Lecture 2. Nonlinear Filtering in Discrete Time

Let X, Z be two independent r.v. on $(\Omega, \mathcal{F}, P)$. Assume that both have densities, say f_X, f_Z and that $X \in L^2$.

It is required to estimate X, given the observed realization of the sum $Y = X + Z$, in other words, to find a Borel measurable function $\bar{g} : \mathbb{R} \to \mathbb{R}$, such that

$$E[(X - \bar{g}(Y))^2] \leq E[(X - g(Y))^2] \tag{2.1}$$

for any other measurable function g.

Note that such $\bar{g}$ satisfies $\bar{g}(Y) \in L^2$, since (2.1) with $g = 0$ and

$$\bar{g}^2(Y) \leq 2X^2 + 2(X - \bar{g}(Y))^2,$$

(Check that $a^2 \leq 2b^2 + 2(a-b)^2$ $\forall a, b \in \mathbb{R}$)

imply $E[\bar{g}(Y)^2] \leq 4 E[X^2] < +\infty$

Moreover, if $\bar{g}$ satisfies

$$E[(X - \bar{g}(Y)) g(Y)] = 0, \tag{2.2}$$

for any measurable g such that $E[g(Y)^2] < +\infty$, then (2.1) also holds.

In fact, if $E[(X - g(Y))^2] = \infty$, the claim is trivial and if $E[(X - g(Y))^2] < +\infty$, then

$$E[g(Y)^2] \leq 2E[X^2] + 2E[(X - g(Y))^2]$$

and we can write

$$E[(X-g(Y))^2] = E[(X-\bar{g}(Y)+\bar{g}(Y)-g(Y))^2] \quad (2.2)$$

$$= E[(X-\bar{g}(Y))^2] + E[(\bar{g}(Y)-g(Y))^2] \geq E[(X-\bar{g}(Y))^2]$$

This last inequality also suggests that $\exists\ \tilde{g}$ such that also satisfies (2.1) then $\tilde{g}$ should be equal to $\bar{g}$ on any set A such that $P(Y \in A) > 0$. In other words, $\bar{g}$ and $\tilde{g}$ must be equal on the support of the law of Y.

Does such a function exist?

Yes, we can give an explicit construction using (2.2)

$$E[(X-\bar{g}(Y))g(Y)] = \iint_{\mathbb{R}\mathbb{R}} (x-\bar{g}(x+z))\, g(x+z)\, f_X(x) f_Z(z)\, dx\, dz$$

↑ X and Z have densities and are independent.

$$\begin{pmatrix} u = x+z \\ du = dz \end{pmatrix} = \int_{\mathbb{R}} \int_{\mathbb{R}} (x-\bar{g}(u))\, g(u)\, f_X(x) f_Z(u-x)\, dx\, du$$

$$= \int_{\mathbb{R}} g(u) \left(\int_{\mathbb{R}} (x-\bar{g}(u))\, g(u)\, f_X(x) f_Z(u-x)\, dx \right) du$$

The latter would vanish if

$$\int_{\mathbb{R}} (x-\bar{g}(u))\, f_X(x) f_Z(u-x)\, dx = 0$$

for all u, which leads to choose

$$\bar{g}(u) := \frac{\int_{\mathbb{R}} x\, f_X(x) f_Z(u-x)\, dx}{\int_{\mathbb{R}} f_X(x) f_Z(u-x)\, dx}$$

Hence, the best estimate of X given Y is the r.v.

$$(2.3)\qquad \bar{g}(Y(\omega)) = \frac{\int_{\mathbb{R}} x\, f_X(x)\, f_Z(Y(\omega)-x)\,dx}{\int_{\mathbb{R}} f_X(x)\, f_Z(Y(\omega)-x)\,dx} = E[X|Y](\omega),$$

i.e., the Bayes formula for the conditional expectation of X given Y.

2.1. The conditional expectation: a clever look

2.1.1. Definition and basic properties

Let $(\Omega, \mathcal{F}, P)$ be a prob. space carrying a r.v. $X \geq 0$ and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.

Def 2.1 The conditional expectation of $X \geq 0$ with respect to $\mathcal{G}$ is a r.v., denoted by $E[X|\mathcal{G}](\omega)$, which is $\mathcal{G}$-measurable, i.e.,

$$\{\omega : E[X|\mathcal{G}](\omega) \in A\} \in \mathcal{G}, \quad \forall A \in \mathcal{B}(\mathbb{R})$$

and satisfies

$$E\left[(X - E[X|\mathcal{G}])\, 1_A\right] = 0 \qquad \forall A \in \mathcal{G}.$$

Is this definition correct? Does there exist such r.v.? And is it unique? Yes!!!

Theorem 2.2 (Radon-Nikodym) Let $(X, \mathcal{X})$ be a measurable space, μ be a σ-finite measure on X is a signed measure, absolutely continuous with respect to μ. Then, there exists an $\mathcal{X}$-measurable function $f = f(x)$, taking values in $\mathbb{R} \cup \{\pm\infty\}$, such that

$$\nu(A) = \int_A f(x)\, \mu(dx), \qquad A \in \mathcal{X}.$$

f is called the Radon-Nikodym derivative (a density) of ν with respect to μ and denoted $\frac{d\nu}{d\mu}$. It is unique up to μ-null sets.

Now consider the measurable space $(\Omega, \mathcal{G})$, $\mathcal{G}$ a sub-sigma algebra of $\mathcal{F}$, and define a nonnegative set function on $\mathcal{G}$

$$(2.4) \qquad Q(A) = \int_A X \, P(d\omega) = E[X 1_A], \quad A \in \mathcal{G}$$

This set function is a nonnegative σ-finite measure: take the partition $D_j = \{ j \le X < j+1 \}$, $j = 0, 1, \dots$, then $Q(D_j) = E[X 1_{\{X \in [j, j+1)\}}] < +\infty$ even if $E[X] = \infty$.

That $Q \ll P$ (i.e. $P(A) = 0 \Rightarrow Q(A) = 0$) follows from the basic properties of the Lebesgue integral, but an easy verification is as follows. Let $(X_j)_{j \ge 1}$ be a sequence of simple r.v., such that $X_j \uparrow X$, for example

$$X_j = \sum_{\ell=1}^{j 2^j} \frac{\ell - 1}{2^j} 1_{\left\{ \frac{\ell-1}{2^j} \le X(\omega) < \frac{\ell}{2^j} \right\}} + j 1_{\{X(\omega) \ge j\}}$$

(or more generally

$$X_j = \sum_k x_k^j 1_{B_k^j}, \quad B_k^j \in \mathcal{F}, \quad x_k^j \in \mathbb{R}_+$$

Since $E[X_j 1_A] = \sum_k x_k^j P(B_k^j \cap A) = 0$

by monotone convergence $Q(A) = E[X 1_A] = \lim_{j \to \infty} E[X_j 1_A] = 0$

Applying the Radon-Nikodym theorem $\exists !$ (up to P-null sets) r.v. ξ, measurable with respect to $\mathcal{G}$, such

Let $$Q(A) = \int_A \xi \, dP \quad , \forall A \in \mathcal{G}.$$

This ξ is said to be a version of $E[X|\mathcal{G}]$ to emphasize its uniqueness up to P-null sets,

$$E[X|\mathcal{G}](\omega) = \frac{dQ}{dP}(\omega).$$

For general r.v. X define $X^+ = \max(0, X)$

$$E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}] \qquad X^- = \max(0, -X)$$

if no $\infty - \infty$ occurs with positive probability. ($\infty - \infty$ is allowed on the P-null sets)

Note that $E[X|\mathcal{G}]$ may be well defined even if when $E[X]$ is not.

For example, let $\mathcal{G} = \mathcal{F}^X = \mathcal{F}(X)$ and suppose that $E[X^+] = E[X^-] = \infty$, so $E[X]$ is not well defined, but we still have

$$E[X|\mathcal{G}] = E[X^+|\mathcal{G}] - E[X^-|\mathcal{G}] = X^+ - X^- = X,$$

note that $\{X^+ = +\infty\} \cap \{X^- = \infty\}$ is a null set.

