ and under the 'real world' probability P, therefore, the state process X has dynamics

$$X_{k+1} = AX_k + V_{k+1}$$

where $E[V_{k+1}|\mathcal{G}_k] = E[V_k|X_k] = O \in R^N$

and $$y_{k+1} = c(X_k) + \sigma(X_k)\omega_{k+1}.$$

We choose to work under a Reference Probability $\overline{P}$. Under this probably X is still a Markov chain with transition matrix A. Therefore,

$$X_{k+1} = AX_k + V_{k+1}$$

where $\overline{E}\left[V_{k+1}|\mathcal{G}_k\right] = O \in R^N.$

However, under $\overline{P}$ the $y_1, y_2, \ldots$ are a sequence of i.i.d. random variables each of which is $N(0,1)$. That is, under

$$\overline{P} : \overline{P}(y_k \leq \alpha | \mathcal{G}_{k-1}) = \overline{P}(y_k \leq \alpha) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} e^{-x^2/2} dx.$$

We shall describe how to construct the real world probability P from $\overline{P}$.

Write $\phi(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}$ for the $N(0,1)$ density.

For $\ell = 1, 2, \ldots,$ write

$$\overline{\lambda}_\ell = \frac{\phi\big(\langle \sigma, X_{\ell-1} \rangle^{-1} (y_\ell - \langle c, X_{\ell-1} \rangle)\big)}{\langle \sigma, X_{\ell-1} \rangle \; \phi(y_\ell)} \,,$$

$$\overline{\Lambda}_0 = 1$$

$$\text{and} \quad \overline{\Lambda}_k = \prod_{\ell=1}^{k} \overline{\lambda}_\ell, \quad k \geq 1.$$

Clearly this definition requires that $\langle \sigma, X_k \rangle \neq 0$; the assumption that the observation process has non-singular noise is standard in filtering theory. (If $\langle \sigma, X_k \rangle = 0$ and the components of c are distinct then $y_{k+1} = c_r$ implies

$X_k = e_r$.)

Definition. A probability P is defined by setting

$$\frac{dP}{d\overline{P}}\Big|_{\mathcal{G}_k} = \overline{\Lambda}_k.$$

Lemma. *Under* P *the random variables* $\omega_1, \omega_2, \ldots$ *are a sequence of* i.i.d. $N(0,1)$ *random variables, where*

$$\omega_{k+1} := \langle \sigma, X_k \rangle^{-1} (y_{k+1} - \langle c, X_k \rangle).$$

Proof. $P(\omega_{k+1} \leq \alpha | \mathcal{G}_k)$

$$= E[I(\omega_{k+1} \leq \alpha) \,|\, \mathcal{G}_k]$$

$$= \frac{\overline{E}\,[\,\overline{\Lambda}_{k+1} I(\omega_{k+1} \leq \alpha) \,|\, \mathcal{G}_k]}{\overline{E}\,[\,\overline{\Lambda}_{k+1} \,|\, \mathcal{G}_k]}$$

by the Bayes' Theorem

$$= \frac{\overline{E}\,[\,\overline{\lambda}_{k+1} I(\omega_{k+1} \leq \alpha) \,|\, \mathcal{G}_k]}{\overline{E}\,[\,\overline{\lambda}_{k+1} \,|\, \mathcal{G}_k]}\ .$$

Consider the denominator

$$\overline{E}\,[\,\overline{\lambda}_{k+1}\,|\,\mathcal{G}_k] = \overline{E}\left[\,\frac{\phi\big(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle)\big)}{\langle\sigma, X_k\rangle\,\phi(y_{k+1})}\,\Big|\,\mathcal{G}_k\right]$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\phi\big(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle)\big)}{\langle\sigma, X_k\rangle\,\phi(y_{k+1})}\,\phi(y_{k+1})dy_{k+1}.$$

Write $\qquad \omega := \langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle\sigma, X_k\rangle)$

this is

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\phi(\omega)d\omega = 1.$$

Similarly, the numerator

$$\overline{E}\,[\,\overline{\lambda}\,I(\omega_{k+1} \le \alpha)\,|\,\mathcal{G}_k]$$

$$= \overline{E}\left[\,\frac{\phi\big(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle\,)\big)}{\langle\sigma, X_k\rangle\,\phi(y_{k+1})}\,I(\omega_{k+1} \le \alpha)\,|\,\mathcal{G}_k\right]$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\frac{\phi\big(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle\,)\big)}{\langle\sigma, X_k\rangle\,\phi(y_{k+1})}$$

$$\times\,\phi(y_{k+1})I(\omega_{k+1} \le \alpha)dy_{k+1}$$

$$= \frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\alpha}\phi(\omega)d\omega.$$

Therefore,

$$P(\omega_{k+1} \leq \alpha \mid \mathcal{G}_k) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\alpha} \phi(\omega) d\omega$$
$$= P(\omega_{k+1} \leq \alpha)$$

and the result follows.

Remarks. We wish to estimate X, given the observations under the 'real world' probability P. However, $\overline{P}$ is a nicer measure under which to work. Note that, if

$$\widehat{p}\,^i_k = P[X_k = e_i \mid \mathcal{Y}_k]$$
$$= E[\langle X_k, e_i \rangle \mid \mathcal{Y}_k]$$
$$= E[I(X_k = e_i) \mid \mathcal{Y}_k]$$
$$\text{and} \quad \widehat{p}_k = (\widehat{p}\,^1_k, \ldots, \widehat{p}\,^N_k)$$
$$\text{then} \quad \widehat{p}_k = E[X_k \mid \mathcal{Y}_k]$$

is the conditional distribution of X_k given $\mathcal{Y}_k$.

Using the Bayes' Theorem

$$E[X_k \mid \mathcal{Y}_k] = \frac{\overline{E}[\,\overline{\Lambda}_k X_k \mid \mathcal{Y}_k]}{\overline{E}\,[\,\overline{\Lambda}_k \mid \mathcal{Y}_k]}\ .$$

We write $q_k := \overline{E}\,[\,\overline{\Lambda}_k X_k \mid \mathcal{Y}_k]$ for the unnormalized conditional distribution of X_k given $\mathcal{Y}_k$.

Recall

$$\sum_{i=1}^{N} \langle\, X_k, e_i\, \rangle = 1$$

so

$$\sum_{i=1}^{N} \overline{E}\,[\,\langle\, \overline{\Lambda}_k X_k, e_i \rangle \mid \mathcal{Y}_k] = \overline{E}\,[\,\overline{\Lambda}_k \sum_{i=1}^{N} \langle\, X_k, e_i\, \rangle \mid \mathcal{Y}_k]$$

$$= \overline{E}\,[\,\overline{\Lambda}_k \mid \mathcal{Y}_k] = \sum_{i=1}^{N} \langle\, q_k, e_i\, \rangle$$

$$\text{and} \quad \widehat{p}_k = q_k \Big/ \Big(\sum_{i=1}^{N} \langle\, q_k, e_i\, \rangle \Big).$$

We now derive a recursive filter for q.

Lemma. *Write* $B(y_{k+1})$ *for the diagonal matrix with en-*

tries $\frac{\phi\left(\sigma_i^{-1}(y_{k+1}-c_i)\right)}{\sigma_i\,\phi(y_{k+1})}$. *Then*

$$q_{k+1} = AB(y_{k+1})q_k.$$

Proof. $q_{k+1} = \overline{E}\,[\,\overline{\Lambda}_{k+1}X_{k+1}\,|\,\mathcal{Y}_{k+1}]$

$$= \overline{E}\,[\,\overline{\Lambda}_k\overline{\lambda}_{k+1}(AX_k + V_{k+1})\,|\,\mathcal{Y}_{k+1}]$$

$$= \overline{E}\left[\overline{\Lambda}_k\ \frac{\phi\left(\langle\sigma, X_k\rangle^{-1}(y_{k+1} - \langle c, X_k\rangle)\right)}{\langle\sigma, X_k\rangle\,\phi(y_{k+1})}\ (AX_k + V_{k+1})\,|\,\mathcal{Y}_{k+1}\right]$$

$$= \sum_{i=1}^{N}\overline{E}\,[\,\overline{\Lambda}_k\langle\, X_k, e_i\rangle\,|\,\mathcal{Y}_k]\ \frac{\phi\left(\sigma_i^{-1}(y_{k+1} - c_i)\right)}{\sigma_i\phi(y_{k+1})}\ Ae_i$$

$$= \sum_{i=1}^{N}\langle q_k, e_i\rangle\ \frac{\phi\left(\sigma_i^{-1}(y_{k+1} - c_i)\right)}{\sigma_i\phi(y_{k+1})}\ Ae_i$$

$$= AB(y_{k+1})q_k. \qquad \square$$

Remarks. This is a discrete time form of the Zakai equation.

The parameters of this model are the transition matrix

$A = (a_{ji})$ and the functions (vectors) $c = (c_i)$, $\sigma = (\sigma_i)$.

To estimate these we need estimates of the processes:

$$\begin{aligned} J_k^{rs} &= \sum_{n=1}^{k} \langle\, X_{n-1}, e_r \,\rangle \, \langle\, X_n, e_s \,\rangle \\ &= \text{ the number of jumps from state } \quad r \quad \text{to} \\ &\qquad \text{state} \quad s \quad \text{in time} \quad k. \end{aligned}$$

$$\begin{aligned} O_k^{r} &= \sum_{n=1}^{k} \langle\, X_n, e_r \,\rangle \\ &= \text{ the amount of time} \quad X \quad \text{has spent in state} \quad r \\ &\qquad \text{up to time} \quad k. \end{aligned}$$

We also need an estimate of the process

$$T_k^r(f) = \sum_{n=1}^{k} \langle\, X_{n-1}, e_r \,\rangle \; f(y_n),$$

where $f(y) = y$ or y^2.

