1. Abstract

Recently, significant progress has been made developing kernel mean expressions for Bayesian inference. An important success in this domain is the nonparametric kernel Bayes’ smoother (nKBS) filter, which can be used for sequential inference in state space models. We expand upon this work by introducing a smoothing algorithm, the nonparametric kernel Bayes’ smoother (nKBS-smoother) which relies on kernel Bayes’ inference through the kernel sum rule and kernel Bayes’ rule. We report experimental results that compare the nKBS-smoother to previous parametric and nonparametric approaches to Bayesian filtering and smoothing.

2. Preliminaries: Kernel Bayesian Inference

2.1 Positive-definite (p.d.) kernel

Let $X$ be a nonempty set. A symmetric function $k : X \times X \to \mathbb{R}$ is called a positive-definite kernel if for $n \geq 1$ and $x_1, \ldots, x_n \in X$, the matrix $G = (k(x_i, x_j))_{i,j \in [1,n]}$ is positive-semidefinite.

2.2 Reproducing kernel Hilbert space (RKHS)

The RKHS $H$ associated with kernel $k$ on a set $X$ is a Hilbert space consisting of functions $f : X \to \mathbb{R}$, which has the reproducing property:

$$f(x) = \langle f, k(x, \cdot) \rangle_H \quad \forall f \in H, \quad \forall x \in X,$$

where $\langle \cdot, \cdot \rangle_H$ denotes the inner product of $H$.

2.3 Kernel mean

The kernel mean of a probability distribution $P$ is defined as $m_P = \mathbb{E}_{X \sim P}[k(X, \cdot)] \in H$. Even when $P$ is complex-shaped, the kernel function $k$ models, and estimates, $m_P$. Kernel Bayesian inference infers the representation $m_P$ of $P$ in $H$, which implies a predictive distribution, posterior distribution and so on. The kernel mean $m_P$ has the following expectation property:

$$\langle m_P, f \rangle_H = \mathbb{E}_{X \sim P}[f(x)] \quad \forall f \in H.$$

2.4 Nonparametric kernel mean estimation

Let $D_n = \{(x_i, y_i) \in X \times \mathbb{R} \mid i = 1, \ldots, n\}$ be a data set. Let $N_{n,d} = \text{Span}(k_{x_1}, \ldots, k_{x_n})$ be its finite dimensional span. We estimate $m_P$ by an empirical mean $\hat{m}_P = \frac{1}{n} \sum_{i=1}^n k(x_i, \cdot)$, if $D_n$ is i.i.d. Then a kernel mean estimator is its sample mean $\hat{m}_P = \frac{1}{n} \sum_{i=1}^n k(x_i, \cdot)$; otherwise $w$ is non-uniform. Once we obtain weights $w$, expressing $m_P$, any expectation of any function $f$ in $H$ can be estimated as:

$$\mathbb{E}_{X \sim P}[f(x)] = \langle w, f \rangle_H = \sum_{i=1}^n w_i f(x_i).$$

2.5 Kernel sum rule (KSR) & Kernel Bayes’ rule (KB)

Kernel sum rule (KSR) employs sum rule in RKHSs. Kernel Bayes’ rule (KB) employs Bayes’ rule in RKHSs. Kernel sum rule is defined by marginalization $g(y) = \int g(y|x)p(x|y)dx$, Bayes’ rule is defined by $p(y|x) = p(y|x)p(x)/\mathbb{P}(x)$, with likelihood $p(y|x)$ and prior $p(x)$.

Let $(X, Y, k_{x}, k_{y}, k_{xy})$ be RKHSs. KSR computes marginal kernel mean $m_{g(Y)} = \mathbb{E}_{X \sim p(X|Y)}[g(Y)]$. Given input kernel mean $m_{g(Y)} = \mathbb{E}_{X \sim p(X|Y)}[g(Y)]$, KSR computes posterior kernel mean $m_{g(Y|X)} = \mathbb{E}_{Y \sim p(Y|X)}[g(Y)]$ given $u$ and prior kernel mean $m_{g(Y)}$.

2.6 Nonparametric KSR (nKSR) & nonparametric KB (nKB)

Let $(X, Y, k_{x}, k_{y}, k_{xy})$ be a joint sample drawn i.i.d. from $p(x, y)$ and $(X, Y)$ be another i.i.d sample in $X$. nKSR estimates kernel mean $m_{g(Y)} = \sum_i g(Y_i)$ given an input kernel mean $m_{g(Y)} = \sum_i g(Y_i)$. Its weight $w$ is computed by a matrix multiplication $w = M^{-1}m_{g(Y)}$ using the nKSR matrix $M$.

\[ M = \mathbb{E}_{X \sim p(X|Y)}[k_{xy}]^{-1} \text{ and } m_{g(Y)} = \mathbb{E}_{X \sim p(X|Y)}[g(Y)] \]

\[ m_{g(Y|X)} = \mathbb{E}_{Y \sim p(Y|X)}[g(Y)] \text{ given } u \text{ and prior kernel mean } m_{g(Y)} \]

nKB estimates posterior kernel mean $m_{g(Y|X)} = \sum_i g(Y_i|x_i)$ given a prior kernel mean $m_{g(Y)} = \sum_i g(Y_i)$ and $u$. Its weight $w$ is computed by a matrix multiplication $w = M^{-1}m_{g(Y)}$ using the nKB matrix $M$.

\[ M = \mathbb{E}_{X \sim p(X|Y)}[k_{xy}]^{-1} \text{ and } m_{g(Y)} = \mathbb{E}_{X \sim p(X|Y)}[g(Y)] \]

\[ m_{g(Y|X)} = \mathbb{E}_{Y \sim p(Y|X)}[g(Y)] \text{ given } u \text{ and prior kernel mean } m_{g(Y)} \]

4. Experimental Results

4.1 Synthetic: Cluttered environment on $\mathbb{R}^2$

Results on the problem of "Tracking a Single Object with Cluttered Measurements" in the RBMCDA toolbox.

4.2 Real-World: Slotcar State Estimation

Estimation of the position of a miniature car (1:32 scale) racing around a 1:16 track (a similar dataset was used in [24]).

The observations are noisy estimates of 3D acceleration and velocity of the car, collected at 10Hz. Ground truth positional information is captured by an overhead camera and then converted to a 2D position that tracks the car's progress around the circular manifold of the track.

5. Extension to General Tree Graphs

Marginal kernel mean inference on state space models can be extended to on general tree graphs.