Example 2.3

Let $\mathcal{G}$ be (finite) σ-algebra generated by the finite partition

$D_j \in \mathcal{F}$, $j = 1, \dots, n$, $\bigcup D_j = \Omega$, $P(D_j) > 0$.

Any $\mathcal{G}$-measurable r.v. ξ is necessarily constant on each set D_j: Suppose (on the contrary) that it takes two values on (e.g.) D_j, say $x' < x''$, then $\{\omega : X(\omega) \le x'\} \cap D_j$ and $\{\omega : X(\omega) \ge x''\} \cap D_j$ are disjoint subsets of D_j and, hence, not in any other D_i, $i \neq j$. Therefore both events cannot be in $\mathcal{G}$, which contradicts the $\mathcal{G}$-measurability of X.

So for any r.v. X, we get

$$E[X|\mathcal{G}](\omega) = \sum_{j=1}^{n} a_j \mathbb{1}_{D_j}(\omega)$$

the constants a_j are found from

$$E\left[\left(X - \sum_{j=1}^{m} a_j 1_{D_j}\right) 1_{D_i}\right] = 0, \quad i = 1, \dots, m.$$

which leads to

$$E[X|\mathcal{G}](\omega) = \sum_{j=1}^{m} \frac{E[X 1_{D_j}]}{P(D_j)} 1_{D_j}(\omega)$$

Of special interest is when we condition with respect σ-algebras generated by r.v., i.e.,

$$\mathcal{F}^Y := \sigma(Y) := \sigma\left\{ Y^{-1}(B), B \in \mathcal{B}(\mathbb{R}) \right\}$$

Given (X, Y) a pair of r.v., $E[X|\sigma(Y)]$ is written sometimes as $E[X|Y]$.

Moreover, it can be shown that for any $\sigma(Y)$-measurable r.v. Z, there exists a Borel function ξ such that $Z(\omega) = \xi(Y(\omega))$. This applies, in particular, to $E[X|Y]$ and this function is denoted sometimes by $E[X|Y=y]$.

Main properties of the conditional expectation

Let $X, Y, \{X_n\}_{n \in \mathbb{N}}$ be random variables in L^1, and let $\mathcal{G}$ and $\mathcal{H}$ be sub-σ-algebras of $\mathcal{F}$. Then

a) (linearity) $E[aX + bY|\mathcal{G}] = a E[X|\mathcal{G}] + b E[Y|\mathcal{G}]$, $a, b \in \mathbb{R}$.

b) (monotonicity) $X \leq Y$ a.s., implies $E[X|\mathcal{G}] \leq E[Y|\mathcal{G}]$, a.s.

c) (identity on $L^1(\mathcal{G})$) If X is $\mathcal{G}$-measurable, then $X = E[X|\mathcal{G}]$, a.s. In particular, $c = E[c|\mathcal{G}]$, for any constant $c \in \mathbb{R}$.

d) (conditional Jensen's inequality) If $\psi : \mathbb{R} \to \mathbb{R}$ is convex and $E[|\psi(X)|] < +\infty$, then

$$E[\psi(X)|\mathcal{G}] \geq \psi(E[X|\mathcal{G}]), \text{ a.s.}$$

Important Corollary: $\|E[X|\mathcal{G}]\|_p \leq \|X\|_p$, $p \geq 1$

e) (pulling out what is known). If Y is $\mathcal{G}$-measurable and $XY \in L^1$, then

$$E[XY|\mathcal{G}] = Y E[X|\mathcal{G}], \text{ a.s.}$$

f) (L^2-projection) If $X \in L^2$, then $\xi^* = E[X|\mathcal{G}]$ minimizes $E[(X-\xi)^2]$ over all $\mathcal{G}$-measurable random variables $\xi \in L^2$.

g) (tower property) If $\mathcal{H} \subseteq \mathcal{G}$, then

$$E[E[X|\mathcal{G}]|\mathcal{H}] = E[X|\mathcal{H}], \text{ a.s.}$$

h) (irrelevance of independent information) If $\mathcal{H}$ is independent of $\sigma(\mathcal{G}, \sigma(X))$ then

$$E[X|\sigma(\mathcal{G},\mathcal{H})] = E[X|\mathcal{G}], \text{ a.s.}$$

In particular, if X is independent of $\mathcal{H}$, then

$$E[X|\mathcal{H}] = E[X], \text{ a.s.}$$

i) (Substitution property or freezing lemma). Let $h: \mathbb{R}\times\mathbb{R} \to \mathbb{R}$ be a Borel function. If X is $\mathcal{G}$-measurable, $h(X,Y) \in L^1$, and Y is independent of $\mathcal{G}$ then

$$E[h(X,Y)|\mathcal{G}] = E[h(a,Y)]\big|_{a=X}$$

j) (C MON) If $0 \leq X_n \leq X_{n+1}$ (a.s.) for all $n \in \mathbb{N}$ and $X_n \xrightarrow{a.s.} X \in L^1$, then

$$E[X_n|\mathcal{G}] \uparrow E[X|\mathcal{G}], \text{ a.s.}$$

k) (C FATOU) If $X_n \geq 0$ (a.s.) for all $n \in \mathbb{N}$, and $\liminf X_n \in L^1$, then

$$E[\liminf X_n|\mathcal{G}] \leq \liminf E[X_n|\mathcal{G}], \text{ a.s.}$$

l) (c DOM) If $|X_n| \le Z$, for all $n \in \mathbb{N}$ and some $Z \in L^1$, [2.8]
and if $X_n \to X$, a.s., then

$$E[X_n|\mathcal{G}] \to E[X|\mathcal{G}], \text{ a.s. and in } L^1.$$

Let $(A_j)_{j\ge 1}$ be a sequence of disjoint events, then

(2.5) $$P\Big(\biguplus_j A_j \,\Big|\, \mathcal{G}\Big] = \sum_j P(A_j|\mathcal{G}).$$

So one is tempted to think that for any fixed $\omega \in \Omega$, $P(A|\mathcal{G})(\omega) = E[1_A|\mathcal{G}](\omega)$ is a measure on $\mathcal{A}$. In general, this is wrong because (2.5) only holds P-a.s. and we need to check (2.5) for an uncountable number of sequences. Hence, the intersection of all null sets for which (2.5) fails may pile up to a set of positive probability. Nevertheless, in most interesting situations we can show that this does not happen.

Definition 2.4 (Transition Kernels)

Let $(G_1, \mathcal{G}_1)$ and $(G_2, \mathcal{G}_2)$ be two measurable spaces. An unnormalized transition kernel from $(G_1, \mathcal{G}_1)$ to $(G_2, \mathcal{G}_2)$ is a function $Q : G_1 \times \mathcal{G}_2 \to [0, \infty]$ that satisfies

i) for all $x \in G_1$, $Q(x, \cdot)$ is a measure on $(G_2, \mathcal{G}_2)$

ii) for all $A \in \mathcal{G}_2$, the function $x \mapsto Q(x, A)$ is $\mathcal{G}_1$-measurable.

If $Q(x, G_2) = 1$, for all $x \in G_1$, Q is called a transition kernel (or simply a kernel)

If $G_1 = G_2$ and $Q(x, G_1) = 1$ for all $x \in G_1$, then Q is called a Markov transition kernel on (G_1, G_1).

An (unnormalized) transition kernel Q is said to admit a density with respect to the positive measure μ on G_2 if there exists a non-negative function $q : G_1 \times G_2 \to [0, \infty]$, measurable with respect to $G_1 \otimes G_2$, such that

$$Q(x, A) = \int_A q(x, y)\, \mu(dy), \quad A \in G_1$$

the function q is then called an (unnormalized) transition density function.