Consider, for example,

$$\widehat{J}_k^{rs} = E[J_k^{rs} \,|\, \mathcal{Y}_k].$$

Using the Bayes' theorem this is

$$\overline{E}\,[\,\overline{\Lambda}_k\,J_k^{rs}\,|\,\mathcal{Y}_k]\Big/\,\overline{E}\,[\,\overline{\Lambda}_k\,|\,\mathcal{Y}_k].$$

$\overline{P}$ is an easier measure under which to work. Now it turns out there is not a recursive expression for

$$\overline{E}\,[\,\overline{\Lambda}_k\,J_k^{rs}\,|\,\mathcal{Y}_k].$$

However, if we consider the VECTOR process

$$\overline{E}\,[\,\overline{\Lambda}_k\,J_k^{rs}X_k\,|\,\mathcal{Y}_k]$$

we can find a recursive formula. Then

$$\overline{E}\,[\,\overline{\Lambda}_k\,J_k^{rs}\,|\,\mathcal{Y}_k] = \sum_{i=1}^{N}\overline{E}\,[\,\overline{\Lambda}_k\,J_k^{rs}\,\langle X_k, e_i\rangle\,|\,\mathcal{Y}_k]$$

$$\text{and}\quad E[J_k^{rs}\,|\,\mathcal{Y}_k] = \widehat{J}_k^{\,rs} = \frac{\overline{E}\,[\,\overline{\Lambda}_k\,J_k^{rs}\,|\,\mathcal{Y}_k]}{\langle\, q_k, 1\,\rangle}\,.$$

Notation. For any $\mathcal{G}$-adapted process Z we shall write $\widehat{Z}_k = E[Z|\mathcal{Y}_k]$ and $\sigma(Z)_k$ for $\overline{E}\,[\,\overline{\Lambda}_k Z_k\,|\,\mathcal{Y}_k]$. We shall derive a recursion for $\sigma(J^{rs}X)_k$.

Lemma. *With* $B(y_{k+1})$ *the diagonal matrix with entries* $\frac{\phi\left(\sigma_i^{-1}(y_{k+1}-c_i)\right)}{\sigma_i \phi(y_{k+1})}$

$$\sigma(J^{rs}X)_{k+1} = AB(y_{k+1})\sigma(J^{rs}X)_k$$

$$+ \langle q_k, e_r \rangle \frac{\phi\left(\sigma_r^{-1}(y_{k+1} - c_r)\right)}{\sigma_r \phi(y_{k+1})} a_{sr} e_s.$$

Proof. $\sigma(J^{rs}X)_{k+1} = \overline{E}\,[\,\overline{\Lambda}_{k+1} J^{rs}_{k+1} X_{k+1} \mid \mathcal{Y}_{k+1}]$

$$= \overline{E}\,[\,\overline{\Lambda}_k \overline{\lambda}_{k+1} (J_k^{rs} + \langle\, X_k, e_r \,\rangle \, \langle X_{k+1}, e_s \,\rangle) X_{k+1} \,|, \mathcal{Y}_{k+1}]$$

$$= \overline{E}\left[\overline{\Lambda}_k \, \frac{\phi\left(\langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle)\right)}{\langle \sigma, X_k \rangle \, \phi(y_{k+1})} \, J_k^{rs}(AX_k + V_{k+1}) \mid \mathcal{Y}_{k+1}\right]$$

$$+ \overline{E}\left[\overline{\Lambda}_k \, \frac{\phi\left(\langle \sigma, X_k \rangle^{-1}(y_{k+1} - \langle c, X_k \rangle)\right)}{\langle \sigma, X_k \rangle \, \phi(y_{k+1})}\right.$$

$$\left. \times \langle X_k, e_r \rangle \, \langle AX_k + V_{k+1}, e_s \rangle e_s \,|\mathcal{Y}_{k+1}\right]$$

$$= \sum_{i=1}^{N} \overline{E}\,[\,\overline{\Lambda}_k \, \langle X_k, e_i \rangle \, J_k^{rs} \mid \mathcal{Y}_k] \, \frac{\phi\left(\sigma_i^{-1}(y_{k+1} - c_i)\right)}{\sigma_i \phi(y_{k+1})} \, Ae_i$$

$$+ \overline{E}\,[\,\overline{\Lambda}_k \, \langle X_k, e_r \rangle \mid \mathcal{Y}_k] \, \frac{\phi\left(\sigma_r^{-1}(y_{k+1} - c_r)\right)}{\sigma_r \phi(y_{k+1})} \, a_{sr} e_s$$

$$= \sum_{i=1}^{N} \langle \sigma(J^{rs}X)_k, e_i \rangle \, \frac{\phi\big(\sigma_i^{-1}(y_{k+1} - c_i)\big)}{\sigma_r \phi(y_{k+1})} \, Ae_i$$

$$+ \langle q_k, e_r \rangle \, \frac{\phi\big(\sigma_r^{-1}(y_{k+1} - c_r)\big)}{\sigma_r \phi(y_{k+1})} \, a_{sr} e_s$$

$$= AB(y_{k+1})\sigma(J^{rs}X)_k + \langle q_k, e_r \rangle \, \frac{\phi\big(\sigma_r^{-1}(y_{k+1} - c_r)\big)}{\sigma_r \phi(y_{k+1})} \, a_{sr} e_s$$

where $B(y_{k+1})$ is the matrix with the terms $\frac{\phi\left(\sigma_i^{-1}(y_{k+1}-c_i)\right)}{\sigma_i \phi(y_{k+1})}$ on its diagonal. $\square$

Similar calculations give

Lemma.

$$\sigma(O^r X)_{k+1} = AB(y_{k+1})\sigma(O^r X)_k + \langle \, AB(y_{k+1})q_k, e_r \, \rangle \, er$$

$$\sigma\big(T^r(f)X\big)_{k+1} = AB(y_{k+1})\sigma\big(T^r(f)X\big)_k$$

$$+ \langle \, q_k, e_r \, \rangle \, \frac{\phi\big(\sigma_r^{-1}(y_{k+1} - c_r)\big)}{\sigma_r \phi(y_{k+1})} \, Ae_r.$$

Parameter Estimation

Suppose $\{P_\theta, \theta \in \Theta\}$ is a family of probability measures on a measurable space $(\Omega, \mathcal{F})$, each of which is absolutely continuous with respect to some fixed probability measure P_0. Suppose $\mathcal{Y} \subset \mathcal{F}$.

The likelihood function for computing an estimate of θ based on the information given in $\mathcal{Y}$ is

$$L(\theta) = E_0\left[\frac{dP_\theta}{dP_0}\,\Big|\mathcal{Y}\right].$$

The maximum likelihood estimate of θ is then

$$\widehat{\theta} \in \arg\max_{\theta \in \Theta} L(\theta).$$

The reasoning behind this choice is that the most likely value of θ is the one which maximizes this conditional expectation of the density. However, the MLE is hard to compute. The Expectation Maximization (EM) algorithm provides an alternative approximate method:

Step 1: Set $p = 0$ and choose $\widehat{\theta}_0$.

Step 2: (E-step) Set $\theta^* = \widehat{\theta}_p$ and compute

$$Q(\theta, \theta^*) = E_{\theta^*}\left[\log \frac{dP_\theta}{dP_{\theta^*}} \Big| \mathcal{Y}\right].$$

Step 3: (M-step) Find

$$\widehat{\theta}_{p+1} \in \arg\max_{\theta \in \Theta} Q(\theta, \theta^*).$$

Step 4: Replace p by $p+1$ and repeat from Step 2 until a stopping criterion is satisfied.

The sequence $\{\widehat{\theta}_p\}$ gives non-decreasing values of the likelihood function to a local maximum of the likelihood function. From Jensen's inequality:

$$\log L(\widehat{\theta}_{p+1}) - \log L(\widehat{\theta}_p) \geq Q(\widehat{\theta}_{p+1}, \widehat{\theta}_p)$$

with equality only when $\widehat{\theta}_{p+1} = \widehat{\theta}_p$.

Our model is determined by the parameters

$$\theta := \{a_{ji}, c_i, \sigma_i,\ 1 \leq i, j \leq N\}.$$

Further $a_{ji} \geq 0$, $\sum_{j=1}^{N} a_{ji} = 1$ and $\sigma_i > 0$. Suppose our model is given by such a set of parameters and we wish to determine a new set

$$\widehat{\theta} = \{\widehat{a}_{ji},\ \widehat{c}_i,\ \widehat{\sigma}_i,\ 1 \leq i,\ j \leq N\}$$

which maximizes the analogs of the Q functions.

Consider first the parameter a_{ji}. We recall the form of the change of measure described in the first lecture. Under P_θ X is a Markov chain with transition matrix $A = (a_{ji})$. We wish to introduce a new probability measure $P_{\widehat{\theta}}$ so that under $P_{\widehat{\theta}}$ X is a Markov chain with transition matrix $\widehat{A} = (\widehat{a}_{ji})$. That is, $P_{\widehat{\theta}}(X_{k+1} = e_j | X_k = e_i) = \widehat{a}_{ji}$, so $\widehat{a}_{ji} \geq 0$ and $\sum_{j=1}^{N} \widehat{a}_{ji} = 1$. Define

$$\Lambda_0 = 1$$

$$\Lambda_k = \prod_{\ell=1}^{k} \Big(\sum_{r,s=1}^{N} \left(\frac{\widehat{a}_{sr}}{a_{sr}} \right) \langle X_\ell, e_s \rangle \langle X_{\ell-1}, e_r \rangle \Big).$$

In case $a_{ji} = 0$ take $\widehat{a}_{ji} = 0$ and $\frac{\widehat{a}_{ji}}{a_{ji}} = 1$. Define $P_{\widehat{\theta}}$

by setting $\frac{dP_{\widehat{\theta}}}{dP_\theta}\Big|_{\mathcal{F}_k} = \Lambda_k$.

Lemma. *Under* $P_{\widehat{\theta}}$ X *is a Markov chain with transition matrix* $\widehat{A} = (\widehat{a}_{ji})$.

