確率数理要論8

● Gauss過程

Def

$\mathbb{R}^m$値確率変数$(X_1, \dots, X_m)$が多変量正規分布に従うとは、$^\forall t_1, \dots, t_m \in \mathbb{R}$に対して
$$t_1X_1 + t_2X_2 + \cdots + t_mX_m$$
が$\mathbb{R}$上の正規分布に従うことと定義する。
ただし、分散0でも正規分布であると言う。

平均： $\mu = (E[X_1], \dots, E[X_m])^T \in \mathbb{R}^m$
共分散： $\Sigma = (\underbrace{E[X_iX_j] - E[X_i]E[X_j]}_{Cov(X_i, X_j)})_{i,j} \in \mathbb{R}^{m \times m}$

を用いて、$N(\mu, \Sigma)$と書く。 //

※ この定義ならΣがランク落ちしていても良い。
（通常の密度関数を用いた定義では扱えない。）

実際、$\Sigma > 0$のときに、密度関数を用いた定義と同値であることをチェックする。
$\mu = 0$としても一般性を失わない。
$t_1X_1 + \cdots + t_mX_m$の平均は0，分散は$t^T\Sigma t$である。
これが正規分布に従うので、その特性関数は
$$\phi_t(s) = E[e^{is(t_1X_1 + \cdots + t_mX_m)}] = \exp\left(-\frac{1}{2}s^2(t^T\Sigma t)\right) \quad \text{—— } (*)$$
である。$s=1$とすると、$\phi_t(1)$は$(X_1, \dots, X_m)^T$の特性関数に他ならない。
実際、$\phi(t) = E[e^{it^TX}] = E[e^{i(t_1X_1 + \cdots + t_mX_m)}]$
$$= \exp\left(-\frac{1}{2}t^T\Sigma t\right) \quad (\because *)$$
これは平均0，分散共分散Σの多変量正規分布の特性関数。
特性関数の一意性より、両者の定義が同値であることがわかる。

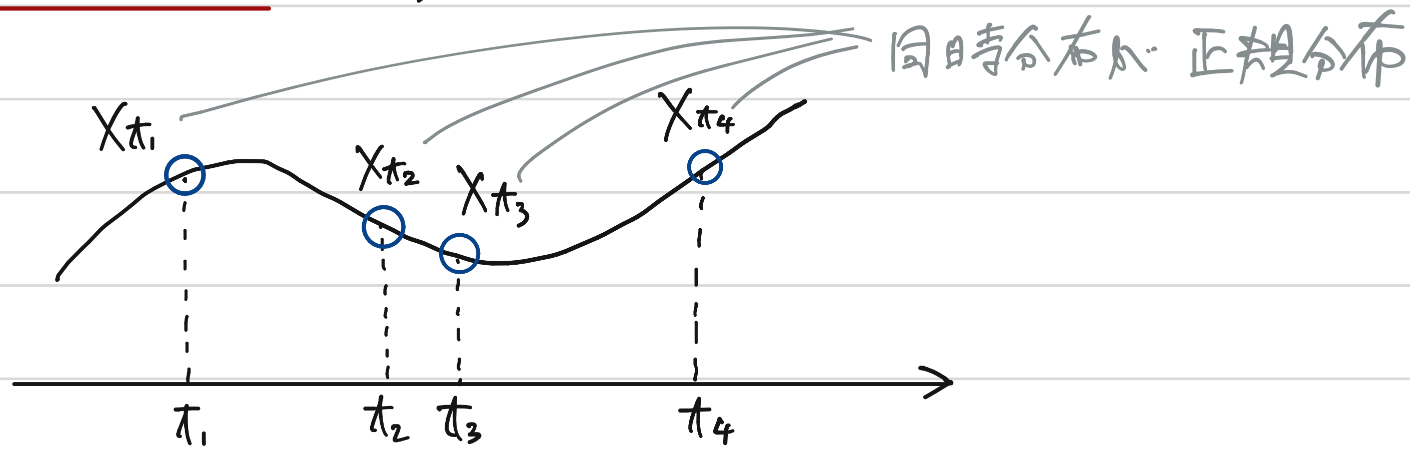
Def (Gauss過程, Gaussian process)

T : インデックスの集合 ($T=\mathbb{R}$ とは限らない. 一般の集合)

確率過程 $X=(X_t)_{t\in T}$ が任意の $t_1,\dots,t_n\in T$ に対し.