Definition 2.5

On a probability space $(\Omega, \mathcal{F}, P)$, let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$. Let $(S, \mathcal{S})$ be a measurable space, and let $X : \Omega \to S$ be a random element in S. A transition kernel $Q = Q_{X|\mathcal{G}} : \Omega \times \mathcal{S} \to [0,1]$ is called a regular conditional distribution of X given $\mathcal{G}$, if

$$Q(\omega, B) = P(X \in B \mid \mathcal{G})(\omega), \text{ p.a.s., } \forall B \in \mathcal{S}$$

Remark 2.6

1. When $(S, \mathcal{S}) = (\Omega, \mathcal{F})$ and $X(\omega) = \omega$, the regular conditional distribution of X (if it exists) is called the regular conditional probability on $\mathcal{F}$ given $\mathcal{G}$. In this case,

$$P_{\mathcal{G}}(\omega, B) := Q_X|_{\mathcal{G}}(\omega, B) = P(X \in B(\mathcal{G})(\omega) = P(B|\mathcal{G})(\omega)$$

2. Regular conditional distributions do not need to exist when $(S, \mathcal{S})$ is "too large".

Definition 2.7

A measurable space $(S, \mathcal{S})$ is said to be a Borel space (or a nice space) if it is isomorphic to a Borel subset of $\mathbb{R}$, i.e., if there exists a bijection $\ell : S \to \mathbb{R}$ such that both ℓ and ℓ^{-1} are measurable.

Any Polish space (complete, separable metric space) is a Borel space.

Theorem 2.8

Let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$, and let $(S, \mathcal{S})$ be a Borel space. Any random element $X : \Omega \to S$ admits a regular conditional distribution given $\mathcal{G}$.

Remark 2.9

Combining the previous Theorem with Remark 2.6 1), we have that if the probability space under consideration $(\Omega, \mathcal{F}, P)$ is the "canonical space" of Borel space-valued random variable, then regular conditional probabilities given any $\mathcal{G} \subseteq \mathcal{F}$ can always be defined on the whole $\mathcal{F}$.

For more general $(\Omega, \mathcal{F}, P)$, we still can always define r.c.p. but on certain smaller σ-algebras.

For example, let X be a real valued random variable defined on $(\Omega, \mathcal{F}, P)$, and $\mathcal{G} \subseteq \mathcal{F}$. We can always

define $P_G : \Omega \times \mathcal{F}(X) \to [0,1]$ by

$$P_G(\omega, X^{-1}(B)) := Q_{X|G}(\omega, B) \qquad B \in \mathcal{B}(\mathbb{R}).$$

P_G is called the conditional probability of X given G.

And this will suffice if we are only interested in computing conditional expectations of functions applied to X.

Proposition 2.9

Let X be a random vector, $G \subseteq \mathcal{F}$, and let $h : \mathbb{R}^n \to \mathbb{R}$ be a Borel measurable function such that $h(X) \in L^1$. Then

$$E[h(X) \mid G](\cdot) = \int_{\mathbb{R}^n} h(x)\, Q_{X|G}(\cdot, dx), \quad P\text{-a.s.}$$

$$= \int_{\Omega} h(X(\hat{\omega}))\, P_G(\cdot, d\hat{\omega}), \quad P\text{-a.s.}$$

Conditioning on the value of a r.v.

If X and Y are two r.v. with $E[|X|] < +\infty$.
We know that since $E[X \mid Y] := E[X \mid \mathcal{F}(Y)]$ is $\mathcal{F}(Y)$ measurable then $E[X \mid Y](\omega) = h(Y(\omega))$ for some Borel function h.
The conditional expectation of X given the value of Y may be defined by $E[X \mid Y = y] = h(y)$.

that is, $E[X \mid Y=y]$ may be regarded as a version of the conditional expectation induced on $\mathbb{R}$ by the transformation $Y(\omega)$ (and thus Borel measurable, rather than $\mathcal{F}(Y)$-measurable)

If $B \in \mathcal{B}(\mathbb{R})$

$$\int_B E[X \mid Y=y]\, P_Y(dy) = \int_B h(y)\, P_Y(dy)$$

$$= \int_{Y^{-1}(B)} h \circ Y(\omega)\, P(d\omega) = \int_{Y^{-1}(B)} E[X \mid Y](\omega)\, P(d\omega)$$

$$= \int_{Y^{-1}(B)} X(\omega)\, P(d\omega)$$

Since in particular $\int_B h(y)\, P_Y(dy) = \int_{Y^{-1}(B)} X(\omega)\, P(d\omega)$,

any two choices of $h(y)$ have the same integral for every $B \in \mathcal{B}(\mathbb{R})$ and, hence, must be equal P_Y-a.e..

the conditional probability of $A \in \mathcal{F}$ given $Y=y$ is then defined as

$$P(A \mid Y=y) = E[1_A \mid Y=y] \quad P_Y\text{-a.e.}$$

then $P(A \mid Y=y)$ is a Borel measurable function, P_Y-a.e. unique satisfying

$$P(A \cap Y^{-1}(B)) = \int_B P(A \mid Y=y)\, P_Y(dy)$$

A transition kernel $\hat{Q}_{X|Y}(y, B) : \mathbb{R} \times \mathcal{B}(\mathbb{R}) \to [0,1]$ is called a regular conditional distribution of X given $Y = y$ if

i) $\forall B \in \mathcal{B}(\mathbb{R})$, $\hat{Q}_{X|Y}(\cdot, B)$ is a Borel measurable function

ii) $\forall y$ fixed $\hat{Q}_{X|Y}(y, \cdot)$ is a probability measure on $\mathcal{B}(\mathbb{R})$.

iii) For each $B \in \mathcal{B}(\mathbb{R})$, $\hat{Q}_{X|Y}(y, B) = P(X^{-1}(B) \mid Y = y)$, P_Y a.e.

Theorem 2.10

Let X and Y be two r.v.. Then

a) $\exists$ a regular cond. distribution of X given $Y = y$

ii) If $Q_{X|Y}$ and $\hat{Q}_{X|Y}$ are regular condi. dist. of X given Y and given $Y = y$, respectively then,

$$Q_{X|Y}(\omega, B) = \hat{Q}_{X|Y}(Y(\omega), B) \quad \forall B \in \mathcal{B} \text{ and } \omega \notin N, \text{ with } N \in \mathcal{F}(Y) \text{ and } P(N) = 0.$$

iii) If g is Borel measurable such that $E[|g(X,Y)|] < \infty$,

$$E[g(X,Y) \mid Y = y] = \int_{\mathbb{R}} g(x, Y)\, \hat{Q}_{X|Y}(y, dx), \; P_Y \text{ a.e.}$$

Another application of transition kernels is the following version of Tonelli's theorem.

Definition 2.11

A family of measures $\{\mu_i\}_{i \in I}$ on a measurable space $(E, \mathcal{E})$ is uniformly σ-finite iff there exists a decomposition $E = \bigcup_{n \geq 1} A_n$ with $A_n \in \mathcal{E}$, and constants $K_n > 0$, such that $\forall n \in \mathbb{N}, \forall i \in I$, $\mu_i(A_n) \leq K_n$.

Theorem 2.12 (Generalised Fubini)

Let $(E_1, \mathcal{E}_1, \mu)$ be a measure space with μ being σ-finite. Let $(E_2, \mathcal{E}_2)$ be a measurable space. Let $\tau : E_1 \times \mathcal{E}_2 \to \mathbb{R}_+$ be an unnormalised transition kernel such that the family of measures $\{\tau(e_1, \cdot)\}_{e_1 \in E_1}$ is uniformly σ-finite. Then

(1) $\exists!$ measure ν on $(E_1 \times E_2, \mathcal{E}_1 \otimes \mathcal{E}_2)$ such that

$$\nu(A_1 \times A_2) = \int_{A_1} \tau(e_1, A_2)\, \mu(de_1)$$

(2) If $g : E_1 \times E_2 \to \mathbb{R}$ is a function $(\mathcal{E}_1 \otimes \mathcal{E}_2, \mathcal{B}(\mathbb{R}))$-measurable and the integral of g with respect to ν exists, then the function

$$e_1 \longmapsto \int_{E_2} g(e_1, e_2)\, \tau(e_1, de_2)$$

is well defined μ-a.e., it is measurable and its integral with respect to μ exists and is equal to

$$\int_{E_1 \times E_2} g \, d\nu = \int_{E_1} \left(\int_{E_2} g(e_1, e_2)\, \tau(e_1, de_2) \right) \mu(de_1). \tag{2.6}$$

Theorem 2.13 (Generalised Tonelli).