Proof. $E_{\widehat{\theta}}[\,\langle X_{k+1}, e_s\rangle \,|\, \mathcal{F}_k]$

$$= \frac{E_\theta[\Lambda_{k+1}\,\langle X_{k+1}, e_s\rangle \,|\, \mathcal{F}_k]}{E_\theta[\Lambda_{k+1} \,|\, \mathcal{F}_k]}$$

$$= \frac{E_\theta\Big[\Big(\sum_{r,s=1}^N \big(\frac{\widehat{a}_{sr}}{a_{sr}}\big)\,\langle X_{k+1}, e_s\rangle\,\langle X_k, e_r\rangle\Big)\,\langle X_{k+1}, e_s\rangle \,|\, \mathcal{F}_k\Big]}{E_\theta\Big[\sum_{r,s=1}^N \big(\frac{\widehat{a}_{sr}}{a_{sr}}\big)\,\langle X_{k+1}, e_s\rangle\,\langle X_k, e_r\rangle \,|\, \mathcal{F}_k\Big]}.$$

Now $E_\theta\Big[\sum_{r,s=1}^N \Big(\frac{\widehat{a}_{sr}}{a_{sr}}\Big)\,\langle X_{k+1}, e_s\rangle\,\langle X_k, e_r\rangle \,|\, \mathcal{F}_k\Big]$

$$= \sum_{r=1}^N E_\theta\Big[\sum_{s=1}^N \Big(\frac{\widehat{a}_{sr}}{a_{sr}}\Big)\,\langle X_{k+1}, e_s\rangle \,|\, X_k = e_r\Big]\,\langle X_k, e_r\rangle$$

$$= \sum_{r=1}^N \Big(\sum_{s=1}^N \Big(\frac{\widehat{a}_{sr}}{a_{sr}}\Big)\,a_{sr}\Big)\,\langle X_k, e_r\rangle$$

$$= \sum_{r=1}^N \langle X_k, e_r\rangle = 1.$$

Also,

$$E_\theta\Big[\Big(\sum_{r,s=1}^{N}\Big(\frac{\widehat{a}_{sr}}{a_{sr}}\Big)\langle X_{k+1}, e_s\rangle\,\langle X_k, e_r\rangle\Big)\,\langle X_{k+1}, e_s\rangle \mid \mathcal{F}_k\Big]$$

$$= E_\theta\Big[\sum_{r=1}^{N}\Big(\frac{\widehat{a}_{sr}}{a_{sr}}\Big)\langle X_{k+1}, e_s\rangle\,\langle X_k, e_r\rangle \mid \mathcal{F}_k\Big]$$

$$= \widehat{a}_{sr}\,\langle X_k, e_r\rangle\ .$$

Then:

$$P_{\widehat{\theta}}(X_{k+1} = e_s \mid X_k = e_r) = E_{\widehat{\theta}}[\,\langle X_{k+1}, e_s\rangle \mid X_k = e_r]$$

$$= \widehat{a}_{sr} \qquad \square$$

Lemma. *Given the observations up to time* k, $\{y_1, \dots, y_k\}$, *and given the parameter set* $\theta = \{a_{ji},\ c_i,\ \sigma_i,\ 1 \le i,\ j \le N\}$ *the* EM *estimates* $\widehat{a}_{ji}$ *are given by*

$$\widehat{a}_{ji} = \frac{\widehat{J}^{\,ij}_{k}}{\widehat{O}^{\,i}_{k}} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}\ .$$

Proof. $\left.\dfrac{dP_{\widehat{\theta}}}{dP_\theta}\right|_{\mathcal{F}_k} = \prod_{\ell=1}^{k} \Big(\sum_{r,s=1}^{N} \big(\dfrac{\widehat{a}_{sr}}{a_{sr}}\big) \langle X_\ell, e_s\rangle \langle X_{\ell-1}, e_r\rangle \Big)$

so

$$
\begin{aligned}
\log \frac{dP_{\widehat{\theta}}}{dP_\theta} &= \sum_{\ell=1}^{k} \log \Big(\sum_{r,s=1}^{N} \Big(\frac{\widehat{a}_{sr}}{a_{sr}}\Big) \langle X_\ell, e_s\rangle \langle X_{\ell-1}, e_r\rangle \Big) \\
&= \sum_{\ell=1}^{k} \sum_{r,s=1}^{N} \langle X_\ell, e_s\rangle \langle X_{\ell-1}, e_r\rangle (\log \widehat{a}_{sr} - \log a_{sr}) \\
&= \sum_{r,s=1}^{N} J_k^{rs} \log \widehat{a}_{sr} + R(a),
\end{aligned}
$$

where $R(a)$ is independent of the $\widehat{a}_{sr}$. Therefore

$$
L(\widehat{\theta}) = E\Big[\log \frac{dP_{\widehat{\theta}}}{dP_\theta} \,\Big|\, \mathcal{Y}_k\Big] = \sum_{r,s=1}^{N} \widehat{J}_k^{\,rs} \log \widehat{a}_{sr} + \widehat{R}(a). \tag{6}
$$

Recall

$$
\sum_{s=1}^{N} \widehat{a}_{sr} = 1 \tag{7}
$$

and note

$$\sum_{s=1}^{N} J_k^{rs} = O_k^r,$$

so that $$\sum_{s=1}^{N} \widehat{J}_k^{rs} = \widehat{O}_k^r. \tag{8}$$

The optimal estimate of $\widehat{a}_{ji}$ is, therefore, the value which maximizes the right side of (6) subject to (7). Write λ for the Lagrange multiplier and put

$$L(\widehat{a}, \lambda) = \sum_{r,s=1}^{N} \widehat{J}_k^{rs} \log \widehat{a}_{sr} + \widehat{R}(a) + \lambda\Big(\sum_{s=1}^{N} \widehat{a}_{sr} - 1\Big).$$

Differentiating in $\widehat{a}_{ji}$ and λ and equating the derivatives to 0 gives:

$$\frac{1}{\widehat{a}_{ji}} \widehat{J}_k^{ij} + \lambda = 0 \tag{9}$$

$$\sum_{s=1}^{N} \widehat{a}_{si} = 1. \tag{10}$$

From (8), (9) and (10) we see $\lambda = -\widehat{O}_k^r$ so that $\widehat{a}_{ji} = \frac{\widehat{J}_k^{ij}}{\widehat{O}_k^i} = \frac{\sigma(J^{ij})_k}{\sigma(O^i)_k}$ (as $\widehat{J}_k^{ij} = \frac{\sigma(J^{ij})_k}{\langle q_k, 1\rangle}$, for example). $\square$

Remarks. Consider now the parameter $c = (c_1, \dots, c_N)' \in R^N$. To change this parameter to $\widehat{c} = (\widehat{c}_1, \dots, \widehat{c}_N)'$ we consider factors

$$\lambda^*_{\ell+1} := \exp\big(2^{-1}\langle \sigma, X\rangle^{-1}\{\langle c, X_\ell\rangle^2 - \langle \widehat{c}, X_\ell\rangle^2 \\ - 2y_{\ell+1}\langle c, X_\ell\rangle + 2y_{\ell+1}\langle \widehat{c}, X_\ell\rangle\}\big).$$

Write $\Lambda^*_k = \prod_{\ell=1}^{k} \lambda^*_\ell$ and consider a new measure $P_{\widehat{\theta}}$ defined by setting

$$\frac{dP_{\widehat{\theta}}}{dP_\theta}\Big|_{\mathcal{G}_k} = \Lambda^*_k.$$

Therefore, $L(\widehat{\theta}) = E\Big[\log \dfrac{dP_{\widehat{\theta}}}{dP_\theta}\Big|\mathcal{Y}_k\Big]$

$$= E\Big[\sum_{\ell=1}^{k}(2\langle \sigma, X_{\ell-1}\rangle)^{-1}\{\langle c, X_{\ell-1}\rangle^2 - \langle \widehat{c}, X_{\ell-1}\rangle^2 \\ - 2y_\ell\langle c, X_{\ell-1}\rangle + 2y_\ell\langle \widehat{c}, X_{\ell-1}\rangle\}\,|\mathcal{Y}_k\Big]$$

$$= E\Big[\sum_{\ell=1}^{k}\Big(\sum_{r=1}^{N}\langle X_{\ell-1}, e_r\rangle\Big\{\frac{c_r^2 - \widehat{c}_r^{\,2} - 2y_\ell(c_r - \widehat{c}_r)}{2\sigma_r}\Big\}\,|\mathcal{Y}_k\Big]$$

$$= E\Big[\sum_{r=1}^{N}\frac{2T^r_k(y)\widehat{c}_r - O^r_k\,\widehat{c}_r^{\,2}}{2\sigma_r} + R(c)\,|\,\mathcal{Y}_k\Big]$$

where $R(c)$ is independent of $\widehat{c}$,

$$= \sum_{r=1}^{N} \frac{2\widehat{T}\,{}^{r}_{k}(y)\,\widehat{c}_r - \widehat{O}\,{}^{r}_{k}\,\widehat{c}\,{}^{2}_{r}}{2\sigma_r} + \widehat{R}(c).$$

Differentiating this in $\widehat{c}_i$ and equating the derivative to 0 we see that the optimum choice for $\widehat{c}_i$, given the observations $y_1, \ldots, y_k$, is

$$\widehat{c}_i = \frac{\widehat{T}\,{}^{i}_{k}(y)}{\widehat{O}\,{}^{i}_{k}} = \frac{\sigma\big(T^i(y)\big)_k}{\sigma(O^i)_k}\,. \qquad \square$$

Consider now the parameters σ_i, $1 \le i \le N$. To change the parameters $\sigma = (\sigma_1, \ldots, \sigma_N)$ to $\widehat{\sigma} = (\widehat{\sigma}_1, \ldots, \widehat{\sigma}_N)$ we consider factors

$$\lambda_\ell = \sqrt{\frac{\langle \sigma, X_\ell \rangle}{\langle \widehat{\sigma}, X_\ell \rangle}} \cdot \frac{\exp\Big(-\frac{1}{2\langle \widehat{\sigma}, X_\ell \rangle}\,(y_{\ell+1} - \langle c, X_\ell \rangle)^2\Big)}{\exp\Big(-\frac{1}{2\,\langle \sigma, X_\ell \rangle}\,(y_{\ell+1} - \langle c, X_\ell \rangle)^2\Big)}\,.$$

Write $\Lambda_k = \prod_{\ell=1}^{k} \lambda_\ell$ and define $P_{\widehat{\theta}}$ so that $\frac{dP_{\widehat{\theta}}}{dP_\theta}\Big|_{\mathcal{G}_k} = \Lambda_k$.

