Filtering in strong noise

P. Chigansky

Universitè du Maine, France

2007-03-22/ SAPS VI

Outline

Filtering of finite state Markov chains

Let the signal process X_t be a continuous time Markov chain:

- values in a finite set of real numbers $\mathbb{S} = \{a_1, ..., a_d\}$
- transition rates $\Lambda = \{\lambda_{ij}\}_{1 \le i,j \le d}$
- the initial distribution $\nu_i = \mathrm{P}(X_0 = a_i)$, $i = 1, ..., d$

The observation process Y_t is generated by

$$Y_t = \int_0^t h(X_s) ds + \sigma B_t, \quad t \ge 0,$$

where

- $h : \mathbb{S} \mapsto \mathbb{R}$ is a fixed known function
- B_t is a Brownian motion, independent of X
- $\sigma > 0$ is the noise intensity

Filtering of finite state Markov chains

Let the signal process X_t be a continuous time Markov chain:

- values in a finite set of real numbers $\mathbb{S} = \{a_1, ..., a_d\}$
- transition rates $\Lambda = \{\lambda_{ij}\}_{1\le i,j\le d}$
- the initial distribution $\nu_i = \mathrm{P}(X_0 = a_i)$, $i = 1, ..., d$

The observation process Y_t is generated by

$$Y_t = \int_0^t h(X_s)ds + \sigma B_t, \quad t \ge 0,$$

where

- $h : \mathbb{S} \mapsto \mathbb{R}$ is a fixed known function
- B_t is a Brownian motion, independent of X
- $\sigma > 0$ is the noise intensity

Filtering of finite state Markov chains

- The objective of the filtering problem is to calculate (recursively) the conditional probabilities ($i = 1, ..., d$)

$$\pi_t(i) := \mathrm{P}(X_t = a_i | \mathscr{F}_t^Y), \quad \text{where } \mathscr{F}_t^Y := \sigma\{Y_s, s \in [0, t]\}.$$

- Once the vector π_t is computed, it can be used to find various estimates of X_t given the trajectory $\{Y_s, s \le t\}$.

Filtering of finite state Markov chains

- The objective of the filtering problem is to calculate (recursively) the conditional probabilities ($i = 1, ..., d$)

$$\pi_t(i) := \mathrm{P}(X_t = a_i | \mathscr{F}_t^Y), \quad \text{where } \mathscr{F}_t^Y := \sigma\{Y_s, s \in [0,t]\}.$$

- Once the vector π_t is computed, it can be used to find various estimates of X_t given the trajectory $\{Y_s, s \le t\}$.

Filtering of finite state Markov chains

- The minimal probability of error:

$$\inf_{\xi \in L_\infty(\mathscr{F}_t^Y)} \mathsf{P}\big(\xi \neq X_t\big) = \mathsf{P}\big(\hat{X}_t^{MAP} \neq X_t\big) = 1 - \mathsf{E} \max_{a_i \in \mathbb{S}} \pi_t(i)$$

is attained by the Maximum a Posteriori Probability estimate

$$\hat{X}_t^{MAP} = \mathrm{argmax}_{a_i \in \mathbb{S}} \pi_t(i).$$

Filtering of finite state Markov chains

- The mean square error:

$$\inf_{\xi \in L_2(\mathscr{F}_t^Y)} \mathrm{E}\big(\xi - X_t\big)^2 = \mathrm{E}\big(\hat{X}_t^{MSE} - X_t\big)^2 =$$

$$\mathrm{E}\left[\sum_{i=1}^d a_i^2 \pi_t(i) - \Big(\sum_{i=1}^d a_i \pi_t(i)\Big)^2\right]$$

is attained by the conditional mean

$$\hat{X}_t^{MSE} = \sum_{i=1}^d a_i \pi_t(i).$$

The long time behavior

The vector π_t satisfies the Wonham SDE:

$$d\pi_t = \Lambda^* \pi_t dt + \sigma^{-2}\big(\mathrm{diag}(\pi_t) - \pi_t \pi_t^*\big) H\big(dY_t - h^* \pi_t dt\big), \quad \pi_0 = \nu,$$

where

- Λ is the transition rates matrix of the chain
- H is a column vector with the entries $h(a_i)$, $i = 1, ..., d$
- x^* is transposed of x
- $\mathrm{diag}(x)$ with $x \in \mathbb{R}^d$ is the diagonal matrix with x on the diagonal

The limit performance indeces

If X is an ergodic Markov chain, then process π_t is an ergodic diffusion on the simplex of probability vectors $\mathcal{S}^{d-1}$:

- a unique invariant measure φ on $\mathscr{B}(\mathcal{S}^{d-1})$ exists
- the convergence to φ is exponential
- the Law of Large Numbers holds

This implies that the limits of the performance indices exist

$$\mathcal{P}_{err} := \lim_{t\to\infty} \inf_{\xi\in L_\infty(\mathscr{F}_t^Y)} \mathrm{P}(\xi \neq X_t) = 1 - \int_{\mathcal{S}^{d-1}} (\max_i u_i)\varphi(du)$$

$$\mathcal{E}_{mse} := \lim_{t\to\infty} \inf_{\xi\in L_2(\mathscr{F}_t^Y)} \mathrm{E}\big(\xi - X_t\big)^2 = \int_{\mathcal{S}^{d-1}} \Big(u^* a^2 - (u^* a)^2\Big)\varphi(du)$$

The limit performance indeces

If X is an ergodic Markov chain, then process π_t is an ergodic diffusion on the simplex of probability vectors $\mathcal{S}^{d-1}$:

- a unique invariant measure φ on $\mathscr{B}(\mathcal{S}^{d-1})$ exists
- the convergence to φ is exponential
- the Law of Large Numbers holds

This implies that the limits of the performance indices exist

$$\mathcal{P}_{err} := \lim_{t\to\infty} \inf_{\xi\in L_\infty(\mathscr{F}_t^Y)} \mathsf{P}\big(\xi \neq X_t\big) = 1 - \int_{\mathcal{S}^{d-1}} (\max_i u_i)\varphi(du)$$

$$\mathcal{E}_{mse} := \lim_{t\to\infty} \inf_{\xi\in L_2(\mathscr{F}_t^Y)} \mathsf{E}\big(\xi - X_t\big)^2 = \int_{\mathcal{S}^{d-1}} \Big(u^* a^2 - (u^* a)^2\Big)\varphi(du)$$

The limit performance indeces

If X is an ergodic Markov chain, then process π_t is an ergodic diffusion on the simplex of probability vectors $\mathcal{S}^{d-1}$:

- a unique invariant measure φ on $\mathscr{B}(\mathcal{S}^{d-1})$ exists
- the convergence to φ is exponential
- the Law of Large Numbers holds

This implies that the limits of the performance indices exist

$$\mathcal{P}_{err} := \lim_{t\to\infty} \inf_{\xi \in L_\infty(\mathscr{F}_t^Y)} \mathsf{P}\big(\xi \neq X_t\big) = 1 - \int_{\mathcal{S}^{d-1}} (\max_i u_i)\varphi(du)$$

$$\mathcal{E}_{mse} := \lim_{t\to\infty} \inf_{\xi \in L_2(\mathscr{F}_t^Y)} \mathsf{E}\big(\xi - X_t\big)^2 = \int_{\mathcal{S}^{d-1}} \Big(u^* a^2 - (u^* a)^2\Big)\varphi(du)$$

The limit performance indeces

If X is an ergodic Markov chain, then process π_t is an ergodic diffusion on the simplex of probability vectors $\mathcal{S}^{d-1}$:

- a unique invariant measure φ on $\mathscr{B}(\mathcal{S}^{d-1})$ exists
- the convergence to φ is exponential
- the Law of Large Numbers holds

This implies that the limits of the performance indices exist

$$\mathcal{P}_{err} := \lim_{t\to\infty} \inf_{\xi\in L_\infty(\mathscr{F}_t^Y)} \mathsf{P}\big(\xi \ne X_t\big) = 1 - \int_{\mathcal{S}^{d-1}} (\max_i u_i)\varphi(du)$$

$$\mathcal{E}_{mse} := \lim_{t\to\infty} \inf_{\xi\in L_2(\mathscr{F}_t^Y)} \mathsf{E}\big(\xi - X_t\big)^2 = \int_{\mathcal{S}^{d-1}} \Big(u^* a^2 - (u^* a)^2\Big)\varphi(du)$$

The limit performance indeces

If X is an ergodic Markov chain, then process π_t is an ergodic diffusion on the simplex of probability vectors $\mathcal{S}^{d-1}$:

- a unique invariant measure φ on $\mathscr{B}(\mathcal{S}^{d-1})$ exists
- the convergence to φ is exponential
- the Law of Large Numbers holds

This implies that the limits of the performance indices exist

$$\mathcal{P}_{err} := \lim_{t\to\infty} \inf_{\xi\in L_\infty(\mathscr{F}_t^Y)} \mathsf{P}\big(\xi \neq X_t\big) = 1 - \int_{\mathcal{S}^{d-1}} (\max_i u_i)\varphi(du)$$

$$\mathcal{E}_{mse} := \lim_{t\to\infty} \inf_{\xi\in L_2(\mathscr{F}_t^Y)} \mathsf{E}\big(\xi - X_t\big)^2 = \int_{\mathcal{S}^{d-1}} \Big(u^* a^2 - (u^* a)^2\Big)\varphi(du)$$

Expression for the performance indices

What is known about $\mathcal{P}_{err}$ and $\mathcal{E}_{err}$?

