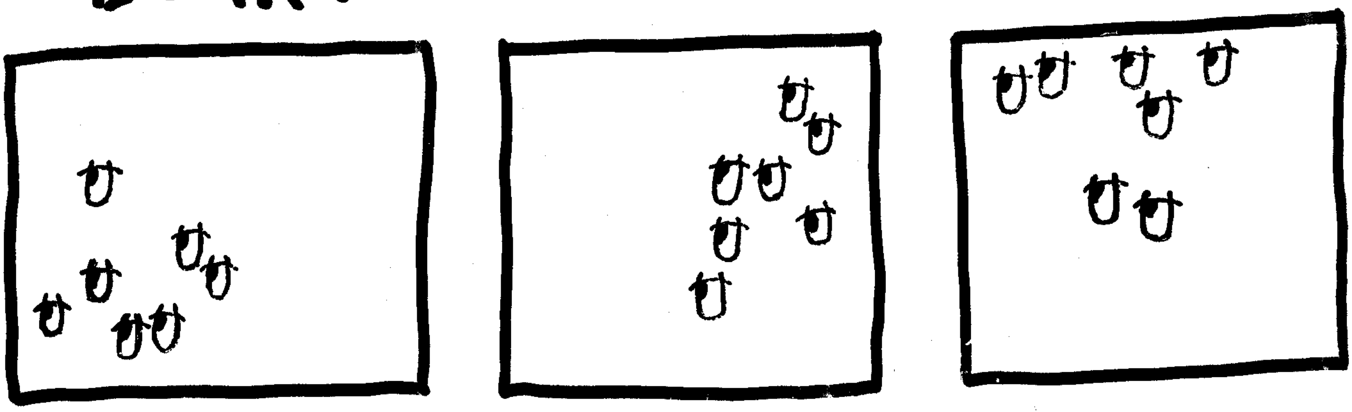
マルコフ連鎖と確率変数の依存性の評価

蒔田恵典（名大多元数理）

動機

空間中をランダムに動く多数の粒子(動物、人、…)を何度も観察する。



粒子の分布に強い相関がみられる(群など…)

?

個々の粒子の動きは、他の粒子の位置や動きと強い相関をもっている。

?を正当化するために…

セッティング

◎ nコのランダムにうごく粒子

→ $E \times E \times \cdots \times E$ 上のマルコフ連鎖

◎ 粒子の分布を何度も観察して得られたデータ

→ マルコフ連鎖の不変測度(定常状態)

個々の粒子の動きと他の粒子の位置や動きの間の相関

が弱い。

せんい確率(Markov核)の性質

これを示すことが目標。

⇓

各粒子の位置の間の相関

が弱い。

確率変数 $X_1, X_2, \cdots, X_n$ の間の'依存性'

まず、これを定義する。

Joint cumulant

$f_1, f_2, \cdots, f_n$: random variables of probability space (E, μ)

Def (Joint cumulant) $n \geq 2$

$$\mathrm{cum}_\mu(f_1, \cdots, f_n) := \sum_{k=1}^{n} (-1)^{k-1}(k-1)! \sum_{\sigma \in S(n,k)} \prod_{P \in \sigma} \mu\Big(\prod_{j \in P} f_j\Big).$$

(notation)

$$S(n,k) := \{\text{partitions of set } \{1,2,\cdots,n\} \text{ with } k \text{ parts}\}$$

ex)

$$S(4,3) = \left\{ \begin{array}{l} \{\{1,2\},\{3\},\{4\}\}, \{\{1,3\},\{2\},\{4\}\}, \{\{1,4\},\{2\},\{3\}\} \\ \{\{2,3\},\{1\},\{4\}\}, \{\{2,4\},\{1\},\{3\}\}, \{\{3,4\},\{1\},\{2\}\} \end{array} \right\}$$

ex)

