キュムラント情報幾何について

公文　雅之　リスク解析戦略研究センター　特任准教授

【概要】

ベクトル値定常時系列の2次および高次キュムラントスペクトルからエントロピー, ダイバージェンス, 相互情報量といった情報諸量がキュムラント空間の情報幾何学的観点より必然的に定義される. 以下の論文にもとづきこれらの定義の情報幾何学的背景を, 現在取り組んでいる2つの研究テーマとの関連を含めて説明する.

M. Kumon. Studies of information quantities and information geometry of higher order cumulant spaces. *Statistical Methodology*, **7**, 152-172, 2010.

Invariant information geometrical structures on system spaces.
ISM Research Memorandum, No. 1210, 2020.

Information geometry of multivariate martingales.

【記号の定義】

column vector-valued stationary time series with real-valued components

$X(t)=(X_a(t);a=1,\dots,m)\quad t=0,\pm1,\pm2,\dots$

joint cumulant function of order k of $X(t)$

$c_{a_1\dots a_k}(u_1,\dots,u_{k-1})=\mathrm{cum}\{X_{a_1}(u_1),\dots,X_{a_{k-1}}(u_{k-1}),X_{a_k}(0)\}$

k-th order cumulant spectrum of $X(t)$

$S_{a_1\dots a_k}(\lambda_1,\dots,\lambda_{k-1})=(2\pi)^{k-1}f_{a_1\dots a_k}(\lambda_1,\dots,\lambda_{k-1})$

$\quad=\sum_{\forall u_j=-\infty}^{\infty}c_{a_1\dots a_k}(u_1,\dots,u_{k-1})\exp\big(-i\sum_{j=1}^{k-1}u_j\lambda_j\big)$

$S_k(\boldsymbol{\lambda})=[S_{\boldsymbol{a}}(\boldsymbol{\lambda})]\quad \boldsymbol{a}=(a_1,\dots,a_k),\ \boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_k)\in P_k$

$\quad P_k=\big\{\boldsymbol{\lambda}=(\lambda_1,\dots,\lambda_k)\big|\sum_{j=1}^{k}\lambda_j\equiv 0\ (\mathrm{mod}\,2\pi)\big\}$

$\langle F(\boldsymbol{\lambda})\rangle=(2\pi)^{-k+1}\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}F(\lambda_1,\dots,\lambda_{k-1})d\lambda_1\cdots d\lambda_{k-1}$

$m\times m$ spectral density matrix

$S_2(\lambda)=[S_{a_1a_2}(\lambda)]\ (a_1,a_2=1,\dots,m)$ rank $S_2(\lambda)=m,\ \langle\log|S_2(\lambda)|\rangle>-\infty$

$\Rightarrow\ S_2(\lambda)=A(e^{i\lambda})A(e^{-i\lambda})'$ spectral factorization

$A(z),z\in\mathbb{C}:m\times m$ matrix (outer function) analytic with $A(z)^{-1}$ in $|z|\le 1$

$X^I(t)=A(e^{i\lambda})^{-1}X(t)\ \Rightarrow\ S_2^I(\lambda)=I_m$

$X^I(t)$: second order white noise

$S_{\boldsymbol{a}}^I(\boldsymbol{\lambda})=\prod_{j=1}^{k}[A(e^{i\lambda_j})^{-1}]_{a_j}^{b_j}S_{\boldsymbol{b}}(\boldsymbol{\lambda})$

$S_{\boldsymbol{a}}^I(\boldsymbol{\lambda})$: normalized k-th order cumulant spectrum

$\sum_{b_1=1}^{m}\cdots\sum_{b_k=1}^{m}[A(e^{i\lambda_1})^{-1}]_{a_1}^{b_1}\cdots[A(e^{i\lambda_k})^{-1}]_{a_k}^{b_k}S_{b_1\dots b_k}(\boldsymbol{\lambda})$

