8.2 確率分布を特徴づける正定値カーネル

茨城大学工学部情報工学科
08T4018Y 小幡智裕

確率変数を特徴づける正定値カーネル

定義

(χ, B_χ) を可測空間、Pをその上の確率速度全体としたとき、
χ 上の有界かつ可測な正定値カーネルkが特性的であるとは、
写像

$$\mathrm{P} \to \mathrm{H}_k, \quad P \mapsto m_P^k$$

が単写であることをいう。
m_P^k は分布 P を持つ χ 上の確率変数 H_k における平均。

確率変数を特徴づける正定値カーネル(続き)

定義(続き)

　分布Pを持つ確率変数Xに対してf(X)の期待値を$E_{X\sim P}[f(X)]$と表すと上の定義は

$$E_{X\sim P}[f(X)] = E_{X\sim Q}[f(X)] \quad (\forall f \in H_k) \Rightarrow P = Q$$

と同値である。

確率変数を特徴づける正定値カーネル（続き）

$\{k(\cdot, y) \mid y \in \chi\}$の線形結合が$H_k$で稠密であることより、条件$m_P^k = m_Q^k$は任意の$y \in \chi$に対し$E_{X \sim P}[k(X, y)] = E_{X \sim Q}[k(X, y)]$が成り立つことと同値である。

$$E_{X \sim P}[k(X, y)] = E_{X \sim Q}[k(X, y)] \quad (\forall y \in \chi) \quad \Leftrightarrow \quad P = Q$$

を成立させる正定値カーネルが特性的な正定値カーネルとなる。

確率変数を特徴づける正定値カーネル（続き）

補題8.6

上記の記法のもと、kが特性的であるための必要十分条件は、任意の確率分布$P \in \mathrm{P}$に対し、$H_k + R$が$m_P^k = m_Q^k$で稠密であることである。

$H_k + R$は再生核ヒルベルト空間としての直和

補題の証明

・十分性

$P, Q \in \mathrm{P}$に対し、$P \neq Q$かつ$m_P^k = m_Q^k$として矛盾を導く。

$P-Q$の全変動を$|P-Q|$で表すとき、仮定から$\mathrm{H}_k + R$は$L^2(|P-Q|)$で稠密なので、χの任意の可測集合Aと任意の$\varepsilon > 0$に対し、$\varphi \in \mathrm{H}_k + R$があって

$$\int |\varphi(x) - I_A(x)| \, d(|P-Q|)(x) < \varepsilon$$

が成り立つ。ここでI_AはAの定義関数である。このとき

$$|(E_{X\sim P}[\varphi(X)] - P(A)) - (E_{X\sim Q}[\varphi(X)] - Q(A)) < \varepsilon$$

である。$m_P^k = m_Q^k$により$E_{X\sim P}[\varphi(X)] = E_{X\sim Q}[\varphi(X)]$なので、$|P(A) - Q(A)| < \varepsilon$であるが、$\varepsilon > 0$は任意なので$P(A) = Q(A)$となり$P \neq Q$に反する。

補題の証明（続き）

・必要性

ある$P \in \mathrm{P}$があって$\mathrm{H}_k + R$が$L^2(P)$で稠密でないと仮定する。

このとき、0でない$f \in L^2(P)$を$\mathrm{H}_k + R$の直交補空間からとると、

$$\int f\varphi dP = 0 \quad (\forall \varphi \in \mathrm{H}_k), \quad \int f dP = 0$$

が成立する。$c = 1/\| f \|_{L^1(P)}$とおき、2つの確率Q_1, Q_2を

$$Q_1(E) \equiv c\int_E | f | dP, \qquad Q_2(E) \equiv c\int_E (| f | - f) dP$$

により定義する。$f \neq 0$により$Q_1 \neq Q_2$であるが、一方任意の$\varphi \in \mathrm{H}_k$に対し

$$E_{X \sim Q_1}[\varphi(X)] - E_{X \sim Q_2}[\varphi(X)] = c\int f\varphi dP = 0$$