- $\mathrm{cum}_\mu(f_1, f_2) = \mu(f_1 f_2) - \mu(f_1)\mu(f_2) = \mathrm{cov}(f_1, f_2)$
- $\mathrm{cum}_\mu(f_1, f_2, f_3) = \mu(f_1 f_2 f_3) - \mu(f_1)\mu(f_2 f_3)$
  $- \mu(f_2)\mu(f_1 f_3) - \mu(f_3)\mu(f_1 f_2)$
  $+ 2\mu(f_1)\mu(f_2)\mu(f_3)$

Properties of joint cumulant $n \geq 2$

- $\mathrm{cum}_\mu(f_1, \cdots, f_i, \cdots, f_j, \cdots, f_n)$
  $= \mathrm{cum}_\mu(f_1, \cdots, f_j, \cdots, f_i, \cdots, f_n)$,

- $\mathrm{cum}_\mu(af_1, f_2, \cdots, f_n)$
  $= a\, \mathrm{cum}_\mu(f_1, f_2, \cdots, f_n)$, $a \in \mathbb{R}$,

- $\mathrm{cum}_\mu(f_1 + g_1, f_2, \cdots, f_n)$
  $= \mathrm{cum}_\mu(f_1, f_2, \cdots, f_n) + \mathrm{cum}_\mu(g_1, f_2, \cdots, f_n)$,

- $\mathrm{cum}_\mu(f_1 + a, f_2, \cdots, f_n)$
  $= \mathrm{cum}_\mu(f_1, f_2, \cdots, f_n)$, $a \in \mathbb{R}$.

definition of dependence $D_\alpha(\mu)$

$(E, \mathcal{E}, \mu)$: probability space.

$\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_n$: family of sub-σ-algebra of $\mathcal{E}$.

$\alpha := \{\mathcal{E}_1, \cdots, \mathcal{E}_n\}$

Def (dependence)

$$D_\alpha(\mu) := \sup_{\substack{A_i \in \mathcal{E}_i \\ i=1,2,\cdots,n}} \left| \sum_{k=1}^{n} (-1)^{k-1} (k-1)! \sum_{\sigma \in \zeta(n,k)} \prod_{p \in \sigma} \mu\left(\bigcap_{j \in p} A_j\right) \right|$$

ex) $\alpha = \{\mathcal{E}_1, \mathcal{E}_2\}$

$$D_\alpha(\mu) = \sup_{\substack{A_1 \in \mathcal{E}_1 \\ A_2 \in \mathcal{E}_2}} \left| \mu(A_1 \cap A_2) - \mu(A_1)\mu(A_2) \right|$$

calculation of $D_\alpha(\mu)$

$E = \{0,1\} \times \{0,1\}$

$\mathcal{E}_1 = \{ \{\phi\}, \{(0.0),(0.1)\}, \{(1.0),(1.1)\}, E \}$

$\mathcal{E}_2 = \{ \{\phi\}, \{(0.0),(1,0)\}, \{(0,1),(1,1)\}, E \}$

$\alpha = \{\mathcal{E}_1, \mathcal{E}_2\}$

μ : probobility measure s.t.

$\mu(1.0) = \frac{1}{4} - a$, $\mu(1,1) = \frac{1}{4} + a$

$\mu(0.0) = \frac{1}{4} - b$, $\mu(0.1) = \frac{1}{4} + b$

$0 \leq a.b \leq \frac{1}{4}$

$\Rightarrow D_\alpha(\mu) = \left|\frac{a-b}{2}\right|$

Dependence $D_\alpha(\mu)$

and joint cumulant $\mathrm{cum}_\mu(f_1, \cdots, f_n)$

$(E, \mathcal{E}, \mu)$: probability space

$\mathcal{E}_1, \cdots, \mathcal{E}_n$: family of sub-σ-algebra of $\mathcal{E}$.

