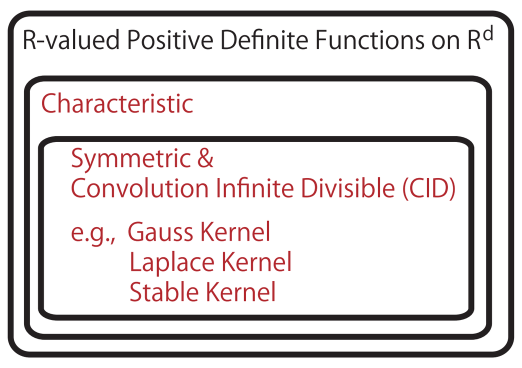
2014年6月13日　統計数理研究所　オープンハウス

カーネル法と確率分布の無限分解可能性

西山　悠　**統計的機械学習研究センター　特任助教**

1　結果：特性的カーネルの十分条件etc

1. $\mathbb{R}^d$上$\mathbb{R}$値有界連続正値関数が対称無限分解可能分布の確率密度関数で与えられるとき特性的カーネルとなる．



2. 無限分解可能分布(α安定分布，0-skewed NIG分布，0-skewed VG分布)のカーネル平均の厳密な表現を与え，正定値カーネルを使ったベイズ推論（カーネルベイズ推論）に組み込むことを考える（確率モデル入りカーネルベイズ推論）．

2　背景：カーネル平均

カーネル平均を使ったノンパラメトリック推定

- HMM [Le Song et al., ICML2010]
- Kernel Bayes' Rule [Fukumizu et al., NIPS2011]
- POMDP [Nishiyama et al., UAI2012]
- Kernel MonteCarlo Filter [Kanagawa et al., AAAI2014]
- 特性的カーネル[Sriperumbudur et al., JMLR2010]

3　畳み込み無限分解可能(CID)カーネル

以下では，有界連続な正定値カーネルを考える．

Definition 1. 関数$\psi : \mathbb{R}^d \to \mathbb{R}_0^+$を無限分解可能な確率測度の確率密度関数であり，かつ対称（$\psi(-x) = \psi(x)$, $x \in \mathbb{R}^d$）なものとする．$k(x,y) = \psi(x-y)$, $x, y \in \mathbb{R}^d$を畳み込み無限分解可能*(Convolution infinitely Divisible; CID)*なカーネルと呼ぶ．

Theorem 1. 畳み込み無限分解可能カーネルは特性的である．

Proof. $(\mathbb{R}^d, \mathcal{B}^d)$上の無限分解可能測度$F$の特性関数$\phi_F(u) = \int_{\mathbb{R}^d} e^{i\langle u,x\rangle} F(dx)$は，Lévy-Khinchin表示を持つ：

$$\begin{aligned}\phi_F(u) &= \exp\left(i\langle u,b\rangle - \frac{1}{2}\langle u,Cu\rangle\right. \\ &\quad \left. + \int_{\mathbb{R}^d}\left(e^{i\langle u,x\rangle} - 1 - i\langle u,x\rangle \mathbf{1}_{\{||x||\le 1\}}(x)\right)\nu(dx)\right) \qquad (1)\end{aligned}$$

$b \in \mathbb{R}^d$, $C \in \mathbb{R}^{d\times d}$は対称半正定値行列，$\mathbb{R}^d$上の$\nu(dx)$は，$\nu(\{\mathbf{0}\}) = 0$と$\int_{R^d}(||x||^2 \wedge 1)\nu(dx) < \infty$を満たすLévy測度．$\mathbb{R}^d$上の確率測度$F$が対称であるのは，特性関数$\phi_F(u)$が実であることと必要十分．より$\phi_F(u) > 0$[Sato 1999].有界連続な正値関数$\psi : \mathbb{R}^d \to \mathbb{R}$により得られる平行移動不変正定値カーネル$\psi(x-y)$が特性的となるための必要十分条件は，$\psi$のフーリエ変換$\Lambda(\omega)$のサポートが全域$\mathbb{R}^n$[Sriperumbudur et al., JMLR2010]. □

Example 1 (畳み込み無限分解可能カーネル)**.** 対称α安定*($S\alpha S$)*分布*(*正規分布*($\alpha = 2$))*，コーシー分布*($\alpha = 1$)*，*Holtsmark*分布*($\alpha = \frac{3}{2}$))*，*student-t*分布*$STU(\nu)$*，対称一般化双曲型*(SGH)*分布*$GH_d(\lambda,\alpha,\beta = 0,\delta,\mu = 0,\Delta)$*.対称*tempered α*-安定*(STαS)*分布．

4　無限分解可能分布のカーネル平均の例

4.1　1変量α安定分布$S_\alpha(\sigma,\beta,\mu)$

対称α安定分布の確率密度関数$d_{stable}(x;\alpha,\sigma,0,0)$が作る平行移動不変カーネルは畳み込み無限分解可能カーネルである：

$$k_{\alpha,\sigma}(x,y) = \frac{1}{2\pi}\int_{\mathbb{R}} e^{-iu(x-y)-|\sigma u|^\alpha} du. \qquad (2)$$

α安定分布$S_\alpha(\sigma,\beta,\mu)$の$\mathcal{H}_{\alpha,\sigma_0}$におけるカーネル平均は，

$$\begin{aligned} m_P &= d_{stable}(\cdot;\alpha,\sigma_0,0,0) * d_{stable}(\cdot;\alpha,\sigma,\beta,\mu) \\ &= d_{stable}(\cdot;\alpha,\tilde{\sigma},\tilde{\beta},\tilde{\mu}) \in \mathcal{H}_{\alpha,\sigma_0}. \end{aligned} \qquad (3)$$