Then

$$\log \frac{dP_{\widehat{\theta}}}{dP_\theta} = \sum_{\ell=1}^{k} \Big(- \frac{1}{2} \log \langle \widehat{\sigma}, X_{\ell-1} \rangle$$

$$- \frac{1}{2\langle \sigma, X_{\ell_1} \rangle} (y_\ell - \langle c, X_{\ell-1} \rangle)^2 + R(c,\sigma),$$

where $R(c,\sigma)$ is independent of $\widehat{\sigma}$. Therefore,

$$E\Big[\log \frac{dP_{\widehat{\theta}}}{dP} \Big| \mathcal{Y}_k\Big] = E\Big[\sum_{\ell=1}^{k} \sum_{r=1}^{N} \Big(- \frac{1}{2} \langle X_{\ell-1}, e_r \rangle \log \widehat{\sigma}_r$$

$$- \frac{\langle X_{\ell-1}, e_r \rangle}{2\widehat{\sigma}_r} (y_\ell^2 - 2c_r y_\ell + c_r^2) \,|\mathcal{Y}_k\Big] + \widehat{R}(c,\sigma)$$

$$= -\frac{1}{2} \sum_{r=1}^{N} \Big\{ \log \widehat{\sigma}_r \cdot \widehat{O}\,{}^r_k$$

$$+ \frac{1}{\widehat{\sigma}_r} \big(\widehat{T}\,{}^r_k(y^2) - 2c_r \widehat{T}\,{}^r_k(y) + c_r^2 \,\widehat{O}\,{}^r_k \big) \Big\} + \widehat{R}(c,\sigma).$$

Differentiating in $\widehat{\sigma}_i$ and setting the derivative equal to 0 we see the optimum choice for $\widehat{\sigma}_i$, given the observations

$\mathcal{Y}_k$, is

$$\widehat{\sigma}_i = \big[\,\widehat{T}\,{}^i_k(y^2) - 2c_i\widehat{T}\,{}^i_k(y) + c_i^2\widehat{O}\,{}^i_k\big]/\widehat{O}\,{}^i_k$$

$$= \big[\sigma\big(T^i(y^2)\big)_k - 2c_i\sigma\big(T^i(y)\big)_k\big]/\sigma(O^i)_k + c_i^2. \qquad \square$$

Our results, therefore, provide not only estimates of the Markov chain X but also of the parameters of the model.

Applications to a Price Process

Suppose in discrete time a price process S has the form

$$S_{k+1} = S_k \exp Y_{k+1}$$

where $Y_{k+1} = c_k + \sigma_k b_{k+1}$. Here $\{b_\ell\}$ is a sequence of i.i.d. $N(0,1)$ random variables.

Suppose (c_k, σ_k) takes values in a finite set $B = \{(c_i, \sigma_i) : \; 1 \le i \le N\}$. Write

$$c = (c_1, c_2, \dots, c_N)'$$

$$\sigma = (\sigma_1, \sigma_2, \dots, \sigma_N)'$$

and suppose that (c_k, σ_k) evolves as a Markov chain with state space B. We can identify B with

$$S = \{e_1, e_2, \dots, e_N\}$$

where, as before, $e_i = (0, \dots, 1, \dots, 0)' \in R^N$. Suppose $\phi : B \to S$ gives this bijection, so that for each i, $1 \le i \le N$,

$$\phi\big((c_i, \sigma_i)\big) = e_i.$$

Write $X_k = \phi\big((c_k, \sigma_k)\big)$ (where k now denotes the time parameter. Then

$$c_k = \langle c, X_k \rangle$$

$$\text{and} \quad \sigma_k = \langle \sigma, X_k \rangle .$$

We suppose X is a Markov chain on $(\Omega, \mathcal{F}, P)$ with state space S and transition matrix A.

The state space S could be quite small and X could represent the state of the economy as 'good', 'bad' or 'average'.

Of course, X is not observed directly. Instead we observe logarithmic increments of the price process:

$$\begin{aligned} Y_{k+1} &= \log \frac{S_{k+1}}{S_k} \\ &= c_k + \sigma_k b_{k+1} \\ &= \langle c, X_k \rangle + \langle \sigma, X_k \rangle \, b_{k+1}. \end{aligned}$$

The $N(0,1)$ random variables b model a purely random noise in the dynamics. The Markov chain X also models some random behaviour, but hopefully random behaviour with some structure.

The results of the first part of this talk can now be applied. For any price process (S_k), $k = 0, 1, 2, \ldots,$ the steps are

1) calculate the sequence of logarithmic increments

$$Y_{j+1} = \log \frac{S_{k+1}}{S_k} \,,$$

2) choose 'appropriate' prior values for $\{(c_i, \sigma_i),\ 1 \leq i \leq N\}$

and for the transition probabilities

$$a_{ji} = P(X_{k+1} = e_j \mid X_k = e_i) \geq 0,$$

with $\sum_{j=1}^{N} a_{ji} = 1.$

3) after n values of Y have been observed, calculate new estimates for c, σ and the a.

4) use these values, iteratively, to re-estimate the c, σ and a. The EM algorithm implies the estimates improve monotonically, in the sense that the expected log-likelihood increases with each re estimation. Consequently, the model is 'self-tuning'. This step is repeated until some stopping criterion is satisfied.

Examples

In the first example, the program was run on a data set consisting of 248 monthly observations on the price of IBM stock. The sample period was from June 1975 to January 1996. The prices were processed in eight groups of 31 prices each; at the end of each pass through the data, parameter estimates were updated using the formulas given in the paper. In the analysis, the size, N, of the state space of the Markov chain was taken to be four.

Table 1 gives the initial values that were assumed for the conditional distribution of the state of the Markov chain, that is, for $E[X_n, Y_n]$, the transition matrix A, and the vectors c and σ. It also gives the re-estimated values of these parameters after the sixth and seventh passes.

To assess the predictive performance of the model, predicted prices for the period February 1991 to June 1993 were calculated using the formula:

$$E[S_{k+1}|Y_k] = S_k \sum_{i=1}^{N} e^{c_i} e^{\sigma i^2/2} \frac{\langle q_k, e_i \rangle}{\langle q_k, \mathbf{1} \rangle} ,$$

$$k = 1, 2, \dots .$$

The formula was evaluated using estimated values of the vectors c, σ, and q_k

$(k = 1, 2, \dots)$ after the seventh pass through the data. We then regressed the actual prices for the period on the predicted prices according to the following model:

$$\text{Actual price} = \alpha + \beta^{*}\text{Predicted price } + \varepsilon.$$

The regression results obtained were assessed on the basis of the 3 criteria for a good model proposed by Fama and Gibbons (1984): (1) conditional unbiasedness, that is, an intercept, α, close to zero, and a regression coefficient, β, close to one; (2) serially uncorrelated residuals; and (3) a low residual standard error. The second column of Table 2 gives the results for IBM stock. These results indicate that at the significance level of 0.05, we may conclude that the intercept, α, equals zero (although its standard error is quite large) and that the regression coefficient, β, equals one. Also, on the basis of the Durbin-Watson test, we may conclude that the residuals do not display first-order serial correlation.

In the second example, a data set of 100 monthly observations on the price of gold was analyzed. In this case, the sample period was January 1988 to April 1996. The prices were processed in four groups of 25 prices each. Again, the size, N, of the state space of the Markov chain was taken to be four.

Table 3 gives the initial and re-estimated values of the model parameters.

We again regressed actual prices for the period May 1992 to March 1994 on the corresponding predicted prices, according to the model stated earlier. Predicted prices were calculated using the estimated values of the g, γ, and q_k vectors after the third pass though the data. The third column of Table 2 gives the regression results for gold. They support the conclusions that the intercept is zero, the slope is one, and residuals are serially uncorrelated.

pageno=56 **3. Discrete Time Non-Linear Filters**

Our processes are defined on a probability space $(\Omega, \mathcal{F}, P)$.

$\{x_\ell\}$, $\ell = 0, 1, 2, \ldots$ is a process with values in some Euclidean space R^d.

$\{w_\ell\}$, $\ell = 0, 1, 2, \ldots$ is a sequence of i.i.d. R^k valued random variables.

Suppose the density function of the w is ψ and that $F : R^d \times R^k \to R^d$ is a measurable function describing the dynamics of x, so that

$$x_{k+1} = F(x_k, \omega_{k+1}).$$

Suppose that x_0 or its density $\pi_0(x)$ is known. Finally, we assume there is an inverse map: $D : R^d \times E^d \to R^k$ such that, if

$$x_{k+1} = F(x_k, \omega_{k+1})$$

$$\text{then} \quad \omega_{k+1} = D(x_{k+1}, x_k).$$

Note this condition is satisfied if the noise ω is additive, so that

$$x_{k+1} = F(x_k) + \omega_{k+1}.$$

We require D to be differentiable in its first variable.

Observations. We suppose the signal process x is not observed directly. Rather there is an observation process $\{y_\ell\}$, $\ell = 0, 1, 2, \ldots$ related to the signal.

We suppose the observation process y is real valued. The extension to vector valued observations is straight forward.

Suppose $\{b_\ell\}$ is a sequence of real i.i.d random variables having a density function ϕ for which $\phi(b) > 0$ for all $b \in R$.