$\alpha = \{\mathcal{E}_1, \cdots, \mathcal{E}_n\}$

Lemma

$$\sup_{\substack{f_i \in \mathcal{E}_i \\ \mathrm{osc}(f_i) \leq C_i \\ i = 1, 2, \cdots, n}} |\mathrm{cum}_\mu(f_1, \cdots, f_n)| = C_1 C_2 \cdots C_n D_\alpha(\mu)$$

(notations)

$f_i \in \mathcal{E}_i \Leftrightarrow f_i$ is $\mathcal{E}_i$-measurable.

$$\mathrm{osc}(f_i) := \sup_{x, y} |f_i(x) - f_i(y)|$$

properties of $D_\alpha(\mu)$

$(E, \mathcal{E}, \mu)$: probability space.

$\mathcal{E}_1, \cdots, \mathcal{E}_n$: sub-σ-algebra of $\mathcal{E}$.

Proposition $\mathcal{E}_{23}$: σ-algebra generated by $\mathcal{E}_2$ & $\mathcal{E}_3$.

$\alpha := \{\mathcal{E}_1, \mathcal{E}_2, \mathcal{E}_3\}$, $\beta := \{\mathcal{E}_1, \mathcal{E}_{23}\}$

$\Rightarrow D_\alpha(\mu) \leqslant D_\beta(\mu)$

Proposition $\alpha := \{\mathcal{E}_1, \cdots, \mathcal{E}_n\}$

$\mathcal{E}_1$ and $\bigvee_{j=2}^{n} \mathcal{E}_j$ are independent or

$\mathcal{E}_1 \vee \mathcal{E}_2$ and $\bigvee_{j=3}^{n} \mathcal{E}_j$ are independent

$\Rightarrow D_\alpha(\mu) = 0$

more practical definitions...

$$\alpha := \{\varepsilon_1, \varepsilon_2, \cdots, \varepsilon_n\}$$

$$\sigma \in \mathfrak{S}(n,m) \qquad m \leq n$$

$$\rightarrow \alpha_\sigma = \{\varepsilon_{\rho_1}, \varepsilon_{\rho_2}, \cdots, \varepsilon_{\rho_m}\}_{\rho_i \in \sigma}$$

where $\varepsilon_\rho := \bigvee_{j \in \rho} \varepsilon_j \; , \; \rho \subset \{1, 2, \cdots, n\}$.

$$D_\alpha^m(\mu) := \sum_{\sigma \in \mathfrak{S}(n,m)} D_{\alpha_\sigma}(\mu) \,,$$

$$\overline{D}_\alpha^m(\mu) := \sup_{\sigma \in \mathfrak{S}(n,m)} D_{\alpha_\sigma}(\mu) \,.$$

Markov kernel (Stochastic transition function)

$(E, \mathcal{E})$: measurable space

$M : E \times \mathcal{E} \to [0,1]$ is Markov kernel

iff $\begin{cases} \cdot \ \forall x \in E, \ \mathcal{E} \ni A \mapsto M(x,A) \text{ is prob. meas.} \\ \cdot \ \forall A \in \mathcal{E}, \ E \ni x \mapsto M(x,A) \text{ is measurable.} \end{cases}$

- Dobrushin's ergodic coefficient

$$\beta(M) := \sup_{x,y \in E} \| M(x,\cdot) - M(y,\cdot) \|_{tv}$$

, where $\|\mu\|_{tv} = \frac{1}{2} \sup_{A,B \in \mathcal{E}} |\mu(A) - \mu(B)|$

for finite signed meas μ on $\mathcal{E}$.

Thm (Dobrushin, 1956)

$\|\mu M\|_{tv} \leq \beta(M) \|\mu\|_{tv}$

$$\beta(M) = \sup \left\{ \frac{\|\mu M\|_{tv}}{\|\mu\|_{tv}} : \mu : \text{finite signed meas s.t. } \mu(E) = 0 \right\}$$

$$= \sup \{ \mathrm{osc}(Mf) : \mathrm{osc}(f) \leq 1 \}$$

Markov kernel and measure

M : Markov kernel on $(E, \mathcal{E})$

for signed measure μ,

$\mu M(A) = \int_E M(x, A)\, \mu(dx)$, $\forall A \in \mathcal{E}$

for bounded function f,

$Mf(x) = \int_E f(y) M(x, dy)$, $\forall x \in E$.

