Stochastic Machine Learning
Chapter 04 - Filtering II

Thorsten Schmidt

Abteilung für Mathematische Stochastik

www.stochastik.uni-freiburg.de
thorsten.schmidt@stochastik.uni-freiburg.de

SS 2024

Filtering - general theory

- To simplify the notation we consider the following explicit form of the setup.
- Let E and F be two polish spaces with the associated Borel σ-algebras. Assume that the signal process $X = (X_n)_{n\geq 0}$ takes values in E and the observation $Y = (Y_n)_{n\geq 0}$ takes values in F with $Y_0 = 0$.
- assume measurable mappings $a : E \times E' \to E$, $A : F \times F' \to F$ are given and that $(\xi_n)_{n\geq 0}$ and $(\eta_n)_{n\geq 1}$ are sequences of globally independent random variables, taking values in the Polish spaces E' and F', respectively.
- We assume that for all $n \geq 1$

$$\begin{aligned} X_n &= a(X_{n-1}, \xi_n) \\ Y_n &= A(X_n, Y_{n-1}, \eta_n). \end{aligned} \tag{1}$$

We denote the initial distribution of X_0 by μ_0.

Finite state space

- Assume, $A(x, y, \eta) = A(x, \eta)$, i.e. Y_n does not depend on Y_{n-1}.
- Denote

$$\begin{aligned} \pi_n(x) &:= \mathbb{P}(X_n = x | Y_1, \ldots, Y_n), \\ P_X^n(x_{n-1}, x) &:= \mathbb{P}(X_n = x | X_{n-1}), \\ P_Y^n(x, y) &:= \mathbb{P}(Y_n = y | X_n = x). \end{aligned}$$

- By Bayes' formula we obtain that

$$\begin{aligned} \mathbb{P}(X_n = x | Y_n = y, Y_1, \ldots, Y_{n-1}) &= \frac{\mathbb{P}(X_n = x, Y_n = y | Y_1, \ldots, Y_{n-1})}{\mathbb{P}(Y_n = y | Y_1, \ldots, Y_{n-1})} \\ &= \frac{\mathbb{P}(X_n = x | Y_1, \ldots, Y_{n-1}) \cdot \mathbb{P}(Y_n = y | X_n = x, Y_1, \ldots, Y_{n-1})}{\mathbb{P}(Y_n = y | Y_1, \ldots, Y_{n-1})}. \end{aligned} \quad (2)$$

- We have,

$$\mathbb{P}(Y_n = y | X_n = x, Y_1, \ldots, Y_{n-1}) = \mathbb{P}(Y_n = y | X_n = x) = P_Y^n(x, y)$$

- and

$$\begin{aligned}
\mathbb{P}(X_n = x | Y_1, \ldots, Y_{n-1}) &= \mathbb{E}\big[\mathbb{P}(X_n = x | X_{n-1}) | Y_1, \ldots, Y_{n-1}\big] \\
&= \sum_{x_{n-1} \in E} P_X^n(x_{n-1}, x)\pi_{n-1}(x) \\
&\propto \sum_{x_{n-1} \in E} P_X^n(x_{n-1}, x)\sigma_{n-1}(x)
\end{aligned}$$

  for all[1] $\sigma_{n-1}(x) \propto \pi_{n-1}(x)$.

- We obtain that the unnormalized conditional probabilities satisfy the recurrence relation

$$\sigma_n(x) = P_Y^n(x, Y_n) \cdot \sum_{x_{n-1} \in E} P_X^n(x_{n-1}, x)\sigma_{n-1}(x). \tag{3}$$

- The conditional distribution can be computed by

$$\pi_n(x) = \frac{\sigma_n(x)}{\sum_{x' \in E} \sigma_n x}.$$

  (Equivalent to (2))

[1]We write $f(x) \propto g(x)$ if there exists a constant C such that $f(x) = Cg(x)$.

Full generality

- Let $\mathscr{G}$ be a sub-σ-algebara of $\mathscr{F}$ and X a random variable with $E[\| X \|] < \infty$. Then any version Z of $E[X|\mathscr{G}]$ satisfies

$$E[X\mathbb{1}_G] = E[Z\mathbb{1}_G], \qquad \text{for all } G \in \mathscr{G}.$$

- Then, we can define the signed measure Q on $(\Omega, \mathscr{G})$ by

$$Q(G) := E[X\mathbb{1}_G], \qquad G \in \mathscr{G}.$$

- This measure is absolutely continuous with respect to P, as $P(G) = 0$ implies that $Q(G) = 0$, i.e. $P \ll Q$. By the Radon-Nikodym theorem there exists a $\mathscr{G}$-measurable function Z such that

$$Q(G) = \int_G Z(\omega)P(d\omega), \qquad G \in \mathscr{G}.$$

- We write simply

$$ZdP = dQ$$

- Then Then Z is a version of the conditional expectation $E[X|\mathscr{G}]$. Indeed, for all $G \in \mathcal{G}$,

$$E[Z\mathbb{1}_G] = \int_G ZdP = Q(G) = E[X\mathbb{1}_G].$$

- If the σ-algebra $\mathscr{G}$ is generated by a random variable, for example $\mathscr{G} = \sigma(Y)$, then $Z = Z(Y)$ is given by a measurable function of Y.