$(X_{t_1},\dots,X_{t_n})$ が多変量正規分布に従う時.

Gauss過程と言う.



$m(t)=E[X_t]$: 平均関数

$k(t,t')=E[(X_t-m(t))(X_{t'}-m(t'))]$: 共分散関数

$GP(m,k)$ と書く.

- 再生核ヒルベルト空間

共分散関数 k は以下の性質を持つ:

- 対称 : $k(t,s) = k(s,t)$
- 正定値: $\forall \alpha_1, \dots, \alpha_m \in \mathbb{R},\ \forall t_1, \dots, t_m \in T,$
$$\sum_{1 \le i,j \le m} \alpha_i \alpha_j k(t_i, t_j) \geq 0$$

この2つの性質を持った関数 $k: T \times T \to \mathbb{R}$ を正定値カーネルと言う.

例: ① 線形カーネル : $k(x,y) = x^T y \quad (x, y \in \mathbb{R}^d)$

② Gaussカーネル : $k(x,y) = \exp\left(-\frac{1}{2\sigma^2}\|x-y\|^2\right)$

Def (再生核ヒルベルト空間, Reproducing kernel Hilbert space, RKHS)

集合 T 上の RKHS H は, T 上の(実数値)関数からなるヒルベルト空間で, $\forall t \in T$ に対して $k_t \in H$ が存在して

(再生成) $\langle f, k_t \rangle_H = f(t) \quad (\forall f \in H)$

が成り立つものとする.

特に, $k(x,y) = \langle k_x, k_y \rangle_H$ を H に付随した再生核 (Reproducing kernel) と呼ぶ. //

※ $k(x,y)$ は正定値カーネルになっている.

つまり, RKHS が1つ与えられれば, 1つ正定値カーネルが定まる.

実は, 以下のようにその逆が成り立つ.

Thm (Moore-Aronszajn)

$k: T \times T \to \mathbb{R}$ を正定値カーネルとする.

$\Longrightarrow$ k を再生核とする RKHS が一意に存在する. //

(略証) $\mathcal{H}_0 = \{ f(\cdot) = \sum_{j=1}^{m} \alpha_j k(\cdot, t_j) \mid \alpha_j \in \mathbb{R}, t_j \in T, m \in \mathbb{N} \}$
とし、
$\mathcal{H}_0$ の元 $f = \sum_{j=1}^{m} \alpha_j k(t_j, \cdot)$, $g = \sum_{j=1}^{m'} \beta_j k(t'_j, \cdot)$ に対し、

$$\langle f, g \rangle_{\mathcal{H}_0} = \sum_{i=1}^{m} \sum_{j=1}^{m'} \alpha_i \beta_j k(t_j, t'_j)$$

と"内積"を定める。
この内積から定まるノルムで $\mathcal{H}_0$ を完備化したものを $\mathcal{H}$ とする。
つまり、$\mathcal{H}$ は $\mathcal{H}_0$ 内のコーシー列の集合で、
$\mathcal{H}_0$ 内のコーシー列 $f = (f_n)_n$, $g = (g_n)_n$ に対して、

$$\langle f, g \rangle_{\mathcal{H}} := \lim_{n\to\infty} \langle f_n, g_n \rangle_{\mathcal{H}_0}$$

とし、$\langle f-g, f-g \rangle_{\mathcal{H}} = 0$ である2つのコーシー列は同一視する。
$f \in \mathcal{H}$ に対して、f に対応するコーシー列 $(f_n)_n$ を用いて、

$$f(t) = \lim_{n\to\infty} \langle f_n, k(t, \cdot) \rangle_{\mathcal{H}_0}$$

とすれば、$f \in \mathcal{H}$ は T 上の実数値関数とみなせる。

実は、この $\mathcal{H}$ は $\langle \cdot, \cdot \rangle_{\mathcal{H}}$ を内積として持つヒルベルト空間になっていることが示せる。
さらに、$\mathcal{H}_0$ 内のコーシー列 $(k_t)_{n=1}^{\infty}$ も $k_t \in \mathcal{H}$ と書くと、
$\mathcal{H}$ は k_t を再生核とする RKHS になっている。 //

$$\boxed{\begin{array}{ccc} k & \xleftrightarrow{\text{一対一対応}} & \mathcal{H} \\ (\text{正定値カーネル}) & & (\text{RKHS}) \end{array}}$$