- For the case $d=2$, both indices are explicitly computable
- Low noise asymptotic $(\sigma\to 0)$: assuming $h(a_i)\neq h(a_j)$,

O.Zeitouni & R.Khasminskii (1996):

$$\mathcal{P}_{err}(\sigma)=\left(\sum_{i=1}^d \mu_i \sum_{j\neq i}\frac{2\lambda_{ij}}{\big(h(a_i)-h(a_j)\big)^2}\right)\cdot\sigma^2\log\frac{1}{\sigma^2}\big(1+o(1)\big)$$

Y.Golubev (2000):

$$\mathcal{E}_{mse}(\sigma)=\left(\sum_{i=1}^d \mu_i \sum_{j\neq i}\frac{2\lambda_{ij}(a_i-a_j)^2}{\big(h(a_i)-h(a_j)\big)^2}\right)\cdot\sigma^2\log\frac{1}{\sigma^2}\big(1+o(1)\big).$$

Expression for the performance indices

What is known about $\mathcal{P}_{err}$ and $\mathcal{E}_{err}$?

- For the case $d = 2$, both indices are explicitly computable
- Low noise asymptotic $(\sigma \to 0)$: assuming $h(a_i) \neq h(a_j)$,

O.Zeitouni & R.Khasminskii (1996):

$$\mathcal{P}_{err}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big)$$

Y.Golubev (2000):

$$\mathcal{E}_{mse}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}(a_i - a_j)^2}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big).$$

Expression for the performance indices

What is known about $\mathcal{P}_{err}$ and $\mathcal{E}_{err}$?

- For the case $d = 2$, both indices are explicitly computable
- Low noise asymptotic $(\sigma \to 0)$: assuming $h(a_i) \neq h(a_j)$,

O.Zeitouni & R.Khasminskii (1996):

$$\mathcal{P}_{err}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big)$$

Y.Golubev (2000):

$$\mathcal{E}_{mse}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}(a_i - a_j)^2}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big).$$

Expression for the performance indices

What is known about $\mathcal{P}_{err}$ and $\mathcal{E}_{err}$?

- For the case $d = 2$, both indices are explicitly computable
- Low noise asymptotic $(\sigma \to 0)$: assuming $h(a_i) \neq h(a_j)$,

O.Zeitouni & R.Khasminskii (1996):

$$\mathcal{P}_{err}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big)$$

Y.Golubev (2000):

$$\mathcal{E}_{mse}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}(a_i - a_j)^2}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big).$$

Expression for the performance indices

What is known about $\mathcal{P}_{err}$ and $\mathcal{E}_{err}$?

- For the case $d = 2$, both indices are explicitly computable
- Low noise asymptotic $(\sigma \to 0)$: assuming $h(a_i) \neq h(a_j)$,

O.Zeitouni & R.Khasminskii (1996):

$$\mathcal{P}_{err}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big)$$

Y.Golubev (2000):

$$\mathcal{E}_{mse}(\sigma) = \left(\sum_{i=1}^{d} \mu_i \sum_{j \neq i} \frac{2\lambda_{ij}(a_i - a_j)^2}{\big(h(a_i) - h(a_j)\big)^2} \right) \cdot \sigma^2 \log \frac{1}{\sigma^2} \big(1 + o(1)\big).$$

Another side of the coin: measure concentration

- The filter "converges" to the signal as $\sigma \to 0$
- As $\sigma \to 0$, the invariant measure $\varphi_\sigma(du)$ of the filtering process π_t converges to the point measure

$$\varphi_0(du) = \sum_{i=1}^{d} \mu_i \delta_{e_i}(du),$$

  where $\{e_i\}$ is the standard basis of $\mathbb{R}^d$ ("corners" of $\mathcal{S}^{d-1}$)
- The results of O.Zeitouni/R.Khasminskii and Y.Golubev give partial information about concentration of $\varphi_\sigma(du)$ around $\varphi_0(du)$ as $\sigma \to 0$
- The complete characterization is still unknown (e.g. important for calculation of the entropy rate, etc.) !