$H : R^d \times R \to R$ is measurable. We suppose that

$$y_k = H(x_k, b_k).$$

Again, we shall suppose there is an inverse function G such that if

$$y_k = H(x_k, b_k)$$

$$\text{then} \quad b_k = G(y_k, x_k).$$

This condition is clearly satisfied in the case of additive noise, when

$$y_k = H(x_k) + b_k.$$

We require the derivatives $\frac{\partial H}{\partial x}$ $\frac{\partial G}{\partial y}$ to be non-singular.

In the real world, therefore, the SIGNAL is given by: $x_{k+1} = F(x_k, \omega_{k+1})$ and the OBSERVATIONS are given by: $y_k = H(x_k, b_k)$.

These are defined on $(\Omega, \mathcal{F}, P)$. Consider the filtrations $\{\mathcal{F}_t\}$, $\{\mathcal{G}_t\}$, $\{\mathcal{Y}_t\}$ where

$$\mathcal{F}_t = \sigma\{x_0, x_1, \ldots, x_t\}$$

$$\mathcal{G}_t = \sigma\{x_0, x_1, \ldots, x_t, y_0, y_1, \ldots, y_{t-1}\}$$

$$\mathcal{Y}_t = \sigma\{y_0, y_1, \ldots, y_t\}.$$

Reference Probability: Suppose we have a second probability measure $\overline{P}$ defined on $(\Omega, \mathcal{G}_\infty)$, where $\mathcal{G}_\infty = \bigvee_{t=1}^{\infty} \mathcal{G}_t$, such that, under $\overline{P}$,

1) the $\{x_\ell\}$ are a sequence of random variables satisfying

$$x_\ell = F(x_{\ell-1}, \omega_\ell)$$

   where the ω_ℓ are i.i.d. random variables with densities ψ, and

2) the $\{y_\ell\}$ are i.i.d. random variables with density function ϕ.

In particular, under $\overline{P}$ the y_ℓ are independent of the x_ℓ.

We shall construct a probability measure P on $(\Omega, \mathcal{G}_\infty)$ such that, under P,

$$b_k := G(y_k, x_k), \quad k = 1, 2, \ldots,$$

is a sequence of independent random variables with positive densities $\phi(b)$.

Consequently, it is under P that the signal and observation are related by the 'real world' dynamics. Write:

$$\lambda_k(x_k, y_k) = \frac{\phi(b_k)}{\phi(y_k)} \cdot \Big(\frac{\partial H}{\partial b}(x_k, b_k)\Big)^{-1}$$

$$\Lambda_0 = 1$$

$$\Lambda_k = \prod_{\ell=1}^{k} \lambda_\ell(x_\ell, y_\ell).$$

P is defined by setting

$$\frac{dP}{d\overline{P}}\Big|_{\mathcal{G}_k} = \Lambda_k.$$

(The existence of P on $\mathcal{G}_\infty$ follows by Kolmogorov's extension theorem.)

Recall $b_k = G(y_k, x_k)$, so that $y_k = H(x_k, b_k)$.

Lemma. *Under* P *the* b_k, $k = 1, 2, \ldots,$ *are* i.i.d. *random variables with density* ϕ.

Proof. $P(b_k \le t | \mathcal{G}_k) = E[I(b_k \le t) | \mathcal{G}_k]$

$$= \frac{\overline{E}[\Lambda_k I(b_k \le t) | \mathcal{G}_k]}{\overline{E}[\Lambda_k | \mathcal{G}_k]}$$

$$= \frac{\overline{E}[\lambda_k I(b_k \le t) | \mathcal{G}_k]}{\overline{E}[\lambda_k | \mathcal{G}_k]} \; .$$

Now

$$\overline{E}[\lambda_k | \mathcal{G}_k] = \overline{E}\Big[\frac{\phi(b_k)}{\phi(y_k)} \cdot \Big(\frac{\partial H}{\partial b} \, (x_k, b_k) \Big)^{-1} \Big| \mathcal{G}_k \Big]$$

with $b_k = G(y_k, x_k)$. The only unknown in this conditioning is y_k so

$$\overline{E}[\lambda_k | \mathcal{G}_k] = \int_{-\infty}^{\infty} \phi(y_k) \, \frac{\phi\big(G(y_k, x_k)\big)}{\phi(y_k)} \, \Big(\frac{\partial H}{\partial b} \, \big(x_k, G(y_k, x_k)\big) \Big)^{-1} dy_k .$$

Write $b = G(y_k, x_k)$ so that

$$y_k = H(x_k, b)$$

$$\text{and} \quad dy_k = \frac{\partial H}{\partial b} \, (x_k, b) db.$$

Then:

$$\overline{E}[\lambda_k|\mathcal{G}_k] = \int_{-\infty}^{\infty} \phi(b)db = 1.$$

Further,

$$\begin{aligned}
&\overline{E}[\lambda_k I(b_k \le t)|\mathcal{G}_k] \\
&\quad = \int_{-\infty}^{\infty} \phi\big(G(y_k, x_k)\big) \left(\frac{\partial H}{\partial b}\big(x_k, G(y_k, x_k)\big)\right)^{-1} I(b \le t)db \\
&\quad = \int_{-\infty}^{t} \phi(b)db.
\end{aligned}$$

That is,

$$\begin{aligned}
P(b_k \le t|\mathcal{G}_k) &= \int_{-\infty}^{t} \phi(b)db \\
&= P(b_k \le t). \qquad \square
\end{aligned}$$

Recursive Estimates: We shall work under $\overline{P}$, so that, under $\overline{P}$ $\{y_k\}$ is a sequence of i.i.d. random variables with densities ϕ and $\{x_k\}$ is a sequence of random variables with dynamics

$$x_k = F(x_{k-1}, \omega_k).$$

Here $\{w_k\}$ is a sequence of random variables with densities ψ. We shall again use the form of Bayes' theorem which states that, for a $\mathcal{G}$-adapted process $\{f_\ell\}$,

$$E[f_k|\mathcal{Y}_k] = \frac{\overline{E}[\Lambda_k f_k|\mathcal{Y}_k]}{\overline{E}[\Lambda_k|\mathcal{Y}_k]}\ .$$

We shall investigate the unnormalized conditional expectation

$$\overline{E}[\Lambda_k f_k|\mathcal{Y}_k].$$

Notation. Write $q_k(z)$, $k = 0, 1, 2, \dots,$ for the unnormalized conditional density such that

$$\overline{E}[\Lambda_k I(x_k \in dz)|\mathcal{Y}_k] = q_k(z)dz.$$

Then
$$\overline{E}[\Lambda_k f(x_k)|\mathcal{Y}_k] = \int_{R^d} f(z)q_k(z)dz.$$

We now derive a recursive update for q. Note the normalized conditional density

$$p_k(z)dz = E[I(x_k \in dz)|\mathcal{Y}_k]$$

is given by $p_k(z) = q_k(z)/(\int_{R^d} q_k(\xi)d\xi)$.

Theorem. *For* $k = 0, 1, 2, \ldots$

$$q_k(\xi) = \phi(y_k)^{-1}\phi\big(G(y_k, \xi)\big)\Big(\frac{\partial H}{\partial b}\,\big(\xi, G(y_k, \xi)\big)\Big)^{-1} \times \int_{R^d} \psi\big(D(\xi, z)\big)\,\frac{\partial D}{\partial \xi}\,(\xi, z)q_{k-1}(z)d\xi.$$

Suppose $f : R^d \to R$ is any Borel 'test' function. Then

$$\overline{E}[f(x_k)\Lambda_k|\mathcal{Y}_k] = \int_{R^d} f(\xi)q_k(\xi)d\xi$$

$$= \overline{E}\Big[f\big(F(x_{k-1},\omega_k)\big)\Lambda_{k-1}\phi\big(G(y_k,x_k)\big)$$

$$\times \Big(\frac{\partial H}{\partial b}\,\big(x_k, G(y_k,x_k)\big)\Big)^{-1}\Big|\mathcal{Y}_k\Big]\,\phi(y_k)^{-1}$$

$$= \overline{E}\Big[\int_{R^k} f\big(F(x_{k-1},\omega)\big)\Lambda_{k-1}\phi\Big(G\big(y_k, F(x_{k-1},\omega)\big)\Big)$$

$$\times \Big(\frac{\partial H}{\partial b}\,\big(F(x_{k-1},\omega), G\big(y_k, F(x_{k-1},\omega)\big)\big)\Big)^{-1}\psi(\omega)d\omega\Big|\mathcal{Y}_k\Big]\,\phi(y_k)^{-1}$$

$$= \int_{R^d}\int_{R^k} f\big(F(z,\omega)\big)\phi\Big(G\big(y_k, F(z,\omega)\big)\Big)$$

$$\times \Big(\frac{\partial H}{\partial b}\,\big(F(z,\omega), G\big(y_k, F(z,\omega)\big)\big)\Big)^{-1}\psi(\omega)q_{k-1}(z)d\omega dz\,\phi(y_k)^{-1}.$$

Write $\xi = F(z,\omega)$. Then $\omega = D(\xi, z)$ and

$$d\omega dz = \left|\det \begin{vmatrix} \frac{\partial\omega}{\partial\xi} & \frac{\partial\omega}{\partial z} \\ \frac{\partial z}{\partial\xi} & \frac{\partial z}{\partial z} \end{vmatrix}\right| d\xi dz$$

$$= \left|\frac{\partial D}{\partial\xi}\,(\xi, z)\right| d\xi dz$$

$$\text{where} \quad \left|\frac{\partial D}{\partial\xi}\right| = \left|\det\frac{\partial D}{\partial\xi}\right|.$$

Therefore

$$\int_{R^d} f(\xi) q_k(\xi) d\xi$$

$$= \int_{R^d} \int_{R^k} f(\xi) \phi\big(G(y_k, \xi)\big) \Big(\frac{\partial H}{\partial b} \left(\xi, G(y_k, \xi)\right) \Big)^{-1}$$

$$\times \, \psi\big(D(\xi, z)\big) q_{k-1}(z) \left| \frac{\partial D}{\partial \xi} (\xi, z) \right| d\xi dz \; \phi(y_k)^{-1}.$$

This holds for all test functions f, so

$$q_k(\xi) = \phi(y_k)^{-1} \phi\big(G(y_k, \xi)\big)$$

$$\times \Big(\frac{\partial H}{\partial b} \left(\xi, G(y_k, \xi)\right) \Big)^{-1} \int_{R^d} \psi\big(D(\xi, z)\big) \left| \frac{\partial D}{\partial \xi} (\xi, z) \right| q_{k-1}(z) dz.$$

$\square$

Remarks. If $\pi(z)$ is the density of x_0 so that $P(x_0 \in A) = \overline{P}(x_0 \in A) = \int_A \pi(z) dz$ then $q_0(z) = \pi(z)$ and updated estimates are given by the theorem.