Einstein's summation convention

【推定問題とキュムラント空間のリーマン計量】

model of the k-th order cumulant spectrum

$M=\{f_{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\ |\boldsymbol{\theta}\in\Theta\ (\text{open in }\mathbb{R}^s)\}$

s-vector valued estimating function

$y^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})=[y^{a_1\cdots a_k}(\boldsymbol{\lambda},\boldsymbol{\theta})]\ y^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})=(y_1^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta}),\dots,y_s^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta}))';(\boldsymbol{\lambda},\boldsymbol{\theta})\in P_k\times\Theta$

deterministic version of estimating equation

$Y(\boldsymbol{\theta})=\int_{-\pi}^{\pi}\cdots\int_{-\pi}^{\pi}y^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})f_{\boldsymbol{a}}(\boldsymbol{\lambda})d\lambda_1\cdots d\lambda_{k-1}=0$

$\Rightarrow$ strong consistency and asymptotic normality

(Cramér-Rao type inequality)

asymptotic covariance matrix Ω is bounded by

$\Omega\ge\Omega_0=k!G(\boldsymbol{\theta}_0)^{-1}\quad G(\boldsymbol{\theta})=[g_{\alpha\beta}(\boldsymbol{\theta})]$

$g_{\alpha\beta}(\boldsymbol{\theta})=\langle T^{\boldsymbol{ab}}(-\boldsymbol{\lambda},\boldsymbol{\theta})\partial_\alpha S_{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\partial_\beta S_{\boldsymbol{b}}(-\boldsymbol{\lambda},\boldsymbol{\theta})\rangle$: Riemannian metric on M

$T^{\boldsymbol{ab}}(\boldsymbol{\lambda},\boldsymbol{\theta})=\prod_{j=1}^{k}S^{a_jb_j}(\lambda_j,\boldsymbol{\theta}),\ \partial_\alpha=\partial/\partial\theta^\alpha\quad[S^{a_jb_j}(\lambda_j,\boldsymbol{\theta})]=[S_{a_jb_j}(\lambda_j,\boldsymbol{\theta})]^{-1}$

optimal estimating function

$y_\beta^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})=T^{\boldsymbol{ab}}(\boldsymbol{\lambda},\boldsymbol{\theta})\partial_\beta S_{\boldsymbol{b}}(\boldsymbol{\lambda},\boldsymbol{\theta});\ \beta=1,\dots,s$: Nonholonomic form

【キュムラント空間のダイバージェンス】

最良推定関数ベクトルが, あるキュムラント型ポテンシャルUの勾配ベクトルとして表わされる. このU-条件を仮定すると パラメータの2変数内積関数$G(\theta,\theta')$の導入を通じて2種類のポテンシャル関数$E(\theta)$, $F(\theta')$の存在が導かれ $D(\theta,\theta')=E(\theta)+F(\theta')-G(\theta,\theta')$ としてダイバージェンス$D(\theta,\theta')$が定義される.

(U-condition) $\quad\partial_\alpha y_\beta^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})=\partial_\beta y_\alpha^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\ \Leftrightarrow\ y_\beta^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})=\partial_\beta U^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})$

inner product function

$G(\boldsymbol{\theta},\boldsymbol{\theta}')=\langle U^{\boldsymbol{a}}(-\boldsymbol{\lambda},\boldsymbol{\theta}')S_{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle$

its two kinds of derivatives

$Y_\beta(\boldsymbol{\theta},\boldsymbol{\theta}')=\partial_\beta G(\boldsymbol{\theta},\boldsymbol{\theta}')=\langle U^{\boldsymbol{a}}(-\boldsymbol{\lambda},\boldsymbol{\theta}')\partial_\beta S_{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle$

$Z_\beta(\boldsymbol{\theta},\boldsymbol{\theta}')=\partial'_\beta G(\boldsymbol{\theta},\boldsymbol{\theta}')=\langle\partial'_\beta U^{\boldsymbol{a}}(-\boldsymbol{\lambda},\boldsymbol{\theta}')S_{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle$