The following proposition gives us the right generalization of Bayes' formula for our purposes. By $P_X(A) := P(X \in A)$ we denote the distribution of X.

Proposition

Consider two random variables X and Y with state spaces E and F, respectively, and assume there exists a measure Q and a measurable function $\Lambda : E \times F \to \mathbb{R}$, such that for P_X-almost all $x \in E$

$$P(Y \in B | X = x) = \int_B \Lambda(x, y) Q(dy), \qquad B \in \mathscr{B}(F). \tag{4}$$

Then, for any measurable and bounded function $f : E \to \mathbb{R}$,

$$E[f(X)|Y] \propto \int f(x)\Lambda(x, Y) P_X(dx), \quad P\text{-a.s.} \tag{5}$$

Proof.

Exercise. □

Examples

- Assume that X and Y have common density $f(x, y)$.
- Then $f_X(x) = \int_F f(x, y)dy$, $f_Y(y) = \int_E f(x, y)dx$.
- Let $f(y|x)$ denote the conditional distribution of y given x and we have

$$P(Y \in B|X = x) = \int_B f(y|x)dy.$$

- Hence,

$$E[g(X)|Y] \propto \int g(x)f(y|x)P_X(dx)$$

- On the other side,

$$\begin{aligned} E[g(X)] &= \int_E \int_F g(x)f(x, y)dx\,dy \\ &= E\big[E[g(x)|Y]\big] = C\int_E \int_F g(x)f(y|x)f_X(x)dx f_Y(y)dy \end{aligned}$$

and we recover that $f(x|y) \propto f(x, y)$.

- Since $\int_f (x|y)dx = 1$,

$$1 = C\int_E f(x, y)dx = Cf_Y(y)$$

and so

$$f(x|y) = \frac{f(x, y)}{f_Y(y)}.$$

- Assume that X has countable state space E and Y has a density.
- In this case we have that

$$P(X = x, Y \in B) = \int_B f(x, y) dy.$$

- Regarding assumption (4) we obtain that for $P(X = x) > 0$,

$$P(Y \in B | X = x) = \int_B \frac{f(x, y)}{P(X = x)} dy.$$

- Applying the proposition gives that

$$P[X = x | Y] \propto \frac{f(x, y)}{P(X = x)} \propto f(x, y).$$

- The conditional distribution of x given Y then computes to

$$P(X = x | Y) = \frac{f(x, y)}{\sum_{x' \in X} f(x', y)} = \frac{f(x, y)}{f(y)}.$$

- Similarly, in the case where Y has discrete state space and X has a density we obtain that

$$f(x|y) = \frac{f(x, y)}{P(Y = y)}.$$

Define for all $A \in \mathscr{B}(E)$ and $B \in \mathscr{B}(F)$

$$\begin{aligned} P^n_{X,Y}(A, B, X_{n-1}, Y_{n-1}) &= P(X_n \in A, Y_n \in B | X_{n-1}, Y_{n-1}) \\ P^n_Y(B, X_n, Y_{n-1}) &= P(Y_n \in B | X_n, Y_{n-1}) \\ P^n_X(A, X_{n-1}) &= P(X_n \in A | X_{n-1}), \\ \mathcal{Y}_n &= \sigma(Y_1, \ldots, Y_n). \end{aligned}$$

Assumption

Assume that for each n there exist a regular conditional distribution Q_n w.r.t. the σ-algebra $\mathcal{Y}_{n-1}$, a $\mathcal{Y}_n \vee \sigma(X_n)$-measurable function Λ_n, and a $\mathcal{Y}_{n-1} \vee \sigma(X_n)$-measurable function C^n such that

$$\begin{aligned} &\mathbb{P}(Y_n \in B | X_n, Y_1, \ldots, Y_{n-1})(\omega) \\ &\qquad = \int_B C^n(Y_1(\omega), \ldots, Y_{n-1}(\omega), y) \Lambda^n(X_n(\omega), Y_1(\omega), \ldots, Y_{n-1}(\omega), y) dQ^n(\omega, dy) \end{aligned}$$

for almost all $\omega \in \mathcal{Y}_{n-1} \vee \sigma(X_n)$.

Theorem

Suppose Assumption 2 holds. Then there exists a sequence $(\sigma_n)_{n \geq 0}$ of random measures such that

(i) $\sigma_0(A) \propto \mu_0(A)$,

(ii) *for each $n > 0$, σ satisfies the recurrence relation*

$$\sigma_n(dx) = \Lambda^n(x, Y_1, \ldots, Y_n) \int_E P^n_X(dx|x') \sigma_{n-1}(dx'),$$

(iii) *and $\pi_n(A) = \sigma_n(A)/\sigma_n(\Omega)$.*

Essentially, the transition from σ_{n-1} to σ_n contains the following two steps:

- **Prediction:** the first step is to predict the movement from any x_{n-1} to values x' by applying the transition kernel of the Markov chain P_X and averaging over all possible x_{n-1} whith the conditional distribution from the previous step:

$$\sigma_{n-1}(dx) \to p_n(dx) = \int_E P_X^n(dx|x')\sigma_{n-1}(dx'),$$

- **Correction:** the second step takes the additional new information into account by an appropriate change of measure:

$$p_n(dx) \to \sigma_n(dx) = \Lambda^n(x, Y_1, \ldots, Y_n)p_n(dx).$$

The correction step only takes the simple Bayes' formula in (2) to this more general level.