Even if x_0 is known, so that $\pi(z)$ is replaced by a delta function, the theorem gives the following unnormalized conditional density for

$$q_1(z) = \phi(y_1)^{-1}\phi\big(G(y,\xi)\big)\Big(\frac{\partial H}{\partial b}\,(\xi, G(y_1,\xi))\Big)^{-1}$$
$$\times\,\psi\big(D(\xi,x_0)\big)\Big|\frac{\partial D}{\partial \xi}\,(\xi,x_0)\Big|.$$

The Kalman Filter

The Kalman filter discusses the case where the dynamics of the signal and observation are linear and the noise processes are Gaussian. That is, consider the case where

$$x_k = Ax_{k-1} + \omega_k \in R^d$$

$$\text{and} \quad y_k = Cx_k + b_k \in R^m.$$

(We shall consider vector observations $y \in R^m$). A and C are matrices of appropriate dimensions. (They could be taken to be time-varying.) The ω_k, b_k and x_0 are normally distributed with means 0 and Gaussian densities

$$(2\pi)^{-d/2}|Q|^{-1/2}\exp\Big(-\frac{1}{2}\,\omega' Q^{-1}\omega\Big)$$

$$(2\pi)^{-m/2}|R|^{-1/2}\exp\Big(-\frac{1}{2}\,b' R^{-1} b\Big)$$

$$\text{and}\quad (2\pi)^{-d/2}|\Sigma_{0|0}|^{-1/2}\exp\Big(-\frac{1}{2}\,x'(\Sigma_{0|0}^{-1})x\Big)$$

respectively.

Suppose at time k $q_k(x)$ is a constant times a Gaussian density of mean μ_k and variance $\sum_k$.

That is, the unnormalized conditional density of x_k given $\mathcal{Y}_k$ is

$$q_k(x) = c\exp\Big(-\frac{1}{2}\,(x-\mu_k)'\big(\Sigma_k\big)^{-1}(x-\mu_k)\Big).$$

As

$$y_k = Cx_k + b_k = H(x_k, b_k)$$

we see $\frac{\partial H}{\partial b}$ is the identity.

Similarly, because $\omega_k = x_k - Ax_{k-1} = D(x_k, x_{k-1})$

$\dfrac{\partial D}{\partial \xi}$ is the identity.

Substituting in the recursion for q_k given in the Theorem we have

$$q_{k+1}(x) = K(x) \int_{R^d} \exp\Big[- \frac{1}{2} \Big\{ (x - A\xi)' Q^{-1} (x - A\xi) + (\xi - \mu_k)' \Big(\Sigma_k \Big)^{-1} (\xi - \mu_k) \Big\} \Big] d\xi$$

where

$$K(x) = \frac{\phi(y_{k+1} - Cx)}{\phi(y_{k+1})} (2\pi)^{-m/2} |Q|^{-1/2} \Big| \Sigma_k \Big|^{-1/2}.$$

Keeping only terms in ξ under the integration we see

$$q_{k+1}(x) = K_1(x) \int_{R^d} \exp\Big[- \frac{1}{2} (\xi' \sigma_{k+1}^{-1} \xi - \beta'_{k+1} \xi) \Big] d\xi$$

where

$$K_1(x) = K(x)\exp\Big[-\frac{1}{2}\Big(x'Q^{-1}x + \mu_k'\Sigma_k^{-1}\mu_k\Big)\Big]$$

$$\sigma_{k+1}^{-1} = A'Q^{-1}A + \Sigma_k^{-1}$$

$$\text{and} \quad \beta_{k+1}' = 2\big(x'Q^{-1}A + \mu_k'\Sigma_k^{-1}\big).$$

Completing the square in the integral

$$q_{k+1}(x) = K_1(x)\exp\Big[-\tfrac{1}{2}\Big(-\ \tfrac{\beta_{k+1}'\sigma_{k+1}\beta_{k+1}}{4}\Big)\Big]$$

$$\times \int_{R^d} \exp\Big[-\tfrac{1}{2}\Big(\xi - \tfrac{\sigma_{k+1}\beta_{k+1}}{2}\Big)'\sigma_{k+1}^{-1}\Big(\xi - \tfrac{\sigma_{k+1}\beta_{k+1}}{2}\Big)\Big]d\xi$$

$$= K_1(x)\exp\Big[\tfrac{1}{2}\Big(\tfrac{\beta_{k+1}'\sigma_{k+1}\beta_{k+1}}{4}\Big)\Big]|\sigma_{k+1}|^{-1/2}(2\pi)^{m/2}$$

$$= \phi(y_{k+1})^{-1}|Q|^{-1/2}|\Sigma_k|^{-1/2}|\sigma_{k+1}|^{-1/2}$$

$$\times \exp\Big[-\tfrac{1}{2}\Big\{\big(y_{k+1} - Cx\big)'R^{-1}(y_{k+1} - Cx) + x'Q^{-1}x$$

$$+\mu_k'\Sigma_k^{-1}\mu_k - (x'Q^{-1}A + \mu_k'\Sigma_k^{-1})\sigma_{k+1}\big(\Sigma_k^{-1}\mu_k + A'Q^{-1}x)\Big\}\Big]$$

$$= M(y_{k+1}, \Sigma_k, \mu_k)\exp\Big[-\ \tfrac{1}{2}\big\{x'(C'R^{-1}C + Q^{-1}$$

$$-Q^{-1}A\sigma_{k+1}A'Q^{-1}\big)x - 2x'\big(C'R^{-1}y_{k+1} + Q^{-1}A\sigma_{k+1}\Sigma_k^{-1}\mu_k\big)\Big].$$

Completing the square in the exponential we see that $q_{k+1}(x)$ is an unnormalized Gaussian density with mean μ_{k+1} and variance Σ_{k+1} where

$$\mu_{k+1} = E[X_{k+1}|\mathcal{Y}_{k+1}]$$

$$= \Sigma_{k+1}^{-1}\big(Q^{-1}A\sigma_{k+1}\Sigma_k^{-1}\mu_k + C'R^{-1}y_{k+1}\big)$$

and $\Sigma_{k+1}^{-1} = C'R^{-1}C + Q^{-1} - Q^{-1}A\sigma_{k+1}A'Q^{-1}.$

The Kalman filter is usually presented in terms of the one-step ahead prediction

$$\mu_{k+1|k} = E[x_{k+1}|\mathcal{Y}_k] = A\mu_k.$$

Similarly,

$$\Sigma_{k+1|k} = E[(x_{k+1} - \mu_{k+1|k})(x_{k+1} - \mu_{k+1|k})'|\mathcal{Y}_k]$$

$$= A\Sigma_k A' + Q.$$

Then, with $K_k := \Sigma_{k|k-1}C'(C\Sigma_{k|k-1}C' + R)^{-1}$ use of the matrix inversion lemma gives the Kalman filter equations in the form:

$$\mu_k = A\mu_{k|k-1} + K_k(y_k - C\mu_{k|k-1})$$

$$\Sigma_k = \Sigma_{k|k-1} - \Sigma_{k|k-1}C'(C\Sigma_{k|k-1}C' + R)^{-1}C\Sigma_{k|k-1}.$$

Matrix Inversion Lemma. *Assuming the required inverses exist, the matrix inversion lemma states that:*

$$(A_{11} - A_{12}A_{22}^{-1}A_{21})^{-1}$$
$$= A_{11}^{-1} + A_{11}^{-1}A_{12}(A_{22} - A_{21}A_{11}^{-1}A_{12})^{-1}A_{21}A_{11}^{-1}.$$

Volatility Estimation

Suppose a price process S evolves in discrete time, $t = 0, 1, 2, \ldots,$ with dynamics

$$S_t = S_{t-1}\exp\left(\mu - \frac{\sigma_t^2}{2} + \sigma_t b_t\right).$$

Here b_t is a sequence of $N(0,1)$ random variables and σ_t represents the volatility of the price change between time

$(t-1)$ and t.

$$E[S_t|S_{t-1}] = S_{t-1} \exp \mu.$$

The price sequence $S_0, S_1, \ldots$ is observed as are the logarithmic increments

$$y_t = \log \frac{S_t}{S_{t-1}} = \mu - \frac{\sigma_t^2}{2} + \sigma_t b_t.$$

Let us suppose that $\log \sigma_t$ has dynamics $\log \sigma_t = a + b \log \sigma_{t-1} + \theta w_t$. Here again w_t is a sequence of i.i.d. $N(0,1)$ random variables. Writing $x_t = \log \sigma_t$, so that $\sigma_t = e^{x_t}$, we see

$$x_t = a + bx_{t-1} + \theta w_t$$

$$\text{and} \quad y_t = \mu - e^{2x_t}/2 + e^{2x_t} b_t.$$

The equations for the observed process y and the unobserved process $x_t = \log \sigma_t$, therefore, put us in the framework where we can apply our nonlinear filter.