Another side of the coin: measure concentration

- The filter "converges" to the signal as $\sigma \to 0$
- As $\sigma \to 0$, the invariant measure $\varphi_\sigma(du)$ of the filtering process π_t converges to the point measure

$$\varphi_0(du) = \sum_{i=1}^{d} \mu_i \delta_{e_i}(du),$$

where $\{e_i\}$ is the standard basis of $\mathbb{R}^d$ ("corners" of $\mathcal{S}^{d-1}$)

- The results of O.Zeitouni/R.Khasminskii and Y.Golubev give partial information about concentration of $\varphi_\sigma(du)$ around $\varphi_0(du)$ as $\sigma \to 0$
- The complete characterization is still unknown (e.g. important for calculation of the entropy rate, etc.) !

Another side of the coin: measure concentration

- The filter "converges" to the signal as $\sigma \to 0$
- As $\sigma \to 0$, the invariant measure $\varphi_\sigma(du)$ of the filtering process π_t converges to the point measure

$$\varphi_0(du) = \sum_{i=1}^{d} \mu_i \delta_{e_i}(du),$$

where $\{e_i\}$ is the standard basis of $\mathbb{R}^d$ ("corners" of $\mathcal{S}^{d-1}$)

- The results of O.Zeitouni/R.Khasminskii and Y.Golubev give partial information about concentration of $\varphi_\sigma(du)$ around $\varphi_0(du)$ as $\sigma \to 0$
- The complete characterization is still unknown (e.g. important for calculation of the entropy rate, etc.) !

Another side of the coin: measure concentration

- The filter “converges” to the signal as $\sigma \to 0$
- As $\sigma \to 0$, the invariant measure $\varphi_\sigma(du)$ of the filtering process π_t converges to the point measure

$$\varphi_0(du) = \sum_{i=1}^{d} \mu_i \delta_{e_i}(du),$$

where $\{e_i\}$ is the standard basis of $\mathbb{R}^d$ (“corners” of $\mathcal{S}^{d-1}$)
- The results of O.Zeitouni/R.Khasminskii and Y.Golubev give partial information about concentration of $\varphi_\sigma(du)$ around $\varphi_0(du)$ as $\sigma \to 0$
- The complete characterization is still unknown (e.g. important for calculation of the entropy rate, etc.) !

Strong noise asymptotic

Theorem (P.Ch., 2006)

Assume X is ergodic, then for any $F : \mathbb{R}^d \mapsto \mathbb{R}$, growing not faster than polynomially,

$$\lim_{\sigma \to \infty} \int_{\mathcal{S}^{d-1}} F\big(\sigma\big(u - \mu\big)\big)\varphi_\sigma(du) = \mathsf{E}F(\xi),$$

where $\xi \sim \mathcal{N}(0, P)$ with the covariance matrix P, solving uniquely the algebraic Lyapunov equation:

$$0 = \Lambda^* P + P\Lambda + \big(\mathrm{diag}(\mu) - \mu\mu^*\big)hh^*\big(\mathrm{diag}(\mu) - \mu\mu^*\big)$$

in the class of nonnegative definite matrices with $\sum_{ij} P_{ij} = 0$.

Strong noise asymptotic

Corollary

In particular,

$$\lim_{\sigma\to\infty}\sigma^2\big(\mathcal{E}_{mse}(\infty)-\mathcal{E}_{mse}(\sigma)\big)=a^*Pa,$$

where $\mathcal{E}_{mse}(\infty)$ is the a priori error and

$$\lim_{\sigma\to\infty}\sigma\big(\mathcal{P}_{err}(\infty)-\mathcal{P}_{err}(\sigma)\big)=\mathsf{E}\max_{j\in\mathcal{J}}Z_j,$$

where $\mathcal{J}=\{i:\mu_i=\max_j\mu_j\}$ and $\mathcal{P}_{err}(\infty)$ is the a priori error.

Strong noise asymptotic

Corollary

In particular,

$$\lim_{\sigma\to\infty} \sigma^2\big(\mathcal{E}_{mse}(\infty) - \mathcal{E}_{mse}(\sigma)\big) = a^* P a,$$

where $\mathcal{E}_{mse}(\infty)$ is the a priori error and

$$\lim_{\sigma\to\infty} \sigma\big(\mathcal{P}_{err}(\infty) - \mathcal{P}_{err}(\sigma)\big) = \mathrm{E}\max_{j\in\mathcal{J}} Z_j,$$

where $\mathcal{J} = \{i : \mu_i = \max_j \mu_j\}$ and $\mathcal{P}_{err}(\infty)$ is the a priori error.