Thm

ある実数値関数のヒルベルト空間 $\mathcal{H}$ を考える。
$\forall t \in T$ において $f \mapsto f(t)$ が $\mathcal{H}$ のノルムについて連続
$\iff$ $\mathcal{H}$ は RKHS. //

(証明は略, Rieszの表現定理より従う)

Thm (Mercerの定理(一般形))

T: ハウスドルフ位相空間

ν: T上の非負Borel測度で $\mathrm{supp}(\nu) = T$ とする。　"Borel測度"の定義に含まれることも多い

ただし、全てのコンパクトな $A \in \mathcal{B}(T)$ において $\nu(A) < \infty$ とする。

$k: T \times T \to \mathbb{R}$ は連続な正定値カーネルで

$$\int k(t,t)\, d\nu(t) < \infty$$

とする。

このとき、kによって決まるRKHS $\mathcal{H}$ は可分で、

$$\begin{cases} \cdot\ \mu_1 \geqq \mu_2 \geqq \cdots \geqq 0 \\ \cdot\ (e_i)_i:\ L^2(\nu) \text{内の ONS (正規直交系)} \end{cases}$$

が存在し、$^\forall t \in T$ において

$$k(t,t') = \sum_{i=1}^{\infty} \mu_i e_i(t) e_i(t') \qquad \text{(Mercer展開)}$$

が $^\forall t' \in T$ で絶対収束する。

さらに、任意のコンパクト集合 $A \subset T$ に対し、$A \times A$ 上で x, x' に関して一様収束する。 //

元々は、Mercer (1909) による。$T = [a,b]$ で示された。

上記の一般化については、

Steinwart & Scovel: Mercer's theorem on general domains: On the interaction between measures, kernels and RKHSs. Constructive Approximation, 35:363–417, 2012.

kに付随して、積分作用素

$$T_k f(x) = \int k(x,y) f(y)\, d\nu(y)$$

が定義できる。Mercer展開はこの積分作用素の固有値・固有関数による展開を与えている。

Mercer展開を用いると、 ($(\sqrt{\mu_i}e_i)_{i=1}^\infty$ が$\mathcal{H}$のCONSなら)

$f\in\mathcal{H}$ は $\sum_{i=1}^\infty \alpha_i^2<\infty$ である $(\alpha_i)_{i=1}^\infty$ を用いて、

$$f=\sum_{i=1}^\infty \alpha_i\sqrt{\mu_i}\,e_i$$

のように書けて、$\mathcal{H}$内の内積は

$$f=\sum_{i=1}^\infty \alpha_i e_i,\quad g=\sum_{i=1}^\infty \beta_i e_i\in\mathcal{H} \text{ に対して、}$$

$$\langle f,g\rangle_{\mathcal{H}}=\sum_{i=1}^\infty \frac{1}{\mu_i}\alpha_i\beta_i$$

で与えられる。

これをβ_iとみなす

特に、 $k(x,\cdot)=\sum_{i=1}^\infty \mu_i e_i(x)\,e_i(\cdot)$ に対して、

$$\langle f,k(x,\cdot)\rangle_{\mathcal{H}}=\sum_{i=1}^\infty \mu_i^{-1}\cdot\alpha_i\cdot\mu_i e_i(x)$$

$f=\sum_{i=1}^\infty \alpha_i e_i$ とする。

$$=\sum_{i=1}^\infty \alpha_i e_i(x)=f(x)$$

であることより、再生性も確かめられる。

<u>Cor</u>

$\int k(x,x)\,d\nu(x)<\infty$ なら、$\mathcal{H}$は$L^2(\nu)$に<u>連続に埋めこまれる</u>。 //

$\mathcal{H}$の単位球が
$L^2(\nu)$の有限半径の球に含まれる。

(略証)