In fact, let us assume that under the reference probability $\overline{P}$ both $\{x_t\}$ and $\{y_t\}$ are sequences of i.i.d. $N(0,1)$ random variables. Write

$$\mathcal{G}_t = \sigma\{x_0, \ldots, x_t, y_0, \ldots, y_{t-1}\}$$

and $\lambda_k := \dfrac{\phi\big(\theta^{-1}(x_k - a - bx_{k-1})\big)}{\theta\phi(x_k)} \, \dfrac{\phi\big(e^{-x_k}(y_k - \mu + \frac{1}{2}e^{2x_k})\big)}{e^{x_k}\phi(y_k)}$

for $k = 1, 2, \ldots$

$$\text{Set} \quad \lambda_0 = \frac{\phi\big(e^{-x_0}(y_0 - \mu + \frac{1}{2}\,e^{2x_0})\big)}{e^{x_0}\phi(y_0)}$$

$$\Lambda_t = \prod_{k=0}^{t} \lambda_k.$$

Define a new probability P, (the 'real world' probability), by setting

$$\left.\frac{dP}{d\overline{P}}\right|_{\mathcal{G}_t} = \Lambda_t.$$

We can then show that under P $\{w_t\}$, $\{b_t\}$, $t = 0, 1, 2, \ldots$ are sequences of i.i.d. $N(0,1)$ random variables where

$$w_t := \theta^{-1}(x_t - a - bx_{t-1})$$

$$b_t := e^{-x_t}\Big(y_t - \mu + \frac{1}{2}\,e^{2x_t}\Big).$$

From Bayes' theorem, for any Borel measurable function f

$$E[f(x_t)|\mathcal{Y}_t] = \frac{\overline{E}\,[\Lambda_t f(x_t)|\mathcal{Y}_t]}{\overline{E}\,[\Lambda_t|\mathcal{Y}_t]}\ .$$

The numerator defines a measure; suppose it has a density $q_t(\cdot)$ so that

$$\overline{E}\,[\Lambda_t f(x_t)|\mathcal{Y}_t] = \int_{-\infty}^{\infty} f(z)q_t(z)dz$$

Similarly to the Theorem of the previous section we have:

Theorem.

$$q_t(z) = \frac{e^{-z}\phi\Big(e^{-z}(y_t-\mu+\frac{1}{2}e^{2z})\Big)}{\theta\phi(y_t)}\int_{-\infty}^{\infty}\phi\big(\theta^{-1}(z-a-bx)\big)q_{t-1}(x)dx.$$

This gives the formula for updating the unnormalized conditional density of $x_t = \log\sigma_t$ given $\mathcal{Y}_t$.

Putting $f(x) = 1$ in ( ) we see

$$\overline{E}\left[\Lambda_t | \mathcal{Y}_t\right] = \int_{-\infty}^{\infty} q_t(z)dz$$

so the normalized conditional density of $x_t = \log \sigma_t$ given $\mathcal{Y}_t$ is

$$p_t(z) = q_t(z) \Big/ \int_{-\infty}^{\infty} q_t(x)dx.$$

Furthermore, taking $f(x_t) = x_t$ we see

$$E[x_t|\mathcal{Y}_t] = \int_{-\infty}^{\infty} zq_t(z)dz \Big/ \int_{-\infty}^{\infty} q_t(z)dz.$$

This is the optimal estimate of the logarithm of the volatility given the observations of the prices.

Calibration. Suppose H, F, G are integrable functions. Consider

$$S_t := \sum_{k=1}^{t} H(y_k)F(x_k)G(x_{k-1}).$$

We wish to estimate

$$E[S_t|\mathcal{Y}_t].$$

Consider an associated measure and suppose there is a density $L_t(z)$ such that

$$\overline{E}\left[\Lambda_t S_t f(x_t)|\mathcal{Y}_t\right] = \int_{-\infty}^{\infty} f(z) L_t(z) dz$$

for any integrable function f. We can derive the following formula for updating L_t.

Theorem

$$L_t(z) = \Delta(y_t, z)\Big[\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z-a-bx)\big) L_{t-1}(x)dx$$

$$+ H(y_t)F(z)\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z-a-bx)\big) G(x) q_{t-1}(x)dx\Big]$$

where $\Delta(y_t, z) = \dfrac{\phi\big(e^{-z}(y_t - \mu + \frac{1}{2}\,e^{2z})\big)}{\theta e^z \phi(y_t)}$.

Proof. $\overline{E}\left[\Lambda_t S_t f(x_t)|\mathcal{Y}_t\right] = \int_{-\infty}^{\infty} f(z) L_t(z) dz$

$$= \overline{E}\left[\Lambda_{t-1} \frac{\phi\big(\theta^{-1}(x_t - a - bx_{t-1})\big)}{\theta\phi(x_t)} \frac{\phi\big(e^{-x_t}(y_t - \mu + \frac{1}{2}e^{2x_t})\big)}{e^{x_t}\phi(y_t)}\right.$$

$$\left. \times\, S_{t-1} f(x_t)|\mathcal{Y}_t\right]$$

$$
\begin{aligned}
&+ \overline{E}\Big[\Lambda_{t-1} \frac{\phi\big(\theta^{-1}(x_t - a - bx_{t-1})\big)}{\theta\phi(x_t)} \frac{\phi\big(e^{-x_t}(y_t - \mu + \frac{1}{2}e^{2x_t})\big)}{e^{x_t}\phi(y_t)} \\
&\quad \times H(y_t)F(x_t)G(x_{t-1})f(x_t)|\mathcal{Y}_t\Big] \\
&= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z - a - bx)\big) \frac{\phi\big(e^{-z}(y_t - \mu + \frac{1}{2}e^{2z})\big)}{\theta e^{z}\phi(y_t)} \\
&\quad \times f(z)L_{t-1}(x)dx\,dz \\
&+ \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z - a - bx)\big) \frac{\phi\big(e^{-z}(y_t - \mu + \frac{1}{2}e^{2z})\big)}{\theta e^{z}\phi(y_t)} \\
&\quad \times H(y_t)F(z)G(x)f(z)q_{t-1}(x)dx\,dz.
\end{aligned}
$$

This equation holds for all integrable f and the result follows.

Corollary. *Taking* $f(z) \equiv 1$ *we see*

$$
\overline{E}\left[\Lambda_t S_t|\mathcal{Y}_t\right] = \int_{-\infty}^{\infty} L_t(z)dz.
$$

Further, from Bayes' formula:

$$E[S_t|\mathcal{Y}_t] = \frac{\overline{E}\,[\Lambda_t S_t|\mathcal{Y}_t]}{\overline{E}\,[\Lambda_t|\mathcal{Y}_t]}$$

$$= \int_{-\infty}^{\infty} L_t(z)dz \Big/ \int_{-\infty}^{\infty} q_t(z)dz.$$

Special Cases:

1) For $S_t^1 = \sum_{k=1}^{t} x_k$ a measure γ_t^1 is defined by

$$\overline{E}\,[\Lambda_t S_t' f(x_t)|\mathcal{Y}_t] = \int_{-\infty}^{\infty} f(z)\gamma_t^1(z)dz.$$

This is updated by the formula

$$\gamma_t^1(z) = \Delta(y_t, z)\Big[\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z-a-bx)\big)\gamma_{t-1}^1(x)dx$$

$$+ z\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z-a-bx)\big)q_{t-1}(x)dx\Big].$$

Then $E[S_t'|\mathcal{Y}_t] = \int_{-\infty}^{\infty} \gamma_t^1(z)dz \Big/ \int_{-\infty}^{\infty} q_t(z)dz.$

2) For $S_t^2 = \sum_{k=1}^{t} x_{k-1}$ the corresponding measure $\gamma_t^2(z)$ is updated by

$$\gamma_t^2 = \Delta(y_t, z)\Big[\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z - a - bx)\big)\gamma_{t-1}^2(x)dx$$

$$+ \int_{-\infty}^{\infty} x\phi\big(\theta^{-1}(z - a - bx)\big)q_{t-1}(x)dx\Big].$$

3) For $J_t = \sum_{k=1}^{t} x_k x_{k-1}$ the corresponding measure $\beta_t^1(z)$ is updated by

$$\beta_t^1(z) = \Delta(y_t, z)\Big[\int_{-\infty}^{\infty} \phi\big(\theta^{-1}(z - a - bx)\big)\beta_{t-1}^1(x)dx$$

$$+ z\int_{-\infty}^{\infty} x\phi\big(\theta^{-1}(z - a - bx)\big)q_{t-1}(x)dx\Big].$$

4) Similar formulae, which are all special cases of the expression for $L_t(z)$ are obtained for updating the measures:

$$\beta_t^2(z) \quad \text{associated with} \quad \sum_{k=1}^{t} x_{k-1}^2$$

$$\beta_t^3(z) \quad \text{associated with} \quad \sum_{k=1}^{t} x_k^2$$

$$\nu_t^1(z) \quad \text{associated with} \quad \sum_{k=1}^{t} y_k e^{-2x_k}$$

$$\nu_t^2(z) \quad \text{associated with} \quad \sum_{k=1}^{t} e^{-2x_k}.$$

In all cases the conditional expectation of the sum, given the observations, is obtained by normalizing the integral of the associated measure. For example

$$E\Big[\sum_{k=1}^{t} y_k e^{-2x_k} \,\Big|\, \mathcal{Y}_t\Big] = \int_{-\infty}^{\infty} \nu_t^1(z)dz \Big/ \int_{-\infty}^{\infty} q_t(z)dz.$$

Parameter Estimation. Estimates of the sums above can be used to apply to EM (Expectation Maximization) algorithm. This was developed by Baum and Welch and is described in our book on Hidden Markov Models (Springer, 1995). Parameters in our model can be re-estimated recursively and, further, one parameter at a time can be updated.

For example, suppose after some iteration a parameter set (a, b', θ, μ) is obtained and we wish to re-estimate the parameter b, given the observations $y_0, y_1, \dots, y_t$.