Strong noise asymptotic

As $\sigma \to \infty$, the measure $\varphi_\sigma(du)$ converges to the a priori stationary distribution of the signal:

$$\lim_{\sigma\to\infty} \varphi_\sigma(du) = \delta_\mu(du)$$

and the Theorem states that the concentration is Gaussian under appropriate scaling (a CLT type result).

Strong noise asymptotic

Fait curieux:

$\mathcal{P}_{map}(\sigma)$ approaches the a priori limit $\mathcal{P}_{map}(\infty)$ much faster, if the maximal probability among all μ_i's is unique, namely for any $p \geq 1$

$$\lim_{\sigma\to\infty} \sigma^p \big(\mathcal{P}_{err}(\infty) - \mathcal{P}_{err}(\sigma)\big) = 0$$

if $|J| = 1$. Otherwise, i.e. if $|J| > 1$,

$$\lim_{\sigma\to\infty} \sigma \big(\mathcal{P}_{err}(\infty) - \mathcal{P}_{err}(\sigma)\big) = C$$

with $C > 0$!

The proof idea

- The Wonham SDE is regularly perturbed, when $\sigma \to \infty$:

$$d\pi_t^\sigma = \Lambda^* \pi_t^\sigma dt + \sigma^{-1}\big(\mathrm{diag}(\pi_t^\sigma) - \pi_t^\sigma \pi_t^{\sigma *}\big) H d\bar{B}_t, \quad \pi_0 = \nu,$$

  where $\bar{B}_t$ is the *innovation* Brownian motion.
- Obviously, $\lim_{\sigma\to\infty} \pi_t^\sigma = \mu$ in the stationary case
- The scaled error process $\sigma(\pi_t^\sigma - \mu)$ satisfies a linear (and hence Gaussian) SDE $\Longrightarrow$ the Lyapunov equation
- Some care should be taken of interchanging the limits $\sigma \to \infty$ and $t \to \infty$.

The proof idea

- The Wonham SDE is regularly perturbed, when $\sigma \to \infty$:

$$d\pi_t^\sigma = \Lambda^* \pi_t^\sigma dt + \sigma^{-1}\big(\mathrm{diag}(\pi_t^\sigma) - \pi_t^\sigma \pi_t^{\sigma *}\big) H\bar{B}_t, \quad \pi_0 = \nu,$$

  where $\bar{B}_t$ is the *innovation* Brownian motion.
- Obviously, $\lim_{\sigma\to\infty} \pi_t^\sigma = \mu$ in the stationary case
- The scaled error process $\sigma(\pi_t^\sigma - \mu)$ satisfies a linear (and hence Gaussian) SDE $\Longrightarrow$ the Lyapunov equation
- Some care should be taken of interchanging the limits $\sigma \to \infty$ and $t \to \infty$.

The proof idea

- The Wonham SDE is regularly perturbed, when $\sigma \to \infty$:

$$d\pi_t^\sigma = \Lambda^* \pi_t^\sigma dt + \sigma^{-1}\big(\mathrm{diag}(\pi_t^\sigma) - \pi_t^\sigma \pi_t^{\sigma *}\big) H\bar{B}_t, \quad \pi_0 = \nu,$$

  where $\bar{B}_t$ is the *innovation* Brownian motion.
- Obviously, $\lim_{\sigma\to\infty} \pi_t^\sigma = \mu$ in the stationary case
- The scaled error process $\sigma(\pi_t^\sigma - \mu)$ satisfies a linear (and hence Gaussian) SDE $\implies$ the Lyapunov equation
- Some care should be taken of interchanging the limits $\sigma \to \infty$ and $t \to \infty$.

The proof idea

- The Wonham SDE is regularly perturbed, when $\sigma \to \infty$:

$$d\pi_t^\sigma = \Lambda^* \pi_t^\sigma dt + \sigma^{-1} \big(\mathrm{diag}(\pi_t^\sigma) - \pi_t^\sigma \pi_t^{\sigma *}\big) H\bar{B}_t, \quad \pi_0 = \nu,$$

  where $\bar{B}_t$ is the *innovation* Brownian motion.
- Obviously, $\lim_{\sigma\to\infty} \pi_t^\sigma = \mu$ in the stationary case
- The scaled error process $\sigma(\pi_t^\sigma - \mu)$ satisfies a linear (and hence Gaussian) SDE $\implies$ the Lyapunov equation
- Some care should be taken of interchanging the limits $\sigma \to \infty$ and $t \to \infty$.