確率場の幾何

統計的発見のための積分幾何学

栗 木　哲

1. はじめに

ベクトル $t=(t_1,\ldots,t_n)$ を添字に持つ確率変数 $X(t)$, $t\in M\subset\mathbb{R}^n$ (n 次元ユークリッド空間) は，確率場 (random field) とよばれる．R. Alder は 1981 年の先駆的な著書[1] で，関数 $t\mapsto X(t)$ が確率 1 でなめらかな関数である場合に，確率場の「幾何化」を通して，確率場にまつわる確率分布，特に最大値の分布の近似計算が可能であることを示した．本稿では，この確率場の幾何の考え方と，そのデータ解析への応用について紹介する．

2. 変化点問題

データ解析への応用の一つは，信号 (変化点) の検出問題である．2013 年 3 月 14 日に，最後の素粒子であるヒッグス粒子 (Higgs boson) の発見がアナウンスされた．図 1 は CERN ホームページで公表されている実験データである．横軸を素粒子の質量として，図の上側には素粒子の観測数とその平滑値，また下側には，観測に伴う系統誤差を差し引いた残差が示されている．下の図において，0 から大きい乖離が見られるとき素粒子の存在が示唆されるが，そこで問題となるのは，どのくらいの乖離があれば大きいと判断してよいか，という閾値の決定問題である．

このような実験データの解析でしばしば行われるのは，その乖離が 2σ (標準誤差 σ の 2 倍) を超えたとき，有意と判断する手順である．しかしながら新素粒子の質量は事前には特定されていないため，広い領域の探索が必要となり，その理由のためしばしば見せかけの発見が生じることが知られている．この現象は，実験物理では LEE (look-elsewhere effect; どこでも効果) とよばれているが，統計学においても検定の多重性として古くから認識されている．

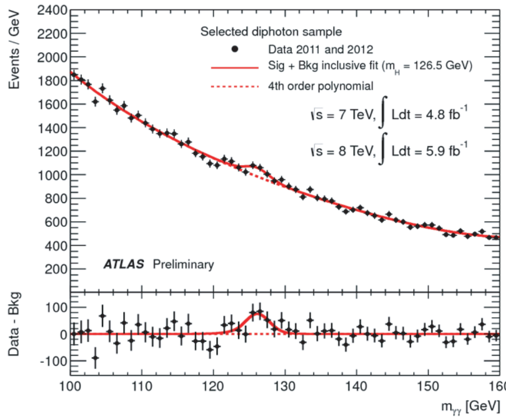


図 1　ヒッグス粒子の探索 (Copyright CERN)

この問題に対して統計的有意性を保証できるような乖離の閾値は，探索する領域の大きさを考慮する形で，統計量の探索領域での最大値 $\max_{t\in M} X(t)$ の確率分布に基づき決定できる．いまの場合 $X(t)$ は図 1 の下図の残差過程である．素粒子実験では経験的に 5σ が閾値として用いられている．本稿で説明する幾何的な方法で最大値 $\max_{t\in M} X(t)$ の

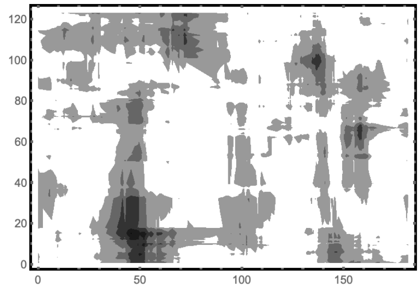


図 2　イネの生殖隔離障壁の探索
(統計量の値を濃淡で表示,
横軸：第 1 染色体，縦軸：第 6 染色体, 単位：cM)

近似分布を求めると，検定の多重性を考慮した p 値の評価ができる[6)].

次に添字集合が 2 次元の確率場の例を図 2 に示す．遺伝学の基本的な仮説として，いくつかの遺伝子が特定の遺伝子型をとる場合，その個体は生育できないような組合せが存在するというものがある．この現象は生殖隔離障壁とよばれ，生物学的種を定義するものと考えられている．イネは 12 の染色体を持つ．ここで紹介する研究では，12 の染色体の 994 遺伝子座について，すべての組合せについて，実際に生存するイネの個体数の，当該組合せが致死遺伝子でないという仮定の下での期待度数からの乖離を，自由度 4 のカイ 2 乗統計量として定量化した．図 2 は第 1 染色体と第 6 染色体の組み合わせについて，カイ 2 乗統計量をプロットしたものである．検証実験によって，図中に濃く示された最大値は真の生殖隔離障壁であり，それ以外の極大点は偽陽性であることが判明している．ここでも確率場の最大値分布を求めることによって，有意性を見積もることができる[3)].