により$m_{Q_1}^k = m_{Q_2}^k$である。したがってkは特性的でない。

Bochnerの定理

Bochnerの定理

ϕをR^n 上の複素数値連続関数とする。このときϕが正値であるための必要十分条件は、R^n 上の有限な非負Borel測度Λが

$$\phi(x) = \int e^{\sqrt{-1}\omega^T x} d\Lambda(\omega)$$

と表されることである。

Bochnerの定理はR^n 上のすべての連続な正値関数をフーリエ変換で特徴づける。

特性的であるための条件（Bochnerの定理）

平行移動不変な正定値カーネル$k(x,y)=\phi(x-y)$に対し、確率PのH_kにおける平均m_P^kは

$$m_P^k(x)=\int k(x,y)dP(y)=\int \phi(x-y)dP(y)=(\phi * P)(x)$$

と表され、特性的であることは

$$\phi * P=\phi * Q \quad \Rightarrow \quad P=Q$$

と同値である。

特性的であるための条件（Bochnerの定理）（続き）

定理8.7

ϕをR^n 上の連続な複素数値正値関数とし、Λ をBochnerの定理の表示

$$\phi(x) = \int e^{\sqrt{-1}\omega^T x} d\Lambda(\omega)$$

を与える有限非負Borel測度とする。このとき、$Supp(\Lambda) = R^n$ であれば $\phi(x-y)$は特性的な正定値カーネルである。

位相空間Ω上の非負Borel測度μのSupp(μ)は

$$Supp(\mu) = \{x \in \Omega \mid x\text{を含む任意の開集合}U\text{に対し}\mu(U) > 0\}$$

特性的であるための条件（Bochnerの定理）（続き）

証明

有限な実測度μが$\mu * \phi = 0$を満たすとき、$\mu = 0$を示す。

*Fubini*の定理を用いると

$$\int (\mu * \phi)(x) d\mu(x) = \int\int \phi(x-y) d\mu(y) d\mu(x)$$

$$= \int\int\int e^{\sqrt{-1}(x-y)^T \omega} \Lambda(\omega) d\mu(y) d\mu(x)$$

$$= \int\int e^{\sqrt{-1}x^T \omega} d\mu(x) \int e^{-\sqrt{-1}y^T \omega} d\mu(y) d\Lambda(\omega) = \int |\hat{\mu}(\omega)|^2 \, d\Lambda(\omega)$$

である。ここで$\mu * \phi = 0$より

$$\int |\hat{\mu}(\omega)|^2 \, d\Lambda(\omega) = 0$$

を得るが、$\hat{\mu}$がR^n上連続であり$Supp(\Lambda) = R^n$であることから$\hat{\mu} = 0$となる。

フーリエ変換の一意性により$\mu = 0$を得る。

特性的であるための条件（Bochnerの定理）（続き）

定理8.8

ϕをR^n 上の連続な複素数値正値関数とし、ΛをBochnerの定理の表示

$$\phi(x) = \int e^{\sqrt{-1}\omega^T x} d\Lambda(\omega)$$

を与える有限非負測度とする。このとき、$\phi(x-y)$が特性的な正定値カーネルであるための必要十分条件は$Supp(\Lambda) = R^n$である。

特性的であるための条件（Bochnerの定理）（続き）

証明

集合$A\subset R^n$に対して$-A=\{-a\in R^n \mid a\in A\}, A-A=\{a-b\in R^n \mid a,b\in A\}$と表す。

定理8.7より必要性のみ示せばよい。

$k(x,y)=\phi(x-y)$が特性的であるとき、$Supp(\Lambda)\neq R^n$と仮定し矛盾を導く。

そのために2つの異なる確率分布のフーリエ変換の差として表される関数hで

$Supp(h)\bigcap Supp(\Lambda)=\emptyset$となるものを構成する。

まずϕが実関数であることから、任意の$Borel$集合Eに対して$\Lambda(-E)=\Lambda(E)$となる。

実際、$\tilde{\Lambda}(E)=\Lambda(-E)$とおくと

$$\int e^{\sqrt{-1}\omega^T x}d\tilde{\Lambda}(\omega)=\int e^{-e\sqrt{-1}\omega^T x}d\Lambda(\omega)=\overline{\phi(x)}=\phi(x)$$