ここで$\tilde{\sigma} = ||\boldsymbol{\sigma}||_\alpha$, $\boldsymbol{\sigma} = (\sigma_0,\sigma)$, $\tilde{\beta} = \left(\frac{\sigma}{\tilde{\sigma}}\right)^\alpha \beta$, $\tilde{\mu} = \mu$である．

4.2　多変量一般化双曲型分布$GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)$

一般化双曲型分布$GH_d(\lambda,\alpha,\beta,\delta,\mu,\Delta)$は，$\lambda$：分布のクラス，$\alpha$：形，$\beta$：歪度, μ：位置，δ：スケール，Δ：正定値行列のパラメータを持つ確率分布族である．K_νを第三種の変形ベッセル関数とする．

1. Normal inverse Gaussian (NIG)分布$GH_d(-\frac{1}{2},\alpha,\beta,\delta,\mu,\Delta)$.

$$d_{NIG_d(\alpha,\beta,\delta,\mu,\Delta)}(x) \propto \left(\sqrt{\delta^2 + ||x-\mu||^2_{\Delta^{-1}}}\right)^{-\frac{d+1}{2}} K_{\frac{d+1}{2}}\left(\alpha\sqrt{\delta^2 + ||x-\mu||^2_{\Delta^{-1}}}\right) e^{\langle\beta,x-\mu\rangle}.$$

2. Vavriance-Gamma (VG)分布$GH_d(\lambda > 0,\alpha,\beta,\delta \to 0,\mu,\Delta)$.

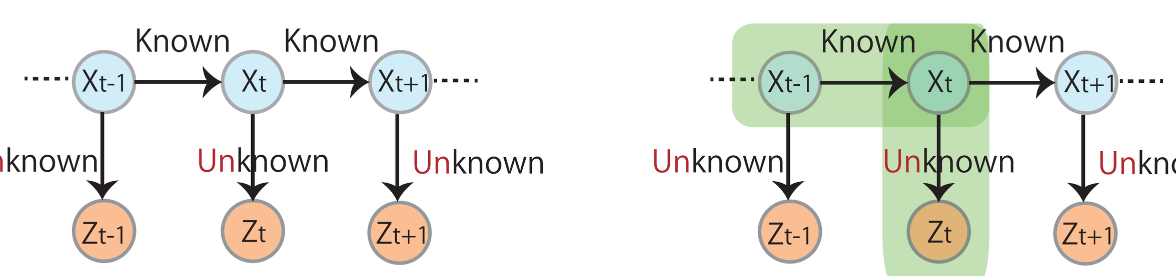
$$d_{VG_d(\lambda,\alpha,\beta,\mu,\Delta)}(x) \propto ||x-\mu||^{\lambda-\frac{d}{2}}_{\Delta^{-1}} K_{\lambda-\frac{d}{2}}\left(\alpha||x-\mu||_{\Delta^{-1}}\right) e^{\langle\beta,x-\mu\rangle}.$$

Proposition 1. 確率密度pの再生核ヒルベルト空間$\mathcal{H}$におけるカーネル平均m_pを$(p,\mathcal{H},m_p)$と書く．

1. $(d_{NIG_d(\alpha,0,\delta,\mu,\Delta)}, \mathcal{H}_{SNIG_d(\alpha,\delta_0,\Delta)}, d_{NIG_d(\alpha,0,\delta_0+\delta,\mu,\Delta)}(x))$.
2. $(d_{VG_d(\lambda,\alpha,0,\mu,\Delta)}, \mathcal{H}_{SVG_d(\lambda_0,\alpha,\Delta)}, d_{VG_d(\lambda_0+\lambda,\alpha,0,\mu,\Delta)})$.
3(a). $(d_{NIG_d(\alpha,0,\delta,\mu,\Delta)}, \mathcal{H}_{GH_d(1/2,\alpha,0,\delta_0,0,\Delta)}, d_{GH_d(1/2,\alpha,0,\delta_0+\delta,\mu,\Delta)})$.
3(b). $(d_{GH_d(1/2,\alpha,0,\delta,\mu,\Delta)}, \mathcal{H}_{GH_d(1/2,\alpha,0,\delta_0,0,\Delta)}, d_{GH_d(1/2,\alpha,0,\delta_0+\delta,\mu,\Delta)})$.
4(a). $(d_{GH_d(-\lambda,\alpha,0,\delta,\mu,\Delta)}, \mathcal{H}_{GH_d(\lambda,\alpha,0,0,0,\Delta)}, d_{GH_d(\lambda,\alpha,0,\delta,\mu,\Delta)})$.
4(b). $(d_{GH_d(\lambda,\alpha,0,0,\mu,\Delta)}, \mathcal{H}_{GH_d(\lambda,\alpha,0,\delta,0,\Delta)}, d_{GH_d(\lambda,\alpha,0,\delta,\mu,\Delta)})$.

4.3　確率モデルのカーネル平均の利用

- 確率モデルありカーネル和公式と状態空間モデルフィルタリング[西山 et al., IBIS2012, Kanagawa et al., AAAI2014]



カーネル平均m_pから確率密度pを復元[Smola et al., ALT2007; Song et al., ICML2008; McCalman et al., ICRA2013].

$$\min_\beta ||m[p] - m[\hat{p}]||^2_{\mathcal{H}}, \text{ subj to } \beta^T\mathbf{1} = 1, \beta \ge 0. \qquad (4)$$