3. スタイナーの公式とミンコフスキー汎関数

応用から一旦離れて，これから確率場の幾何の数理を概観していこう．まずは古典的な積分幾何学の言葉を準備する．M を $\mathbb{R}^N$ の有界な閉集合とする．M からの距離が ρ 以下であるような $\mathbb{R}^N$ の点の全体を，M のまわりの半径 ρ のチューブ

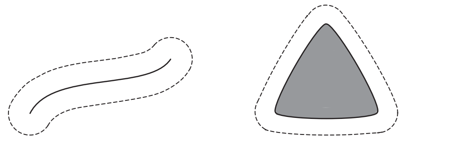


図 3　チューブ

$$\mathrm{Tube}(M,\rho) = \left\{ x \in \mathbb{R}^N \mid \min_{y \in M} \|y - x\| \le \rho \right\}$$

という (図 3).

もし M が凸集合でないならば，ρ の値を 0 から増やしていくと，ある値を境に自己交差が生じる (図 4)．そのような値を臨界半径 ρ_c とよぶ.

半径 ρ が ρ_c 以下の場合，チューブ $\mathrm{Tube}(M,\rho)$ の体積は，M が埋め込まれている外側の空間の次元の次数の多項式で与えられることが知られている．これをスタイナーの公式 (Steiner's formula) という：

$$\begin{aligned} \mathrm{Vol}_N(\mathrm{Tube}(M,\rho)) &= \sum_{j=0}^{N} \rho^j \binom{N}{j} \mathcal{M}_j(M) \\ &= \sum_{j=0}^{n} \rho^{N-j} \omega_{N-j} \mathcal{L}_j(M), \quad \rho \in [0, \rho_c]. \end{aligned} \tag{1}$$

ここで $n = \dim M$，また $\omega_d = \pi^{d/2}/\Gamma(d/2+1)$ は $\mathbb{R}^d$ の単位球の体積．多項式の係数 $\mathcal{M}_j(M)$ を M の j 次ミンコフスキー汎関数 (Minkowski functional)，$\mathcal{L}_j(M)$ をリプシッツ・キリング曲率 (Lipschitz-Killing curvature) とよぶ．$\mathcal{L}_j(M)$ は外側の空間の次元 N に依存せず定義されることに注意する．また，ガウス・ボンネの定理より $\mathcal{M}_N(M)/\omega_N = \mathcal{L}_0(M)$ は集合 M のオイラー標数 $\chi(M)$ に一致する.

ここでオイラー標数について簡単に説明する.

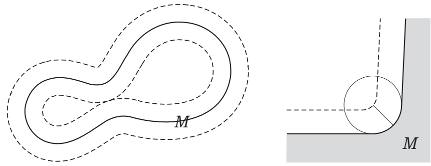


図 4　チューブの自己交差
(左：大域的自己交差，右：局所的自己交差)

有界な閉集合 A のオイラー標数 $\chi(A)$ は，次のように再帰的に定義される：

$$\chi(A)=\begin{cases}0 & (A=\emptyset),\\ 1 & (A\text{ は可縮集合}),\\ \chi(A_1)+\chi(A_2)-\chi(A_1\cap A_2) & \\ & (A=A_1\cup A_2).\end{cases}$$

ここで可縮とは，1 点に連続的に収縮可能 (1 点集合とホモトピー同値) なことをいい，例えば閉球は可縮である．特に集合が 3 次元以下の場合，$\beta_0(A)$ を連結成分数，$\beta_1(A)$ をホールの数 (1 次元集合では 0)，$\beta_2(A)$ を空洞の数 (1,2 次元集合では 0) とおくと，