Assume the same notation of Theorem 2.11. Then, if g is $(\mathcal{E}_1 \otimes \mathcal{E}_2, \mathcal{B}(\mathbb{R}))$-measurable and

$$\int_{E_1} \left(\int_{E_2} |g(e_1, e_2)|\, \tau(e_1, de_2) \right) \mu(de_1) < +\infty,$$

then g is integrable with respect to ν and the formula (2.6) holds.

2.1.2. The Bayes formula: an abstract formulation

Formula (2.3) involved explicit knowledge of the densities of the r.v. involved.
The following result allows to find Bayes formula for more general conditioning σ-algebras.

Theorem 2.14 (Bayes formula)

Let $(\Omega, \mathcal{F}, P)$ be a prob. space, carrying an r.v. X and let $\mathcal{G}$ be a sub-σ-algebra of $\mathcal{F}$.
Assume that there exists a regular conditional probability measure $P(dw | X=x)$ on $\mathcal{G}$ and it has Radon-Nikodym density $\rho(w,x)$ with respect to a σ-finite measure λ (on $\mathcal{G}$):

$$P(B \mid X=x) = \int_B \rho(w,x)\, \lambda(dw).$$

Then for any $\xi : \mathbb{R} \to \mathbb{R}$, such that $E[|\xi(X)|] < +\infty$,

$$E[\xi(X) \mid \mathcal{G}] = \frac{\int_{\mathbb{R}} \xi(u)\, \rho(w,u)\, P_X(du)}{\int_{\mathbb{R}} \rho(w,u)\, P_X(du)}, \tag{2.7}$$

where P_X is the law of X.

Proof

Recall that

$$E[\xi(X) \mid \mathcal{G}](w) = \frac{dQ}{dP}(w), \tag{2.8}$$

where Q is the signed measure defined by

$$Q(B) = \int_B \xi(X(w))\, P(dw), \quad B \in \mathcal{G}.$$

Then, for any $B \in \mathcal{G}$

$$P(B) = E[E[1_B \mid \sigma(X)]] = \int_\Omega P(B \mid \sigma(X))(\omega)\, P(d\omega)$$

$$\overset{(1)}{=} \int_{\mathbb{R}} P(B \mid X = u)\, P_X(du)$$

$$= \int_{\mathbb{R}} \int_B \varrho(\omega, u)\, \lambda(d\omega)\, P_X(du)$$

$$\overset{(2)}{=} \int_B \left(\int_{\mathbb{R}} \varrho(\omega, u)\, P_X(du) \right) \lambda(d\omega) \tag{2.9}$$

(1). Image measure theorem

(2) Fubini's theorem

Also, for any $B \in \mathcal{G}$

$$Q(B) := E[\xi(X) 1_B] = E[\xi(X)\, E[1_B \mid \sigma(X)]]$$

$$\overset{(1)}{=} \int_{\mathbb{R}} \xi(u)\, P(B \mid X = u)\, P_X(du)$$

$$= \int_{\mathbb{R}} \xi(u) \int_B \varrho(\omega, u)\, \lambda(d\omega)\, P_X(du)$$

$$\overset{(2)}{=} \int_B \left(\int_{\mathbb{R}} \xi(u)\, \varrho(\omega, u)\, P_X(du) \right) \lambda(d\omega) \tag{2.10}$$

Note that $Q \ll P$ and by (2.9) $P \ll \lambda$ (on $\mathcal{G}$) and hence also $Q \ll \lambda$.

So for any $B \in \mathcal{G}$

$$Q(B) = \int_B \frac{dQ}{dP}(\omega)\, P(d\omega) = \int_B \frac{dQ}{dP}(\omega)\, \frac{dP}{d\lambda}(\omega)\, \lambda(d\omega)$$

while on the other hand

$$Q(B) = \int_B \frac{dQ}{d\lambda}(\omega)\,\lambda(d\omega), \quad \forall B \in \mathcal{G}.$$

Since B is arbitrary we can conclude

$$\frac{dQ}{d\lambda}(\omega) = \frac{dQ}{dP}(\omega)\,\frac{dP}{d\lambda}(\omega) \quad \lambda\text{-a.s.}$$

Moreover, since

$$P\left(\omega : \frac{dP}{d\lambda}(\omega) = 0\right) = \int_\Omega 1\!\!1_{\left\{\frac{dP}{d\lambda}(\omega)=0\right\}} P(d\omega) =$$

$$= \int_\Omega 1\!\!1_{\left\{\frac{dP}{d\lambda}(\omega)=0\right\}} \frac{dP}{d\lambda}(\omega)\,\lambda(d\omega) = 0$$

we can write

$$\frac{dQ}{dP}(\omega) = \frac{\frac{dQ}{d\lambda}(\omega)}{\frac{dP}{d\lambda}(\omega)}, \qquad P\text{-a.s.}$$

Finally, (2.9) implies that $\frac{dP}{d\lambda}(\omega) = \int_{\mathbb{R}} \varrho(\omega,u)\,P_X(du)$

and (2.10) implies that $\frac{dQ}{d\lambda}(\omega) = \int_{\mathbb{R}} \xi(u)\,\varrho(\omega,u)\,P_X(du)$,

which together with (2.8) allows us to conclude.

Corollary 2.15

Suppose that $\mathcal{G}$ is generated by a random variable Y and there is a σ-finite measure ν on $\mathcal{B}(\mathbb{R})$ and a measurable function $r(y,x) \geq 0$ so that

$$P(Y \in A \mid X = x) = \int_A r(y,x)\,\nu(dy).$$

then, for $\xi(X) \in L^1$,

$$E[\xi(X) \mid \mathcal{G}] = \frac{\int_{\mathbb{R}} \xi(x)\, r(Y(\omega),x)\, P_X(dx)}{\int_{\mathbb{R}} r(Y(\omega),x)\, P_X(dx)} \tag{2.11}$$

Proof

We can write

$$P_Y(A) = P(Y \in A) = E[\, P(Y \in A \mid \sigma(X))\,] = E[\, P(Y \in A \mid X = x)\big|_{x=X}\,]$$

$$= E\left[\int_A r(y, x)\, \nu(dy)\Big|_{x=X}\right] = E\left[\int_A r(y, X)\, \nu(dy)\right]$$

$$\overset{(1)}{=} \int_A E[\, r(y, X)\,]\, \nu(dy)$$

(1) Fubini's theorem.

Define $\bar{r}(y) := E[\, r(y, X)\,]$ and define

$$\ell(\omega, x) = \begin{cases} \dfrac{r(Y(\omega), x)}{\bar{r}(Y(\omega))}, & \bar{r}(Y(\omega)) > 0 \\ 0, & \bar{r}(Y(\omega)) = 0 \end{cases} \tag{2.12}$$

Any $\mathcal{G}$-measurable set is by definition a preimage of some A under $Y(\omega)$, that is, for any $B \in \mathcal{G}$, there is $A \in \mathcal{B}(\mathbb{R})$ such that $B = \{\omega : Y(\omega) \in A\}$. Then,

$$\int_B \ell(\omega, x)\, P(d\omega) = \int_B \frac{r(Y(\omega), x)}{\bar{r}(Y(\omega))}\, P(d\omega)$$

$$= \int_A \frac{r(y, x)}{\bar{r}(y)}\, P_Y(dy) = \int_A \frac{r(y, x)}{\bar{r}(y)}\, \bar{r}(y)\, \nu(dy)$$

$$= \int_A r(y, x)\, \nu(dy) = P(Y \in A \mid X = x) = P(B \mid X = x).$$

Hence (2.11) follows from (2.7) with $\ell(\omega, x)$ given by (2.12) and $\lambda(d\omega) = P(\omega)$. (Note that $\bar{r}(y)$ in the numerator and denominator cancel out)

Remark 2.16

Let $(\dot\Omega, \dot{\mathcal{F}}, \dot P)$ be a copy of the probability space $(\Omega, \mathcal{F}, P)$, then (2.12) reads

$$E[\xi(X)\,|\,\mathcal{G}] = \frac{\dot E[\xi(X(\dot\omega))\,r(Y(\omega), X(\dot\omega))]}{\dot E[r(Y(\omega), X(\dot\omega))]} \tag{2.13}$$

where $\dot E$ denotes expectation on $(\dot\Omega, \dot{\mathcal{F}}, \dot P)$ and $X(\dot\omega)$ is a copy of X defined in this auxiliary probability space.