$f\in\mathcal{H}$ に対し、

$$\int f(x)^2\,d\nu=\int\langle f,k(x,\cdot)\rangle_{\mathcal{H}}^2\,d\nu\le\int\|f\|_{\mathcal{H}}^2\,\|k(x,\cdot)\|_{\mathcal{H}}^2\,d\nu$$

$$=\|f\|_{\mathcal{H}}^2\int k(x,x)\,d\nu\lesssim\|f\|_{\mathcal{H}}^2$$ //

Thm (Karhunen-Loève展開)

- 正定値カーネル $k: T\times T \to \mathbb{R}$ が Mercerの定理の条件を満たしているとする。この時、確率過程

$$X_t = \sum_{i=1}^{\infty} \xi_i \sqrt{\mu_i}\, e_i(t)$$

← RKHSの元の展開と比較せよ。

$$\xi_i \sim N(0,1) \quad (\text{i.i.d.})$$

は $\mathcal{GP}(0,k)$ になる。

- 逆に、$(X_t)_t$ が $\mathcal{GP}(0,k)$ であるとき、

$$\xi_i(\omega) = \frac{1}{\sqrt{\mu_i}} \int X_t(\omega)\, e_i(t)\, d\nu(t)$$

とすると、$(\xi_i)_i$ は独立で $\xi_i \sim N(0,1)$ かつ $E[\xi_i \xi_j] = \delta_{ij}$ で、任意の有限な I と任意の $t\in T$ に対し、

$$E\left[\left(X_t - \sum_{i\in I} \xi_i \sqrt{\mu_i}\, e_i(t)\right)^2\right] = k(t,t) - \sum_{i\in I} \mu_i e_i^2(t)$$

であり、特に

$$\int E\left[\left(X_t - \sum_{i\in I} \xi_i \sqrt{\mu_i}\, e_i(t)\right)^2\right] d\nu(t) = \sum_{i\notin I} \mu_i$$

//

- KL-展開は Gauss過程の具体的な構成法を与えている。
- 2つめの主張より $X_t = \sum_i \xi_i \sqrt{\mu_i}\, e_i(t)$ なる展開は $L^2(P\times\nu)$ 内の収束という意味で成り立つ

確認

$$\begin{aligned} E[X_t X_{t'}] &= E\left[\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \xi_i \xi_j \sqrt{\mu_i}\sqrt{\mu_j}\, e_i(t) e_j(t')\right] \\ &= \sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \underbrace{E[\xi_i\xi_j]}_{=\delta_{ij}} \sqrt{\mu_i}\sqrt{\mu_j}\, e_i(t) e_j(t') \\ &= \sum_{i=1}^{\infty} \mu_i e_i(t) e_i(t') = k(t,t') \end{aligned}$$

//

Thm (Driscollの定理)

Hが無限次元 $\Rightarrow$ 確率1で $X_\cdot \notin H$
Hが有限次元 $\Rightarrow$ (Xのあるバージョン$\tilde{X}$が存在して)
確率1で $\tilde{X}_\cdot \in H$ //

(確認)
KL-展開を用いて
$$\|X_\cdot\|_H^2 = \sum_{i=1}^{\infty} \xi_i^2 \mathbb{1}(\mu_i \neq 0)$$
と書ける.

無限次元: $^\forall i$で$\mu_i \neq 0$なので. $\sum_{i=1}^{\infty} \xi_i^2 = \infty$ (a.s.)
無限

有限次元: $^\exists N$で $^\forall i > N$で$\mu_i = 0$なので. $\sum_{i=1}^{N} \xi_i^2 < \infty$ (a.s.)
有限

例: 線形カーネル $k(x,y) = x^T y$

$$X_x = \sum_{i=1}^{d} \xi_i x_i \qquad (\text{つまり } e_i(x) = x_i \text{ とする})$$
$$(\xi_i \sim N(0,1))$$
とすれば.
$$k(x,y) = \mathbb{E}[X_x X_y] = \mathbb{E}\left[\sum_{i=1}^{d}\sum_{j=1}^{d} \xi_i x_i \xi_j y_j\right]$$
$$= \sum_{i=1}^{d}\sum_{j=1}^{d} \delta_{ij} x_i y_j = \sum_{i=1}^{d} x_i y_i = x^T y$$

νの取り方は色々あるが. 例えば多変量の標準正規分布$N(0,I)$とすれば
$(e_i(x))_{i=1}^{d}$は正規直交基底をなす.