Consider a change of measure which replaces parameter b' in our model by b. This is given by a Radon-Nikodym derivative

$$\Lambda_t^b := \prod_{k=1}^{t} \frac{\phi\big(\theta^{-1}(x_k - a - bx_{k-1})\big)}{\phi\big(\theta^{-1}(x_k - a - b'x_{k-1})\big)}$$

and setting $\frac{dP^b}{dP^{b'}}\Big|_{\mathcal{G}_t} = \Lambda_t^b.$

The maximization step determines the conditional expectation of $\log \Lambda_t^b$ given the observations. That is, consider

$$E[\log \Lambda_t^b | \mathcal{Y}_t] = E\Big[-\frac{1}{2} \sum_{k=1}^{t} \big(\theta^{-1}(x_k - a - bx_{k-1})^2\big) + R(b') | \mathcal{Y}_t \Big],$$

where $R(b')$ does not involve b.

The first order condition gives the maximum value of b as:

$$\widehat{b}_t = \frac{E\big[\sum_{k=1}^{t} x_k x_{k-1} - a \sum_{k=1}^{t} x_{k-1} | \mathcal{Y}_t\big]}{E\big[\sum_{k=1}^{t} x_{k-1}^2 | \mathcal{Y}_t\big]}$$

$$= \int_{-\infty}^{\infty} \big(\beta_t^1(z) - a\gamma_t^2(z)\big) dz \Big/ \int_{-\infty}^{\infty} \beta^2(z) dz.$$

Similar arguments give estimates

$$\widehat{a}_t = \frac{1}{t} E\Big[\sum_{k=1}^{t} x_k - b \sum_{k=1}^{t} x_{k-1} | \mathcal{Y}_k \Big]$$

$$= \int_{-\infty}^{\infty} \big(\gamma_t^1(z) - b\gamma_t^2(z)\big) dz \Big/ t \int_{-\infty}^{\infty} q_t(z) dz$$

$$(\widehat{\theta}_t)^2 = \frac{1}{2t} E\Big[\sum_{k=1}^{t}(x_k - a - bx_{k-1})^2|\mathcal{Y}_t\Big]$$

$$= \frac{\int_{-\infty}^{\infty}[\beta_t^3(z) + b\beta_t^2(z) - 2b\beta_t^1(z) - 2a\gamma_t^1(z) + 2ab\gamma_t^2(z)]dz}{2t\int_{-\infty}^{\infty} q_t(z)dz}$$

$$- \frac{a}{2},$$

$$\widehat{\mu}_t = \frac{\frac{t}{2} + E[\sum_{k=1}^{t} y_k e^{-2x_k}|\mathcal{Y}_t]}{E[\sum_{k=1}^{t} e^{-2x_k}|\mathcal{Y}_t]}$$

$$= \frac{\frac{t}{2}\int_{-\infty}^{\infty} q_t(z)dz + \int_{-\infty}^{\infty}\nu_t^1(z)dz}{\int_{-\infty}^{\infty}\nu_t^1(z)dz}.$$

Filtering Derivative Security Valuations From Market Prices

This problem is discussed in a joint paper with Dilip Madan and Charles LaLaie. The paper appeared in the Cambridge University Press volume 'Mathematics of Derivative Securities' which reports proceedings at the 1995 Bank of England Conference at the Isaac Newton Institute.

The paper models the closing price of the S and P (= Standard and Poors) 500 index as a log-normal price observed in noise.

The prices of options on this index are taken as noisy observations of the 'true' index price.

That is, in discrete time and under the 'real world' probability P we suppose the index price has dynamics

$$S_{t+1} = S_t \exp\Big((\mu - q)h - \frac{\sigma^2 h}{2} + \sigma\sqrt{h}\, w_{t+1}\Big).$$

Here $h = 1/254$, (254 trading days in 1992) and $\{w_t\}$ is a sequence of i.i.d. $N(0,1)$ random variables and q is the dividend rate.

Writing $x_t = \log S_t$

$$x_{t+1} = x_t + (\mu - q)h + \sigma\sqrt{h}\, w_{t+1} - \frac{\sigma^2 h}{2}\,.$$

We suppose the closing price c_t gives the 'true' index price x_t with some (normally distributed) error ε_t. That is, $\log c_t = y_t = x_t + \nu_1 \varepsilon_t - \frac{\nu_1^2}{2}\,.$

Furthermore, we suppose the quoted index futures option prices for different maturities τ_{ti}, $1 \le i \le 16$, and strike prices k_{ti}, $1 \le i \le 16$ are given by the Black-Scholes prices multiplied by an error term $\exp\left(\nu_2 w_{ti} - \frac{\nu_2^2}{2}\right)$.

That is, the quoted option prices y_{ti}, $1 \le i \le 16$, for 16 put and call options at each time t, with different maturities and strikes, are of the form

$$y_{ti} = V(x_t, k_{ti}, \tau_{ti}, u_{ti}, r, q, \sigma) \exp\left(\nu_2 w_{ti} - \frac{\nu_2^2}{2}\right).$$

Here u_{ti} is a type indicator, which is 1 for a call and 0 for a put. Then

$$V(x, k, \tau, u, r, q, \sigma) = (-1)^{1+u}\left[e^{x-q\tau} N(d_1) - k e^{-r\tau} N(d_2)\right]$$

where

$$d_1 = (-1)^{1+u}\left[\frac{x - \log k}{\sigma\sqrt{\tau}} + \left(\frac{r-q}{\sigma} + \frac{\sigma}{2}\right)\sqrt{\tau}\right]$$

$$d_2 = (-1)^{1+u}\left[\frac{x - \log k}{\sigma\sqrt{t}} + \left(\frac{r-q}{\sigma} - \frac{\sigma}{2}\right)\sqrt{\tau}\right].$$

From the observation processes we can write

$$\varepsilon_t = g(c_t, x_t) = \frac{\log c_t - x_t + \nu_1^2/2}{\nu_1}$$

and

$$w_{ti} = g_{ti}(y_{ti}, x_t) = \frac{\log(y_{ti}/V) + \nu_2^2/2}{\nu_2}\ .$$

If we write

$$\Delta_1(y_{t+1}, \xi) = \exp\Big(-\frac{1}{2} g(c_{t+1}, \xi)^2\Big) \prod_{i=1}^{16} \exp\Big(-\frac{1}{2} g_{t+1,i}(y_{t+1}, \xi)^2\Big)$$

the update formula for the unnormalized conditional density q is given by

$$q_{t+1}(z) = \Delta_1(y_{t+1}, z)\Gamma_1(z)$$

where

$$\Gamma_1(z) = \int_{-\infty}^{\infty} \exp\Big(-\ \frac{1}{2}\Big(\frac{z - \xi - (\mu - q)h + \frac{\sigma^2 h}{2}}{\sigma\sqrt{h}}\Big)^2\Big) q_t(\xi) d\xi.$$

This integral and its update are approximated by discretizing q over its support and using Simpson's rule.

The support for the function $q_t(z)$ is taken to be the symmetric interval centered on the current estimate for $x_t = \log S_t$. In fact, the interval was taken to be $[\log S_t - 0.05, \log S_t + 0.05]$.

The filtered prices were then substituted in the Black-Scholes formula and the revised values showed evidence that the 'smile' effect was decreased.

4. Parameter Estimation for Linear Gaussian Systems

We return to the linear Gaussian dynamics for the signal, x, and observation, y. For $k = 0, 1, \ldots,$ on the probability space $(\Omega, \mathcal{F}, P)$:

$$x_{k+1} = Ax_k + Bw_{k+1} \in R^d$$

$$y_k = Cx_k + Dv_k \in R^m$$

x_0 is a Gaussian random variable with zero mean and covariance matrix $B_0^2 \in R^{d \times d}$ $\{w_k\}$ (resp. $\{v_k\}$) is a sequence of independent $N(0, I_d)$ (resp. $N(0, I_m)$) Gaussian random variables. We assume the matrices D and B are non-singular and symmetric. Our results can be extended to the case where B is singular; see Elliott and Krishnamurthy, IEEE Trans. Automatic Control 1999.

Measure Change. We construct our dynamics by starting under a 'reference' probability measure $\overline{P}$.

On the probability space $(\Omega, \mathcal{F}, \overline{P})$ suppose there are two sequences $\{x_k\}$, $x_k \in R^d$, $\{y_k\}$, $y_k \in R^d$, of independent identically distributed random variables.

Under the measure $\overline{P}$ the x_k are i.i.d. $N(0, I_d)$ random variables and the y_k are i.i.d. $N(0, I_m)$ random variables. For $x \in R^d$ and $y \in R^d$ write

$$\psi(x) = (2\pi)^{-d/2} \exp\left(-x'x/2\right)$$

$$\phi(y) = (2\pi)^{-m/2} \exp\left(-y'y/2\right).$$

Define the sigma-fields:

$$\mathcal{G}_k = \sigma\{x_0, x_1, \ldots, x_k, y_0, y_1, \ldots, y_k\}$$

$$\mathcal{Y}_k = \sigma\{y_0, \ldots, y_k\}.$$

For any square matrix B write $|B|$ for the absolute value of its determinant. Write:

$$\lambda_0 = \frac{\phi\big(D^{-1}(y_0 - Cx_0)\big)}{|D|\phi(y_0)}$$

and for $\ell \geq 1$:

$$\lambda_\ell = \frac{\phi\big(D^{-1}(y_\ell - Cx_\ell)\big)}{|D|\phi(y_\ell)} \cdot \frac{\psi\big(B^{-1}(x_\ell - Ax_{\ell-1})\big)}{|B|\phi(x_\ell)} .$$

For $k \geq 0$ set

$$\Lambda_k = \prod_{\ell=0}^{k} \lambda_\ell .$$

A new probability P can be defined, as before, by setting

$$\frac{dP}{d\overline{P}}\Big|_{\mathcal{G}_k} = \Lambda_k .$$

The following result can then be proved:

Lemma. *Write*

$$v_\ell = D^{-1}(y_\ell - Cx_\ell), \qquad \ell = 0, 1, \ldots$$

$$w_\ell = B^{-1}(x_\ell - Ax_{\ell-1}), \quad \ell = 1, 2, \ldots .$$

Then, under the measure P, $\{v_\ell\}, \{w_\ell\}$ *are sequences of independent* $N(0, I_d)$ *and* $N(0, I_m)$ *random variables, respectively.*