Remark 2.17

Formulas (2.12) and (2.13) hold when X and Y are random vectors.

Remark 2.18

The following notation is often used

$$P(X \in dx \,|\, Y = y) = \frac{r(y,x)\,P_X(dx)}{\int_{\mathbb{R}} r(y,x)\,P_X(dx)}$$

for the regular conditional distribution of X given $\sigma(Y)$. Note that it absolutely continuous with respect to the law of X.

2.2. The nonlinear filter via the Bayes formula

Let $(X_j, Y_j)_{j \geq 0}$ be a pair of random sequences with the following structure:

- X_j is a Markov process with the transition kernel $\Lambda(x, du)$ and initial distribution $p(du)$, that is,

$$P(X_j \in B \mid \mathcal{F}^X_{j-1} \vee \mathcal{F}^Y_{j-1}) = \int_B \Lambda(X_{j-1}, du), \quad P\text{-a.s.}$$

where $\mathcal{F}^X_{j-1} := \sigma(X_0, X_1, \dots, X_{j-1})$

$$P(X_0 \in B) = \int_B p(du), \quad \forall B \in \mathcal{B}(\mathbb{R})$$

- Y_j is a random sequence, such that for all $j \geq 0$

$$P(Y_j \in B \mid \mathcal{F}^X_j \vee \mathcal{F}^Y_{j-1}) = \int_B \Gamma(X_j, du), \quad P\text{-a.s.}$$

with a Markov kernel $\Gamma(x, du)$, which has density $\gamma(x, u)$ with respect to some σ-finite measure $\nu(du)$ on $\mathcal{B}(\mathbb{R})$.

Theorem 2.19

Let $\pi_j(dx)$ be the solution of the recursive equation

$$\pi_j(dx) = \frac{\int_{\mathbb{R}} \gamma(x, Y_j)\, \Lambda(u, dx)\, \pi_{j-1}(du)}{\int_{\mathbb{R}} \int_{\mathbb{R}} \gamma(x, Y_j)\, \Lambda(u, dx)\, \pi_{j-1}(du)}, \quad j \geq 1 \tag{2.14}$$

subject to

$$\pi_0(dx) = \frac{\gamma(x, y_0)\, p(dx)}{\int_{\mathbb{R}} \gamma(u, y_0)\, p(du)} \tag{2.15}$$

Let $f : \mathbb{R} \to \mathbb{R}$ be a Borel measurable function, such that $E[|f(X_j)|] < +\infty$ for each $j \geq 0$. Then

$$E[f(X_j) \mid \mathcal{F}_j^Y] = \int_{\mathbb{R}} f(x)\, \pi_j(dx), \quad P\text{-a.s.}$$

Proof

(X_j, Y_j) is Markov with transition kernel

$$\Lambda(x, du)\, \gamma(u, v)\, V(dv) :$$

$$P(X_j \in A, Y_j \in B \mid \mathcal{F}_{j-1}^X \vee \mathcal{F}_{j-1}^Y)$$

$$= E[E[\mathbb{1}_{\{X_j \in A\}} \mathbb{1}_{\{Y_j \in B\}} \mid \mathcal{F}_j^X \vee \mathcal{F}_{j-1}^Y] \mid \mathcal{F}_{j-1}^X \vee \mathcal{F}_{j-1}^Y]$$

$$= E[\mathbb{1}_{\{X_j \in A\}} E[\mathbb{1}_{\{Y_j \in B\}} \mid \mathcal{F}_j^X \vee \mathcal{F}_{j-1}^Y] \mid \mathcal{F}_{j-1}^X \vee \mathcal{F}_{j-1}^Y]$$

$$= E[\mathbb{1}_{\{X_j \in A\}} \int_B \gamma(X_j, v)\, V(dv) \mid \mathcal{F}_{j-1}^X \vee \mathcal{F}_{j-1}^Y]$$

$$= \int_A \int_B \gamma(u, v)\, V(dv)\, \Lambda(X_{j-1}, du).$$

The regular conditional distribution of $\{Y_0, \dots, Y_j\}$ given $\mathcal{F}_j^X$ is given by

$$P(Y_0 \in A_0, \dots, Y_j \in A_j \mid \mathcal{F}_j^X) = E[\mathbb{1}_{\{Y_0 \in A_0\}} \cdots \mathbb{1}_{\{Y_{j-1} \in A_{j-1}\}} P(A_j \mid \mathcal{F}_j^X \vee \mathcal{F}_{j-1}^Y) \mid \mathcal{F}_j^X]$$

$$= E[\mathbb{1}_{\{Y_0 \in A_0\}} \cdots \mathbb{1}_{\{Y_{j-1} \in A_{j-1}\}} \int_{A_j} \gamma(X_j, u_j)\, V(du_j) \mid \mathcal{F}_j^X]$$

$$= \int_{A_0} \cdots \int_{A_j} \gamma(x_0, u_0) \cdots \gamma(x_j, u_j)\, \nu(du_0) \cdots \nu(du_j)$$

then, by Remark 2.17

$$E[\xi(X_j) \mid \mathcal{F}_j^Y](\omega) = \frac{\dot{E}\left[\xi(X_j(\dot{\omega})) \prod_{i=0}^{j} \gamma(X_i(\dot{\omega}), Y_i(\omega))\right]}{\dot{E}\left[\prod_{i=0}^{j} \gamma(X_i(\dot{\omega}), Y_i(\omega))\right]}$$

We introduce the notation

$$L_j(X(\dot{\omega}), Y) = \prod_{i=2}^{j} \gamma(X_i(\dot{\omega}), Y_i)$$

and note that

$$\dot{E}\left[\xi(X_j(\dot{\omega}))\, L_j(X(\dot{\omega}), Y) \mid \mathcal{F}_{j-2}^X\right]$$

$$= L_{j-2}(X(\dot{\omega}), Y)\, \dot{E}\left[\xi(X_j(\dot{\omega}))\, \gamma(X_j(\dot{\omega}), Y_j) \mid \mathcal{F}_{j-2}^X\right]$$

$$= L_{j-2}(X(\dot{\omega}), Y) \int_{\mathbb{R}} \xi(u)\, \gamma(u, Y_j)\, \Lambda(X_{j-2}(\dot{\omega}), du)$$

then,

$$E[\xi(X_j) \mid \mathcal{F}_j^Y] = \frac{\dot{E}\left[\xi(X_j(\dot{\omega}))\, L_j(X(\dot{\omega}), Y)\right]}{\dot{E}\left[L_j(X(\dot{\omega}), Y)\right]}$$

$$= \frac{\dot{E}\left[L_{j-2}(X(\dot{\omega}), Y) \int_{\mathbb{R}} \xi(u)\, \gamma(u, Y_j)\, \Lambda(X_{j-2}(\dot{\omega}), du)\right]}{\dot{E}\left[L_{j-2}(X(\dot{\omega}), Y) \int_{\mathbb{R}} \gamma(u, Y_j)\, \Lambda(X_{j-2}(\dot{\omega}), du)\right]}$$

$$= \frac{\dot{E}\left[L_{j-2}(X(\dot\omega), Y) \int_{\mathbb{R}} \xi(u)\gamma(u, Y_j)\Lambda(X_{j-2}(\dot\omega), du)\right] / \dot{E}[L_{j-2}(X(\dot\omega), Y]}{\dot{E}\left[L_{j-2}(X(\dot\omega), Y) \int_{\mathbb{R}} \gamma(u, Y_j)\Lambda(X_{j-2}(\dot\omega), du)\right] / \dot{E}[L_{j-2}(X(\dot\omega), Y)]}$$

$$= \frac{E\left[\int_{\mathbb{R}} \xi(u)\gamma(u, Y_j)\Lambda(X_{j-2}, du) \mid \mathcal{F}^Y_{j-2}\right]}{E\left[\int_{\mathbb{R}} \gamma(u, Y_j)\Lambda(X_{j-2}, du) \mid \mathcal{F}^Y_{j-2}\right]} = (*)$$