以下、$T=\mathbb{R}_+$ の状況を考える。伊藤積分に話を向けてゆく。

Def (Brown運動, Brownian motion, Wiener process)

確率過程 $B=(B_t)_{t\geq 0}$ が次をみたすとき、(標準)Brown運動 (standard Brownian motion) と言う。

(1) B は Gauss過程

(2) $\forall t,s\in[0,\infty)$, $E[B_t]=0$, $E[B_tB_s]=t\wedge s$

(3) a.s. $\omega\in\Omega$ に対して、$t\mapsto B_t(\omega)$ は連続

(連続なパスを持つと言う) //

Cor

B を Brown運動とする。

(1)′ $B_0=0$ a.s.

(2)′ B_t は加法過程

(3)′ $\forall t\geq 0, \forall s>0,\quad B_{t+s}-B_t\sim N(0,s)$

特に、Brown運動は Lévy過程になる。

(1)′〜(3)′ と (3) を Brown運動の定義としても良い。

Proof

(1)′ $E[B_0^2]=0$ より、O.K.

(2)′ $0\leq a<b\leq c<d$ に対し、

$$E[(B_d-B_c)(B_b-B_a)]$$
$$=d\wedge b-c\wedge b-d\wedge a+c\wedge a$$
$$=b-b-a+a=0$$

より、B_d-B_c と B_b-B_a は独立。

(共分散 0 の正規分布ペアは独立)

一般の n についても同様。

(3)′ $B_{t+s}-B_t$ が平均 0 の正規分布に従うことは明らか。

$$E[(B_{t+s}-B_t)^2]=(t+s)\wedge(t+s)-2(t+s)\wedge t+t\wedge t$$
$$=t+s-2t+t=s$$ //

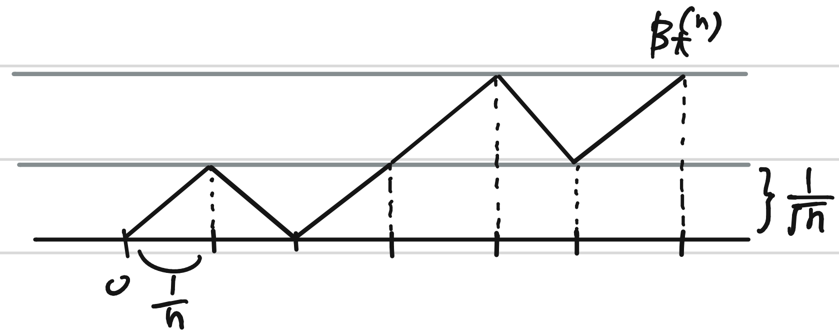
B_t の直感的な構成

$X_1, X_2, \dots$ は i.i.d. で

$$P(X_i=1) = P(X_i=-1) = \frac{1}{2}$$

とする.

$$B_t^{(n)} := \frac{1}{\sqrt{n}}\left(\sum_{k=1}^{\lfloor nt\rfloor} X_k + \left(t - \frac{\lfloor nt\rfloor}{n}\right)\cdot X_{\lfloor nt\rfloor+1}\right) \quad (0 \le t \le 1)$$



(2)' は $n\to\infty$ で $X_{\lfloor nt\rfloor+1}$ 項が無視できて成り立つことがわかる.

ほぼ ns 個

(3)' は $$B_{t+s}^{(n)} - B_t^{(n)} = \frac{1}{\sqrt{n}}\left(\sum_{k=\lfloor nt\rfloor+1}^{\lfloor n(t+s)\rfloor} X_k + O(1)\right)$$

$$\rightsquigarrow N(0,s) \quad (\because \text{中心極限定理})$$

から示される.

(厳密に扱うには $B_t^{(n)}$ が確率過程として B_t に収束することを示す (連続関数空間上の分布収束))

- $(B_t)_{t\in[0,1]}$ の KL-展開

$$k(t,s) = t\wedge s \quad (t,s\in[0,1])$$

より,

$$\int_0^1 k(t,s)e(s)ds = \mu e(t)$$

を解けば良い.