Then at the diagonal $\boldsymbol{\theta}=\boldsymbol{\theta}'$

$\partial_\alpha Y_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=\partial_\beta Y_\alpha(\boldsymbol{\theta},\boldsymbol{\theta}),\ \partial_\alpha Z_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=\partial_\beta Z_\alpha(\boldsymbol{\theta},\boldsymbol{\theta})$

$\Leftrightarrow\ \exists E(\boldsymbol{\theta}),\ \exists F(\boldsymbol{\theta})$ such that

$Y_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=\langle U^{\boldsymbol{a}}(-\boldsymbol{\lambda},\boldsymbol{\theta})\partial_\beta S_{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle=\partial_\beta E(\boldsymbol{\theta})$

$Z_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=\langle\partial_\beta U^{\boldsymbol{a}}(-\boldsymbol{\lambda},\boldsymbol{\theta})S_{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle=\partial_\beta F(\boldsymbol{\theta})$

$\Rightarrow\ D(\boldsymbol{\theta},\boldsymbol{\theta}')=E(\boldsymbol{\theta})+F(\boldsymbol{\theta}')-G(\boldsymbol{\theta},\boldsymbol{\theta}')$

□ 2次キュムラント空間のダイバージェンス

$y_\beta(\lambda,\boldsymbol{\theta})=\partial_\beta[-S_2(\lambda,\boldsymbol{\theta})^{-1}],\ U(\lambda,\boldsymbol{\theta})=-S_2(\lambda,\boldsymbol{\theta})^{-1}$

$G(\boldsymbol{\theta},\boldsymbol{\theta}')=-\langle\mathrm{tr}[S_2(\lambda,\boldsymbol{\theta}')^{-1}S_2(\lambda,\boldsymbol{\theta})]\rangle\quad G(\boldsymbol{\theta},\boldsymbol{\theta})=-m$

$Y_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=-\partial_\beta\langle\log|S_2(\lambda,\boldsymbol{\theta})|\rangle\quad E(\boldsymbol{\theta})=-\langle\log|S_2(\lambda,\boldsymbol{\theta})|\rangle$

$Z_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=-\partial_\beta\langle\log|S_2(\lambda,\boldsymbol{\theta})^{-1}|\rangle\quad F(\boldsymbol{\theta})=-\langle\log|S_2(\lambda,\boldsymbol{\theta})^{-1}|\rangle$

$F(\boldsymbol{\theta})=-E(\boldsymbol{\theta})$: mutual duality

Divergence (Bregman type)

$D(\boldsymbol{\theta},\boldsymbol{\theta}')=\langle\mathrm{tr}[S_2(\lambda,\boldsymbol{\theta}')^{-1}S_2(\lambda,\boldsymbol{\theta})]\rangle-m-\langle\log|S_2(\lambda,\boldsymbol{\theta}')^{-1}S_2(\lambda,\boldsymbol{\theta})|\rangle$

For two m-vector series $X(t),Y(t)$

$D_2(X,Y)=-H_2(X)+H_2(Y)-G_2(X,Y)-m$

$X(t)$から$Y(t)$への2次のダイバージェンス

$H_2(X)=-E(X)=\langle\log|S_2^X(\lambda)|\rangle\quad H_2(Y)=F(Y)=\langle\log|S_2^Y(\lambda)|\rangle$

$X(t),Y(t)$の2次のエントロピー

$G_2(X,Y)=-\langle\mathrm{tr}[S_2^Y(\lambda)^{-1}S_2^X(\lambda)]\rangle$

$X(t)$と$Y(t)$の2次の内積関数

□ 高次キュムラント空間のダイバージェンス

When $k\ge3,\ S_2^I(\lambda,\boldsymbol{\theta})=I_m\Rightarrow\ T^{\boldsymbol{ab}}(-\boldsymbol{\lambda},\boldsymbol{\theta})=\delta^{\boldsymbol{ab}}=\prod_{j=1}^{k}\delta^{a_jb_j}$

$y_\beta^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})=\partial_\beta[\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})],\ U^{\boldsymbol{a}}(\boldsymbol{\lambda},\boldsymbol{\theta})=\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})$