Denote by $\pi_j(dx)$ the regular conditional distribution of X_j given $\mathcal{F}^Y_j$. Then, the last equation reads

$$\int_{\mathbb{R}} \xi(x)\pi_j(dx) = E[\xi(X_j) \mid \mathcal{F}^Y_j] \overset{(*)}{=} \frac{\int_{\mathbb{R}}\int_{\mathbb{R}} \xi(u)\gamma(u, Y_j)\Lambda(x, du)\pi_{j-2}(dx)}{\int_{\mathbb{R}}\int_{\mathbb{R}} \gamma(u, Y_j)\Lambda(x, du)\pi_{j-2}(dx)}$$

(Definition of $\pi_{j-2}(dx)$; integral with respect to u; integral with respect to x)

(Fubini in the double integral in the numerator)

$$= \int_{\mathbb{R}} \xi(u) \frac{\int_{\mathbb{R}} \gamma(u, Y_j)\Lambda(x, du)\pi_{j-2}(dx)}{\int_{\mathbb{R}}\int_{\mathbb{R}} \gamma(u, Y_j)\Lambda(x, du)\pi_{j-2}(dx)}$$

and, by arbitrariness of ξ, (2.14) follows.

To prove (2.15), we can write

$$\int_{\mathbb{R}} \xi(x)\pi_0(dx) = E[\xi(X_0) \mid \mathcal{F}^Y_0]$$

$$= \frac{\dot{E}[\xi(X_0(\dot\omega))\gamma(X_0(\dot\omega), Y_0)]}{\dot{E}[\gamma(X_0(\dot\omega), Y_0)]} = \frac{\int_{\mathbb{R}} \xi(x)\gamma(x, Y_0)p(dx)}{\int_{\mathbb{R}} \gamma(x, Y_0)p(dx)}$$

which yields (2.15).

2.3. The nonlinear filter by the reference measure approach

Here we will give an alternative proof to (3.14) using the reference measure method. It requires stronger assumptions on (X, Y) but gives additional insight on the structure of (3.14) and it is useful in the continuous time case. It is based on the following lemma

Lemma 2.20

Let $(\Omega, \mathcal{F})$ be a measurable space and let P and $\tilde{P}$ be two equivalent probability measures on $\mathcal{F}$, i.e., $P \sim \tilde{P}$. Denote by $E[\cdot|\mathcal{G}]$ and $\tilde{E}[\cdot|\mathcal{G}]$ the conditional expectations with respect to $\mathcal{G} \subseteq \mathcal{F}$ under P and $\tilde{P}$. Then, for any $X \in L^1$

$$E[X|\mathcal{G}] = \frac{\tilde{E}\left[X \frac{dP}{d\tilde{P}} \middle| \mathcal{G}\right]}{\tilde{E}\left[\frac{dP}{d\tilde{P}} \middle| \mathcal{G}\right]} \tag{2.16}$$

Proof

First we prove that the r.h.s. of (2.16) is well defined P-a.s.

$$P\left(\tilde{E}\left[\frac{dP}{d\tilde{P}} \middle| \mathcal{G}\right] = 0\right) = E\left[1_{\left\{\tilde{E}\left[\frac{dP}{d\tilde{P}} \middle| \mathcal{G}\right] = 0\right\}}\right]$$

$$= \tilde{E}\left[1_{\{\tilde{E}[\frac{dP}{d\tilde{P}}|\mathcal{G}]=0\}} \frac{dP}{d\tilde{P}} \right]$$

$$= \tilde{E}\left[1_{\{\tilde{E}[\frac{dP}{d\tilde{P}}|\mathcal{G}]=0\}} E[\frac{dP}{d\tilde{P}}|\mathcal{G}] \right] = 0$$

($\mathcal{G}$-measurable)

clearly the r.h.s. of (2.16) is $\mathcal{G}$-measurable and for any $A \in \mathcal{G}$

$$E\left[\left(X - \frac{\tilde{E}[X \frac{dP}{d\tilde{P}}|\mathcal{G}]}{\tilde{E}[\frac{dP}{d\tilde{P}}|\mathcal{G}]} \right) 1_A \right]$$

$$= \tilde{E}\left[\left(X - \frac{\tilde{E}[X \frac{dP}{d\tilde{P}}|\mathcal{G}]}{\tilde{E}[\frac{dP}{d\tilde{P}}|\mathcal{G}]} \right) 1_A \frac{dP}{d\tilde{P}} \right]$$

$$= \tilde{E}[X \frac{dP}{d\tilde{P}} 1_A] - \tilde{E}\left[\frac{\tilde{E}[X \frac{dP}{d\tilde{P}}|\mathcal{G}]}{\tilde{E}[\frac{dP}{d\tilde{P}}|\mathcal{G}]} 1_A \tilde{E}[\frac{dP}{d\tilde{P}}|\mathcal{G}] \right]$$

$= 0$ which verifies the claim.

This lemma suggests the following way of computing the conditional distribution: Find a reference measure $\tilde{P}$, equivalent to P, under which the computation of the conditional expectation would be easier (usually one chooses $\tilde{P}$ such that X is independent of $\mathcal{G}$) and use (2.16).

We will assume the following structure for the observation process (all other assumptions remain the same)

- $Y_j = h(X_j) + \xi_j$, where $h : \mathbb{R} \to \mathbb{R}$ is Borel measurable and $\xi = \{\xi_j\}_{j \geq 0}$ is an i.i.d. sequence of r.v., independent of X, such that ξ_1 has positive density $q(u) > 0$ with respect to the Lebesgue measure.

$$P(\xi_1 \leq u) = \int_{-\infty}^{u} q(s)\, ds .$$

Let's verify the claim of Theorem 2.19 under this assumption.

Define $\mathcal{F}_j := \mathcal{F}_j^X \vee \mathcal{F}_j^Y = \mathcal{F}_j^X \vee \mathcal{F}_j^\xi$,

$$\Phi_j(X, Y) := \prod_{i=0}^{j} \frac{q(Y_i)}{q(Y_i - h(X_i))} > 0, \tag{2.17}$$

and

$$\frac{d\tilde{P}}{dP}(\omega) := \Phi_j(X(\omega), Y(\omega)).$$

Hence $\tilde{P}$ is a probability measure (on $\mathcal{F}_j$) which is absolutely continuous with respect to the restriction of P on $\mathcal{F}_j$. Indeed

$$\tilde{P}(\Omega) = E[\Phi_j(X, Y)] = E\left[\prod_{i=0}^{j} \frac{q(Y_i)}{q(Y_i - h(X_i))}\right]$$

$$= E\left[\prod_{i=0}^{j} \frac{q(h(X_i) + \xi_i)}{q(\xi_i)}\right]$$

$$\left(\begin{matrix}\text{independence}\\ \text{of } X \text{ and } \xi\end{matrix}\right) = E\Big[\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{i=0}^{j} \frac{q(h(X_i)+u_i)}{q(u_i)} \prod_{i=0}^{j} q(u_i)\, du_0 \cdots du_j \Big] \qquad (2.27)$$

$$= E\Big[\int_{\mathbb{R}} \cdots \int_{\mathbb{R}} \prod_{i=0}^{j} q(h(X_i)+u_i)\, du_0 \cdots du_j \Big]$$

$$\left(\text{Fubini}\right) = E\Big[\prod_{i=0}^{j} \int_{\mathbb{R}} q(h(X_i)+u_i)\, du_i \Big]$$

$$\left(\begin{matrix} v_i = h(X_i)+u_i \\ dv_i = du_i \end{matrix}\right) = E\Big[\prod_{i=0}^{j} \underbrace{\int_{\mathbb{R}} q(v_i)\, dv_i}_{\| \; 1} \Big] = 1$$

Under the measure $\tilde{P}$, the random process (X, Y) "looks" absolutely different:

i) the distribution of the process $Y = (Y_0, \dots, Y_j)$ coincides with the distribution of $\xi = (\xi_0, \dots, \xi_j)$ under P.

ii) the distribution of the process $X = (X_0, \dots, Y_j)$ is the same under both measures P and $\tilde{P}$.

iii) the process X and Y are independent under $\tilde{P}$.