$\rightarrow \mu M(f) = \mu(Mf)$

※ Probability measure μ is invariant measure of M iff

$\mu M = \mu$.

monotonous decrease of dependence

$(E_i, \mathcal{E}_i)$: measurable spaces $i=1,2,\cdots,n$.

$E := E_1 \times E_2 \times \cdots \times E_n$

$\mathcal{E} := \mathcal{E}_1 \otimes \mathcal{E}_2 \otimes \cdots \otimes \mathcal{E}_n$

M_i : Markov kernels on $(E_i, \mathcal{E}_i)$, $i=1,\cdots,n$.

M : Markov kernel on $(E, \mathcal{E})$ s.t.

$$M((x_1, x_2, \cdots, x_n), \cdot) = M_1(x_1, \cdot) \otimes \cdots \otimes M_n(x_n, \cdot)$$

$\alpha := \{\mathcal{E}_1, \cdots, \mathcal{E}_n\}$ $\quad x_i \in E_i$

Theorem (M)

$\forall \mu$: probability measure on $(E, \mathcal{E})$

$$D_\alpha(\mu M) \leq \beta(M_1) \cdots \beta(M_n) D_\alpha(\mu)$$

<u>proof</u>

$$D_\alpha(\mu M)$$

$$= \sup_{\substack{f_i \in \mathcal{E}_i \\ osc(f_i) \leq 1}} |cum_{\mu M}(f_1, \cdots, f_n)|$$

$$= \sup_{\substack{\shortparallel}} |cum_\mu (M_1 f_1, \cdots, M_n f_n)|$$

$$\leq \sup_{\substack{f_i \in \mathcal{E}_i \\ osc(f_i) \leq \beta(M_i)}} |cum_\mu (f_1, \cdots, f_n)|$$

$$= \beta(M_1) \cdots \beta(M_n) \sup_{\substack{f_i \in \mathcal{E}_i \\ osc(f_i) \leq 1}} |cum_\mu (f_1, \cdots, f_n)|$$

$$= \beta(M_1) \cdots \beta(M_n) D_\alpha(\mu) \quad //$$

Mutual Information.

(Ω, P) : probability space.

E_i : finite set. $i=1,2$.

$X_i : \Omega \to E_i$: random variables. $i=1,2$.

Def (entropy. mutual information)

$$H^P(X_1) := -\sum_{x\in E_1} P(X_1=x)\log P(X_1=x)$$

$$I^P(X_1;X_2) := H^P(X_1) - H^P(X_1 \mid X_2)$$

$$= \sum_{x_1\in E_1}\sum_{x_2\in E_2} P(X_1=x_1, X_2=x_2)\log\frac{P(X_1=x_1, X_2=x_2)}{P(X_1=x_1)P(X_2=x_2)}$$

E : finite set. μ, ν : probability meas. on E.

Def (relative entropy)

$$D(\mu\|\nu) := \begin{cases} \sum_{x\in E}\mu(x)\log\frac{\mu(x)}{\nu(x)}, & \mu \ll \nu, \\ \infty & \text{, otherwise} \end{cases}$$

$$\Rightarrow \quad I^P(X_1;X_2) = D(P_{X_1X_2} \,\|\, P_{X_1}\otimes P_{X_2})$$

Thm (P. Del Moral, M. Ledoux, L. Micro, 2003)

$\forall \mu, \nu$: probability measures on $(E, \mathcal{E})$

M : Markov kernel on $(E, \mathcal{E})$

$$D(\mu M \| \nu M) \leq \beta(M) D(\mu \| \nu)$$

Cor

$\beta(M_1)\beta(M_2) \leq \beta(M)$

$E = E_1 \times E_2$: finite set.

$P_i : E \to E_i$: projections $i = 1, 2$.

M_i : Markov kernels on E_i $i = 1, 2$.

M : Markov kernel on E s.t.