$G(\boldsymbol{\theta},\boldsymbol{\theta}')=\langle\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(-\boldsymbol{\lambda},\boldsymbol{\theta}')S_{\boldsymbol{a}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle=G(\boldsymbol{\theta}',\boldsymbol{\theta})$

$\quad G(\boldsymbol{\theta},\boldsymbol{\theta})=\langle\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(-\boldsymbol{\lambda},\boldsymbol{\theta})S_{\boldsymbol{a}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle=\|S_k^I(\boldsymbol{\lambda},\boldsymbol{\theta})\|^2$

$Y_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=\partial_\beta\left[\frac{1}{2}\langle\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(-\boldsymbol{\lambda},\boldsymbol{\theta})S_{\boldsymbol{a}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle\right]\quad E(\boldsymbol{\theta})=\frac{1}{2}\langle\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(-\boldsymbol{\lambda},\boldsymbol{\theta})S_{\boldsymbol{a}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle$

$Z_\beta(\boldsymbol{\theta},\boldsymbol{\theta})=\partial_\beta\left[\frac{1}{2}\langle\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(-\boldsymbol{\lambda},\boldsymbol{\theta})S_{\boldsymbol{a}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle\right]\quad F(\boldsymbol{\theta})=\frac{1}{2}\langle\delta^{\boldsymbol{ab}}S_{\boldsymbol{b}}^I(-\boldsymbol{\lambda},\boldsymbol{\theta})S_{\boldsymbol{a}}^I(\boldsymbol{\lambda},\boldsymbol{\theta})\rangle$

$F(\boldsymbol{\theta})=E(\boldsymbol{\theta})$: self duality

Divergence (Bregman type)

$D(\boldsymbol{\theta},\boldsymbol{\theta}')=\frac{1}{2}\|S_k^I(\boldsymbol{\lambda},\boldsymbol{\theta})-S_k^I(\boldsymbol{\lambda},\boldsymbol{\theta}')\|^2=D(\boldsymbol{\theta}',\boldsymbol{\theta})$

For two m-vector series $X(t)$, $Y(t)$

$D_k(X,Y)=-H_k(X)-H_k(Y)-G_k(X,Y)$

$X(t)$と$Y(t)$とのk次のダイバージェンス

$H_k(X)=-E(X)=-\frac{1}{2}\|S_k^{XI}\|^2\quad H_k(Y)=-F(Y)=-\frac{1}{2}\|S_k^{YI}\|^2$

$X(t),Y(t)$のk次のエントロピー

$G_k(X,Y)=\langle\delta^{\boldsymbol{ab}}S_{\boldsymbol{a}}^{XI}(\boldsymbol{\lambda})S_{\boldsymbol{b}}^{YI}(-\boldsymbol{\lambda})\rangle$

$X(t)$と$Y(t)$のk次の内積関数

【キュムラント空間の相互情報量】

$X(t)$と$Y(t)$を2つの定常時系列として$\{X(t),Y(t)\}$は定常時系列$Q(t)$で決まるものとする. このとき$X(t)$と$Y(t)$とのk次の相互情報量($k=2,3,\dots$)を $I_k(X,Y)=H_k(X)+H_k(Y)-H_k(Q)$ で定義する. 基本的性質の一つとして高次の相互情報量は負になり得る.

【キュムラント空間の全情報諸量】

total entropy of $X\quad H_\Sigma(X)=\sum_{k\ge2}H_k(X)/k!$

total mutual information between X and $Y\quad I_\Sigma(X,Y)=\sum_{k\ge2}I_k(X,Y)/k!$

total divergence from X to $Y\quad D_\Sigma(X,Y)=\sum_{k\ge2}D_k(X,Y)/k!$

【今後の発展】

キュムラント空間の情報諸量はシステム空間の不変な情報幾何学的構造から導入されており, これらにもとづいてフィードバックを伴う線形システムおよび Volterra 展開・Wiener-Itô 展開される非線形システムの解析への情報幾何学的方法論に発展していく.