Let's check this properties. The proof is based on the following computation.

Let ψ, ϕ be Borel measurable, bounded functions from $\mathbb{R}^{j+1}$ to $\mathbb{R}$. Then,

$$\tilde{E}\big[\psi(\underbrace{X_0, \dots, X_j}_{X})\, \phi(\underbrace{Y_0, \dots, Y_j}_{Y})\big] = E\big[\psi(X)\, \phi(Y)\, \Phi_j(X, Y)\big]$$

$$= E\Big[\psi(X)\, \phi(Y) \prod_{i=0}^{j} \frac{q(Y_i)}{q(Y_i - h(X_i))} \Big]$$

$$= E\Big[\psi(X)\,\phi(h(X_0)+\xi_0,\dots,h(X_j)+\xi_j)\prod_{i=0}^{j}\frac{q(h(X_i)+\xi_i)}{q(\xi_i)}\Big] \quad (2.28)$$

$$= E\Big[\psi(X)\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\phi(h(X_0)+u_0,\dots,h(X_j)+u_j)\prod_{i=0}^{j}\frac{q(h(X_i)+u_i)}{q(u_i)}\prod_{i=0}^{j}q(u_i)\,du_0\cdots du_j\Big]$$

$$= E\Big[\psi(X)\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\phi(h(X_0)+u_0,\dots,h(X_j)+u_j)\prod_{i=0}^{j}q(h(X_i)+u_i)\,du_0\cdots du_j\Big]$$

$$= E[\psi(X)]\int_{\mathbb{R}}\cdots\int_{\mathbb{R}}\phi(u_0,\dots,u_j)\prod_{i=0}^{j}q(u_i)\,du_0\cdots du_j$$

$$= E[\psi(X)]\,E[\phi(S)]$$

Now:

i) Follows by taking $\psi\equiv 1$ and the arbitrariness of ϕ

ii) Follows by taking $\phi\equiv 1$ and the arbitrariness of ψ.

iii) If i) and ii) hold

$$\tilde{E}[\psi(X)\,\phi(Y)] = E[\psi(X)]\,E[\phi(S)]$$

$$= \tilde{E}[\psi(X)]\,\tilde{E}[\phi(Y)],$$

(ii) (i)

which, due to the arbitrariness of ψ and ϕ, yields iii)

By Lemma 2.20, for any bounded measurable function g

$$E[g(X_j)\mid\mathcal{F}_j^Y] = \frac{\tilde{E}[g(X_j)\,\Phi_j^{-1}(X,Y)\mid\mathcal{F}_j^Y]}{\tilde{E}[\Phi_j^{-1}(X,Y)\mid\mathcal{F}_j^Y]}$$

$$(2.18) = \frac{\dot{E}[g(X_j(\omega))\,\Phi_j^{-1}(X(\dot{\omega}),Y(\omega))]}{\dot{E}[\Phi_j^{-1}(X(\dot{\omega}),Y(\omega))]},$$

where $\frac{dP}{d\tilde{P}} = \Phi_j^{-1}(X,Y)$. (2.18) is due to the independence of X and Y under $\tilde{P}$ (we use notations of Remark (2.16))

Next,

$$\dot{E}\left[g(X_j(\dot{\omega})) \Phi_j^{-1}(X(\dot{\omega}), Y(\omega)) \right]$$

$$= \dot{E}\left[g(X_j(\dot{\omega})) E\left[\Phi_j^{-1}(X(\dot{\omega}), Y(\omega) / \sigma(X_j) \right] \right]$$

$$= \int_{\mathbb{R}} g(u) \dot{E}\left[\Phi_j^{-1}(X(\dot{\omega}), Y(\omega)) \mid X_j = u \right] P_{X_j}(du)$$

$$=: \int_{\mathbb{R}} g(u) \ell_j(du).$$

On the other hand

$$\dot{E}\left[g(X_j(\dot{\omega})) \Phi_j^{-1}(X(\dot{\omega}), Y(\omega)) \right]$$

$$= \dot{E}\left[\Phi_{j-2}^{-1}(X(\dot{\omega}), Y(\omega)) \dot{E}\left[g(X_j(\dot{\omega})) \frac{q(Y_j(\omega) - h(X_j(\dot{\omega})))}{q(Y_j(\omega))} / \mathcal{F}_{j-2}^X \right] \right]$$

$$= \dot{E}\left[\Phi_{j-2}^{-1}(X(\dot{\omega}, Y(\omega)) \int_{\mathbb{R}} g(u) \frac{q(Y_j(\omega) - h(u))}{q(Y_j(\omega))} \Lambda(X_{j-2}(\dot{\omega}), du) \right]$$

$$= \int_{\mathbb{R}} g(u) \int_{\mathbb{R}} \frac{q(Y_j(\omega) - h(u))}{q(Y_j(\omega))} \Lambda(s, du) \ell_{j-2}(ds)$$

Since g is arbitrary, we get the recursion

$$\ell_j(du) = \int_{\mathbb{R}} \frac{q(Y_j(\omega) - h(u))}{q(Y_j(\omega))} \Lambda(s, du) \ell_{j-2}(ds) \tag{2.19}$$

Finally, by (2.18)

$$E[g(X_j) \mid \mathcal{F}_j^Y] = \frac{\int_{\mathbb{R}} g(u)\, \ell_j(du)}{\int_{\mathbb{R}} \ell_j(du)}$$

and we can compute $\pi_j(dx)$ from Theorem 2.1 by normalizing

$$\pi_j(dx) = \frac{\ell_j(dx)}{\int_{\mathbb{R}} \ell_j(du)}.$$

Note that in this case $\gamma(u, y_j) = q(y_j - h(u))$.

Remark 2.21

The previous computations suggests that, in order to find $\pi_j(dx)$, we can first solve the linear recursion for $\ell_j(du)$ in equation (2.19). The measure $\ell_j(du)$ is called the unnormalized conditional distribution.

4. The curse of dimensionality and finite dimensional filters

Equations (2.14) (or its unnormalized counterpart) are not very practical. At each time step they require at least two integrations. We are computing (random) measures, which are infinite dimensional objects.

Definition 2.22

The filter is called finite dimensional with respect to a function f, if the right hand side of

$$E[f(X_j) \mid \mathcal{F}_j^Y] = \int_{\mathbb{R}} f(x)\, \pi_j(dx)$$

can be parametrized by a finite number of sufficient statistics of Y.

The evaluation of π_j can be infinite-dimensional, while the integral of π_j against a specific function f may admit a finite dimensional filter.

There is no easy way to determine whether a nonlinear filter is finite dimensional. In fact few dimensional filters are known.

4.2. Hidden Markov Models (HMM)

Suppose $X = \{X_j\}_{j \geq 0}$ is a Markov chain with a finite state space $S = \{a_1, \dots, a_d\}$. Then its Markov kernel is identified with the matrix Λ of transition probabilities

$$\Lambda_{lm} \quad P(X_j = a_m \mid X_{j-1} = a_l).$$

Let p_0 be the initial distribution of X, i.e.,

$$p_0(l) = P(X_0 = a_l), \quad l = 1, \dots, d.$$

Suppose that the observation process $Y = (Y_j)_{j \geq 0}$ satisfies

$$P(Y_j \in A \mid \mathcal{F}_j^X \vee \mathcal{F}_{j-1}^Y) = \int_A \nu_l(du), \quad l = 1, \dots, d.$$

Note that each $\nu_l(du) \ll \nu(du) := \sum_{m=1}^{d} \nu_m(du)$, so no generality is lost by assuming that $\nu_l(du) = f_l(u)\,\nu(du)$ for some σ-finite measure ν on $\mathcal{B}(\mathbb{R})$ and densities f_l.

The conditional distribution $\pi_j(dx) \ll \sum_{i=1}^{d} \delta_{a_i}$ and it can be identified with the density π_j, which is just a vector of conditional probabilities $P(X_j = a_l \mid \mathcal{F}_j^Y)$, $l = 1, \dots, d$.