Proof

- We proceed by induction. The step $n = 1$ is obtained directly with the initial distribution of X_0, μ_0 together with $\mathcal{Y}_0 = \{\emptyset, \Omega\}$ as $Y_0 = 0$. Moreover,

$$\pi_0 = \mu_0 = \frac{\sigma_0}{\sigma_0(\Omega)}$$

  for any $\sigma_0 \propto \mu_0$.
- Assume the claim holds for $n - 1$.
- Under Assumption 2 we may define a regular conditional distribution of X_n and Y_n w.r.t. $\mathcal{Y}_{n-1} \vee X_{n-1}$ by

$$R^n(\omega, A, B) := \int \mathbb{1}_A(x_n) P_X^n(dx_n | X_{n-1}(\omega)) \int \mathbb{1}_B(y_n) Q^n(\omega, dy_n),$$

  $\omega \in \Omega$.
- We have that

$$E^{R^n}[e^{i\langle u, X_n\rangle + i\langle v, Y_n\rangle} | X_{n-1} \vee \mathcal{Y}_{n-1}] = \int e^{i\langle u, x\rangle} P_X^n(dx | X_{n-1}(\omega)) \int e^{i\langle v, y\rangle} Q^n(\omega, dy_n)$$

  such that X_n and Y_n are independent under R^n, conditionally on $\sigma(X_{n-1}) \vee \mathcal{Y}_{n-1}$.
- In particular,

$$E^{R^n}[f(X_n) | X_{n-1}, Y_1, \ldots, Y_{n-1}] = E^{R^n}[f(X_n) | X_{n-1}, Y_1, \ldots, Y_n]$$

  for all integrable functions f.
- Denote $\tilde{\Lambda} = C\Lambda$.

- By Assumption 2

$$
\begin{aligned}
P(X_n \in A, &Y_n \in B | X_{n-1}, \mathcal{Y}_{n-1}) \\
&= E\Big[\mathbb{1}_A(X_n) P\big(Y_n \in B | X_n, X_{n-1}, \mathcal{Y}_{n-1}\big) | X_{n-1}, \mathcal{Y}_{n-1}\Big] \\
&= E\Big[\mathbb{1}_A(X_n) \int_B \tilde{\Lambda}^n(X_n, Y_1, \dots, Y_{n-1}, y) Q^n(\omega, dy) | X_{n-1}, \mathcal{Y}_{n-1}\Big] \\
&= \int_A \int_B \tilde{\Lambda}^n(x, Y_1, \dots, Y_{n-1}, y) Q^n(\omega, dy) P_X(dx | X_{n-1}, \mathcal{Y}_{n-1}),
\end{aligned}
$$

- Moreover,

$$
\begin{aligned}
P(X_n \in A, &Y_n \in B | \mathcal{Y}_{n-1}) \\
&= \int P(X_n \in A, Y_n \in B | X_{n-1} = x', \mathcal{Y}_{n-1})\, \pi_{n-1}(dx' | \mathcal{Y}_{n-1}) \\
&= \int \int_A \int_B \tilde{\Lambda}^n(x, Y_1, \dots, Y_{n-1}, y) Q^n(\omega, dy) P_X(dx | X_{n-1} = x', \mathcal{Y}_{n-1}) \\
&\qquad \pi_{n-1}(dx' | \mathcal{Y}_{n-1}) \qquad (6)
\end{aligned}
$$

and we can concentrate on the inner expression.

- We therefore may apply the generalized Bayes' formula and have to compute the Assumption (4).
- We first rewrite (4) in the following form
$$P(X \in A, Y \in B) = \int_A \int_B \Lambda(x, y)\, Q(dy)\, P_X(dx).$$
Comparing this with Equation (6) we obtain from Equation (5) that
$$\begin{aligned} &P(X_n \in A | Y_n, \mathcal{Y}_{n-1}) \\ &\propto \int_A \tilde{\Lambda}^n(x, Y_1, \ldots, Y_{n-1}, Y_n) P_X(dx | X_{n-1} = x', \mathcal{Y}_{n-1})\, \pi_{n-1}(dx' | \mathcal{Y}_{n-1}). \end{aligned}$$
- By induction, $\pi_{n-1} \propto \sigma_{n-1}$ such that we obtain
$$\begin{aligned} \pi_n(A) &\propto \int_A \tilde{\Lambda}^n(x, Y_1, \ldots, Y_n) P_X(dx|x') \sigma_{n-1}(dx') \\ &\propto \int_A \Lambda^n(x, Y_1, \ldots, Y_n) P_X(dx|x') \sigma_{n-1}(dx') = \sigma_n(A) \end{aligned}$$
and the claim follows.