左辺 $= \int_0^1 (t\wedge s)e(s)ds$ より, 両辺 t について微分すると

$$\int_0^1 \mathbb{1}[s \ge t]e(s)ds = \mu e'(t)$$

$$\underset{\text{さらにビブン}}{\Longrightarrow} \quad -e(t) = \mu e''(t)$$

これを解くと、

$$\begin{cases} \mu_i = \dfrac{4}{(2i-1)^2\pi^2} \\ e_i(t) = \sqrt{2}\sin\left(\dfrac{(2i-1)\pi}{2}t\right) \end{cases} \qquad (i=1,2,\dots)$$

が固有値固有関数になる.

KL-展開より、

$$B_t = \sum_{i=1}^{\infty} \sqrt{\mu_i}\, e_i(t)\, \xi_i$$

$$= \sum_{i=1}^{\infty} \frac{2\sqrt{2}}{\pi(2i-1)} \sin\left(\frac{(2i-1)}{2}\pi t\right) \xi_i$$

各周波数成分が どう足し合わさっているか わかる.

(ただし $\xi_i \sim N(0,1)$ (i.i.d.))

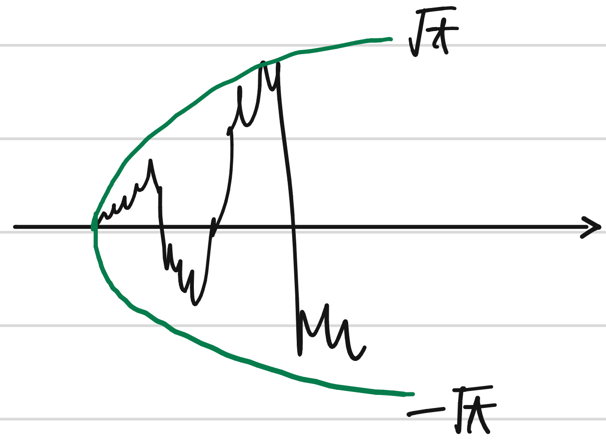
なる表現を持つ

Brown運動の性質

- a.s. $\omega \in \Omega$ に対して、ほとんど全ての t で微分不可能
- $$\lim_{\delta \to 0} \sup_{\substack{|t-s|<\delta \\ t,s\in[0,1]}} \frac{|B_t - B_s|}{|t-s|^{\varepsilon}} = \begin{cases} 0 & (0<\varepsilon<\frac{1}{2}) \\ \infty & (\varepsilon \geq \frac{1}{2}) \end{cases} \quad \text{a.s.}$$

$\longrightarrow$ B_t は大体 $\sqrt{t}$ の大きさで広がる.

(Karatzas and Shreve Thm 2.2.8., Thm 2.9.25)



Cor

$(B_t)_{t\ge 0}$ をBrown運動とする.
すると、以下は全てBrown運動になる:

(1) $X_t = -B_t$

(2) $X_t = B_{t+c} - B_c \qquad (c>0)$

(3) $X_t = \sqrt{c}\, B_{t/c}$

//

(証明は略) ← チェックせよ.

Thm (Lévy-Itoの分解)

Lévy過程 $(X_t)_{t\ge 0}$ に対し、

- 連続関数 $m(t)$ s.t. $m(0)=0$
- 連続・単調非減少関数 $v(t)$ s.t. $s(0)=0$
- Poisson random measure N s.t
  平均測度 μ が
  $$\mu(\{t\}\times(\mathbb{R}-\{0\}))=0,$$
  $$\int_{s\le t}\int_{|u|>0}(u^2\wedge 1)\,\mu(ds\,du)<\infty$$
  をみたす.
- N とは独立なBrown運動 $(B_t)_{t\ge 0}$

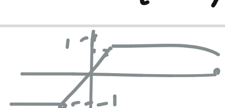
が存在し、

$$X_t = m(t) + B_{v(t)}$$ ← ガウス過程の部分

$$+\lim_{n\to\infty}\iint_{\substack{0\le s\le t\\ |u|\ge \frac{1}{n}}}\bigl(uN(ds\,du)-\phi(u)\mu(ds\,du)\bigr)$$

Poisson過程の部分

と書ける. ただし $\phi(u)=\max\{-1,\min\{u,1\}\}$. //



Lévy過程は大体、Brown運動とPoisson random processで書けるということ.

(伊藤清, 確率論, 定理5.29)