Then formula 2.14 writes as

$$\pi_j = \frac{\operatorname{diag}(f(Y_j|)\, \Lambda^T \pi_{j-1}}{|\operatorname{diag}(f(Y_j|)\, \Lambda^T \pi_{j-1}|}, \quad \text{where } f(Y_j) = (f_1(Y_j), \dots, f_d(Y_j))^T, \ \pi_0 = p_0$$

and $|x| = \sum_{i=1}^{d} |x_i| \quad x \in \mathbb{R}^d$.

Alternatively, one can solve the unnormalised equation

$$\ell_j = \operatorname{diag}(f(Y_j|)\, \Lambda^T \ell_{j-1}, \quad \ell_0 = p_0$$

and then $\pi_j = \ell_j / |\ell_j|$.

4.2. The linear Gaussian case : Kalman-Bucy filter revisited

The Kalman-Bucy filter is a very special case thanks to the properties of Gaussian random vectors.

Definition 2.22

A random vector X, with values in $\mathbb{R}^d$, is Gaussian if

$$E\left[\exp\left\{ i\, \lambda^T X \right\}\right] = \exp\left\{ i\, \lambda^T m - \frac{1}{2}\, \lambda^T K\, \lambda \right\}$$

for all $\lambda \in \mathbb{R}^d$, for a vector m and a nonnegative definite matrix K.

One can check that $m = E[X]$ and $K = \mathrm{cov}(X)$.
Moreover, if the characteristic function of a random vector is the exponential of a quadratic form, this vector is necessarily Gaussian.
Gaussian vectors plays a very special role in probability theory. For stochastic filtering, the following properties of Gaussian vectors are relevant.

Lemma 2.23

Let the random vectors $X \in \mathbb{R}^m$ and $Y \in \mathbb{R}^n$ be such that (X, Y) form a Gaussian vector in $\mathbb{R}^{m+n}$.
Then,

(1) Any affine transformation of (X, Y) is a Gaussian random variable. In particular $Z = b + AX$, with a vector b and a matrix A, is a Gaussian vector with $E[Z] = b + A\,E[X]$ and $\mathrm{cov}(Z) = A\,\mathrm{cov}(X)\,A^T$

(2) If X and Y are orthogonal, they are independent
(the opposite direction is obvious)

(3) The regular conditional distribution of X given Y is Gaussian P-a.s., moreover

$$E[X|Y] = \hat{E}[X|Y]$$

$$cov(X|Y) := E[(X - E[X|Y])(X - E[X|Y])^T | Y]$$

$$= cov(X) - cov(X,Y)\, cov(Y)^{\oplus}\, cov(Y,X)$$

Remark 2.24

Note that $\hat{E}[X|Y]$ is an (affine) function of Y. But $cov(X,Y)$ is not a function of Y, in the sense that it does not depend on the particular value of Y that we observe.

Proof (of Theo. 2.23)

(1) Let $Z = (X,Y) \in \mathbb{R}^{n+m}$ and $M = b + BZ \in \mathbb{R}^k$

$B = (B_X, B_Y) \in \mathbb{R}^{k \times (n+m)}$ $\quad b \in \mathbb{R}^k$

$$E[M] = E[b + BZ] = b + BE[Z] = b + B_X E[X] + B_Y E[Y]$$

$$Cov[M] = cov(BZ) = B\, cov(Z)\, B^T$$

$$= (B_X \; B_Y) \begin{pmatrix} cov(X) & cov(X,Y) \\ cov(Y,X) & cov(Y) \end{pmatrix} \begin{pmatrix} B_X \\ B_Y \end{pmatrix}$$

Let $\lambda \in \mathbb{R}^k$

$$(*) \quad E[e^{i\lambda^T M}] = e^{i\lambda^T (b + BE[Z])}\, E[\exp(i(\lambda^T B)(Z - E[Z]))]$$

(Z is Gaussian) $\quad = e^{i\lambda^T (b + BE[Z])} \exp\left(-\frac{1}{2} \lambda^T B\, cov(Z)\, B^T \lambda\right)$

Using Def. 2.22 we get that M is Gaussian with

mean $E[W]$ and covariance matrix $Cov(W)$.

Taking $B_y = 0$ and $B_x = A$ we get that

$$Z = b + AX \sim \mathcal{N}(b + AE[X], A\,cov(X)A^T)$$

(2) Let $\lambda_x \in \mathbb{R}^m$, $\lambda_y \in \mathbb{R}^n$, then

$$E[e^{i\lambda^T(X,Y)}] = e^{i(\lambda_x E[X] + \lambda_y E[Y])} e^{-\frac{1}{2}(\lambda_x, \lambda_y)^T \begin{pmatrix} cov(X) & cov(X,Y) \\ cov(Y,X) & cov(Y) \end{pmatrix} \begin{pmatrix} \lambda_x \\ \lambda_y \end{pmatrix}}$$

(*) with $B = Id \in \mathbb{R}^{(m+n)\times(m+n)}$

Orthogonality $= e^{i\lambda_x E[X] - \frac{1}{2}\lambda_x^t cov(X)\lambda_x} \cdot e^{i\lambda_y E[Y] - \frac{1}{2}\lambda_y^T cov(Y)\lambda_y}$

$\Rightarrow cov(X,Y) = 0$, $cov(Y,X) = 0$

$\Rightarrow$ X and Y are independent.

(3) Recall that $X - \hat{E}[X|Y]$ is orthogonal to Y, and thus by (2), they are also independent.

Then,

$$E[\exp(i\lambda_x^T(X - \hat{E}[X|Y]))|Y]$$

$$= E[\exp(i\lambda_x^T(X - \hat{E}[X|Y]))]$$

and, on the other hand,

$$E[\exp(i\lambda_x^T(X - \hat{E}[X|Y]))|Y]$$

$$= e^{-i\lambda_x^T \hat{E}[X|Y]} E[\exp(i\lambda_x^T X)|Y]$$

So

$$E[e^{i\lambda_x^T X}|Y] = \exp(i\lambda_x^T \hat{E}[X|Y]) \underbrace{E[\exp(i\lambda_x^T(X - \hat{E}[X|Y]))]}_{(**)}$$

Since $X - \hat{E}[X|Y]$ is in the linear span of (X,Y) we have that

$$(**) = \exp(i\lambda_x^T \underbrace{E[(X - \hat{E}[X|Y])]}_{0} - \frac{1}{2}\lambda_x^T cov(X - \hat{E}[X|Y])\lambda_x)$$

$$= \exp\left(-\frac{1}{2}\,\lambda_x^T \operatorname{cov}\left(X - \hat{E}[X|Y]\right)\lambda_x\right)$$

and

$$E\left[e^{i\lambda_x^T X} \mid Y\right] = \exp\left(i\,\lambda_x^T \underbrace{\hat{E}[X|Y]}_{E[X|Y]} - \frac{1}{2}\,\lambda_x^T \underbrace{\operatorname{cov}(X - \hat{E}[X|Y])}_{\operatorname{cov}(X - E[X|Y])}\lambda_x\right)$$

which is what we wanted to prove.

Consider the linear model (1.10) and (1.11), where the sequences ξ and ε are Gaussian, as well as the initial conditions (X_0, Y_0)

Then, the process (X, Y) are Gaussian and by Lemma 2.23, the conditional distribution of X_j given $\mathcal{F}_j^Y$ is Gaussian too.

Moreover, the parameter (mean and covariance) are governed by the Kalman-Bucy filter equations in Theorem 1.5.

Remark 2.25

The recursions in Theorem 1.5 can be found using the recursions for the nonlinear problem.

In particular for any Borel measurable ξ, such that $E[|\xi(X_j)|] < +\infty$ we have that

$$E[\xi(X_j) \mid \mathcal{F}_j^Y] = \int_{\mathbb{R}} f(u)\,\frac{1}{\sqrt{2\pi P_j}}\exp\left(-\frac{(u - \hat{X}_j)^2}{2P_j}\right)du,$$

where P_j and $\hat{X}_j$ are generated by the Kalman-Bucy equations

More models for which finite dimensional filters are known, but their practical applicability is usually limited.