なので、$Bochner$の定理の一意性から$\tilde{\Lambda}=\Lambda$を得る。

よって$R^n\setminus Supp(\Lambda)$は空でない開集合である。

証明の続き

$\omega_0 \in R^n \setminus Supp(\Lambda)\,(\omega_0 \neq 0)$を固定すると、原点のある開近接$W$があって、
$\pm\omega_0 \notin W - W$かつ$cl(W-W) \pm \omega_0 \subset R^n \setminus Supp(\Lambda)$とできる。
このようなWを固定し、$g = I_w * I_{-W}$と定める。I_Wは集合Wの定義関数である。
関数gは連続で、$Supp(g) \subset cl(W-W)$,さらにgはR^n上の正値関数である。
正値性は

$$\sum_{i,j} c_i \bar{c}_j g(\omega_i - \omega_j) = \sum_{i,j} c_i \bar{c}_j \int I_W(\omega_i - \omega_j - \gamma) I_{-W}(\gamma) d\gamma$$
$$= \sum_{i,j} c_i \bar{c}_j \int I_W(\omega_i - \gamma) I_{-W}(\gamma - \omega_j) d\gamma = \int |\sum_i c_i I_W(\omega_i - \gamma)|^2 \, d\gamma \geq 0$$

よりわかる。したがって$Bochner$の定理よりある有界な非負$Borel$測度μがあって

$$g(\omega) = \int e^{\sqrt{-1}\omega^T x} d\mu(x)$$

が成り立つ。$h(\omega) = g(\omega - \omega_0) + g(\omega + \omega_0)$と定めると

$$h(\omega) = \int e^{\sqrt{-1}\omega^T x} 2\cos(\omega_0^T x) d\mu(x)$$

を得る。

証明の続き

$Supp(g) \subset cl(W-W)$と$cl(W-W) \pm \omega_0 \subset R^n \setminus Supp(\Lambda)$により、
$Supp(h) \bigcap Supp(\Lambda) \neq \emptyset$が成り立つ。また、$\pm\omega_0 \notin W-W$より$h(0)=0$であるから

$$\nu(E) = \int_E 2\cos(\omega_0^T x) d\mu(x)$$

により実の符号付測度νを定義すると、$\nu(R^n)=0$が成立する。
そこで$c = |\nu|(R^n)$とおき、測度μ_1, μ_2を

$$\mu_1 = \frac{1}{c}|\nu|, \qquad \mu_2 = \frac{1}{c}\{|\nu| - \nu\}$$

と定義する。$\nu(R^n)=0$によりμ_1, μ_2はともに確率測度である。$Fubini$の定理を用いると

$$\begin{aligned} c((\mu_1-\mu_2)*\phi)(x) &= \int \phi(x-y) 2\cos(\omega_0^T y) d\mu(y) \\ &= \int 2\cos(\omega_0^T y) \int e^{\sqrt{-1}(x-y)^T\omega} d\Lambda(\omega) d\mu(y) \\ &= \int e^{\sqrt{-1}x^T\omega} h(\omega) d\Lambda(\omega) = 0 \end{aligned}$$

となり、これは$k(x,y)$が特性的であることに矛盾する。

特性的な正定値カーネル

定理8.7より、さまざまな平行移動不変な正定値カーネルが特性的であることがわかる。

特性的なカーネルのフーリエ変換はすべての周波数成分を扱うことができ、特性的でないカーネルはある周波数領域を表すことができない。

→その周波数成分のみ異なる密度関数をもつ確率を区別できない。

普遍性

kをコンパクト位相空間X上の有界かつ連続な正定値カーネルとする。kが定める再生核ヒルベルト空間が、X上の連続関数全体に $\| f \|_\infty = \sup_{x\in\chi} |f(x)|$ というノルムを入れて定義されるバナッハ空間の中で稠密であるとき、kは普遍であるという。

ガウスBRFカーネルやラプラスカーネルなどが、R^m 上の任意のコンパクト集合上で普遍であることが知られている。

普遍性（続き）

命題8.10

コンパクト距離空間上の普遍な正定値カーネルは特性的である。

証明

P,Qをコンパクト距離空間上の確率分布とするとき、任意の有界連続関数fに関して$E_{X\sim P}[f(X)]=E_{X\sim Q}[f(X)]$ならば$P=Q$ であることが知られているので主張は容易に従う。