$M((x_1, x_2), \cdot) = M_1(x_1, \cdot) \otimes M_2(x_2, \cdot)$

$\forall \mu$: probability meas on E

$$I^{\mu M}(P_1 ; P_2) \leq \beta(M) I^{\mu}(P_1 ; P_2)$$

☺ $I^{\mu M}(P_1 ; P_2) = D(\mu M \| P_{1*}(\mu M) \otimes P_{2*}(\mu M))$

$= D(\mu M \| (P_{1*}\mu \otimes P_{2*}\mu) M)$

$\leq \beta(M) D(\mu \| P_{1*}\mu \otimes P_{2*}\mu) = \beta(M) I^{\mu}(P_1 ; P_2)$

//

Notations.

$E = E_1 \times E_2 \times \cdots \times E_n$: at most countable

$P_i : E \to E_i$: projections.

$p \subset \{1, 2, \cdots, n\}$

$\to P_p : E \to \prod_{j \in p} E_j$: projections.

$A_1 \times A_2 \times \cdots \times A_n \subset E_1 \times E_2 \times \cdots \times E_n$

$\to A_p := \prod_{j \in p} A_j$

$\alpha := \{\mathcal{E}_1, \mathcal{E}_2, \cdots, \mathcal{E}_n\}$

where $\mathcal{E}_i := \{P_i^{-1}(A) : A \subset E_i\}$ $i = 1, \cdots, n$.

Proposition

M : Markov kernel on E
with invariant measure μ.

$$\Theta := \sup \left| \sum_{k=1}^{n} (-1)^{k-1} (k-1)! \sum_{\sigma \in \mathcal{S}(n,k)} \left\{ \prod_{\rho \in \sigma} M(y^\sigma_\rho, P_\rho^{-1}(B_\rho)) - \prod_{j=1}^{n} M(z^\sigma_j, P_j^{-1}(B_j)) \right\} \right|$$

sup is taken over

$$\begin{cases} z^\sigma_j \in P_j^{-1}(x^\sigma),\ j = 1, \cdots, n \\ y^\sigma_\rho \in P_\rho^{-1}(x^\sigma),\ \rho \in \sigma \\ x^\sigma \in E,\ \sigma \in \mathcal{S}(n,k),\ k = 1, \cdots, n \\ B_1 \times \cdots \times B_n \subset E_1 \times \cdots \times E_n \end{cases}$$

$$\beta_i := \sup_{x, y \in E} \| M(x, P_i^{-1}(\cdot)) - M(y, P_i^{-1}(\cdot)) \|_{TV}^{E_i} \quad i = 1, 2, \cdots, n.$$

if $\beta_1 \beta_2 \cdots \beta_n \neq 1$ then

$$D_\alpha(\mu) \leq \frac{\Theta}{1 - \beta_1 \beta_2 \cdots \beta_n}$$

$\lrcorner$

Theorem (M)

$E = E_1 \times E_2$

M : Markov kernel on E

with invariant measure μ.

$$\gamma := \sup \left| \underline{M((x_1,x_2), A_1 \times A_2)} \right.$$
$$\left. \underline{- M((x_1,y_2), P_1^{-1}(A_1)) M((y_1,x_2), P_2^{-1}(A_2))} \right|$$

sup is taken over

$y_1 = x_1$, $y_2 = x_2$ とすると $D_\alpha(M((x_1,x_2), \cdot))$

$$\begin{cases} x_1, y_1 \in E_1 \\ x_2, y_2 \in E_2 \\ A_1 \times A_2 \subset E_1 \times E_2 \end{cases}$$

$$\beta_i := \sup_{x,y \in E} \| M(x, P_i^{-1}(\cdot)) - M(y, P_i^{-1}(\cdot)) \|_{tv}^{E_i}$$

$i = 1, 2.$

if $\beta_1 \beta_2 \neq 1$ then

$$D_\alpha(\mu) \leq \frac{\gamma}{1 - \beta_1 \beta_2}$$

┘