$$\chi(A)=\beta_0(A)-\beta_1(A)+\beta_2(A)$$

がなりたつ．

ところでスタイナーの公式 (1) の右辺は，ρ が ρ_c を超えた場合にも意味を持つ．中心を x, 半径 ρ の閉球を $B(x,\rho)$ で表す．x の集合 M からの距離が ρ 以下であることと $M\cap B(x,\rho)$ が空集合でないことは同値であるため，$\mathrm{Vol}_N(\mathrm{Tube}(M,\rho))=\int_{\mathbb{R}^N}\mathbb{1}(M\cap B(x,\rho)\neq\emptyset)\mathrm{d}x$，ここで関数 $\mathbb{1}(\cdot)$ は引数の命題の真偽に応じて 1 または 0 を返す関数である．さらに $\rho\le\rho_c$ の範囲では，集合 $M\cap B(x,\rho)$ は空集合でなければ常に可縮でオイラー標数が 1 となるため，$\mathbb{1}(M\cap B(x,\rho)\neq\emptyset)=\chi(M\cap B(x,\rho))$ となる．したがってスタイナーの公式は

$$\int_{\mathbb{R}^N}\chi(M\cap B(x,\rho))\mathrm{d}x=\sum_{j=0}^{n}\rho^{N-j}\omega_{N-j}\mathcal{L}_j(M),$$

$n=\dim M$ と書き換えられる．実はこれはすべての $\rho\ge 0$ でなりたつ積分幾何学の基本定理 (Kinematic fundamental formula) である．

4. ガウス確率場の幾何

2 次元の添字集合上のガウス確率場 $X(t)$ の一例を図 5 (左) に示す．また，図 5 (右) は，確率場の値がある閾値以上となるような添字の集合である．このような集合を，エクスカーション集合と

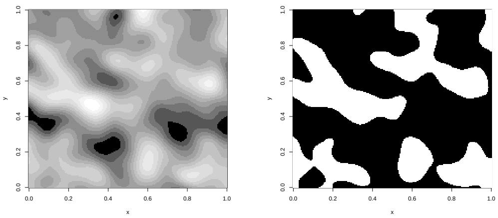


図 5 確率場 (左) とエクスカーション集合 (右)

いう．

より正確には，$X(t)$ の値がある閾値 u 以上となるような添字 t の上側レベル集合

$$M_u=\{t\in M\mid X(t)\ge u\}=X^{-1}([u,\infty))$$

がエクスカーション集合である．当然ランダムな集合である．

これから，確率場 $X(t)$ そのものでなく，このエクスカーション集合に着目する．とくにそのオイラー標数とミンコフスキー汎関数に着目する．

例えば図 5 (右) では，黒地の連結成分数が 3, 白地のホールの数が 1 であり，オイラー標数は $\chi=3-1=2$ である．ここでホールとは，エクスカーション集合に周囲を囲まれた領域である．

これらの幾何量を定量的に扱うために，以下の手順によって添字集合 M に幾何学的な構造を付与する．まず $X(t)$, $t=(t_1,\ldots,t_n)\in M$ を添字集合 $M\subset\mathbb{R}^n$ の上で定義されたガウス確率場とする．ガウス確率場は，平均関数と共分散関数で定まる確率場である．ここではガウス確率場の中で，t を固定したときの $X(t)$ の周辺分布が平均 0，分散 1 の標準正規分布 $N(0,1)$ であるものを考える．共分散関数を $r(s,t)=\mathrm{Cov}(X(s),X(t))$ とおく．

また確率 1 でサンプルパス (関数 $t\mapsto X(t)$) は滑らかであるとする．さらに，微分確率場を $\nabla X(t)=(\partial X(t)/\partial t_i)_{1\le i\le n}$ とおく．

添字集合 M を，その各点で計量

$$g_{ij}(t)=\mathrm{Cov}\Big(\frac{\partial X(t)}{\partial t_i},\frac{\partial X(t)}{\partial t_j}\Big)=\frac{\partial^2 r(s,t)}{\partial s_i t_j}\Big|_{s=t} \tag{2}$$

が付与されたリーマン多様体と考える．この計量

は以下のような理由に基づく．

ψ を $M \subset \mathbb{R}^n$ から $\mathbb{R}^N$ の単位球面 $\mathbb{S}^{N-1}$ への滑らかな関数とする．$\xi = (\xi_1, \ldots, \xi_N)$ を各成分が独立に標準正規分布に従うランダムベクトルとし，M を添字集合とする確率場

$$X(t) = \langle \xi, \psi(t) \rangle_{\mathbb{R}^N}, \quad t = (t_1, \ldots, t_n) \in M$$

を考える．ここで $\langle \cdot, \cdot \rangle_{\mathbb{R}^N}$ は内積である．このとき $X(t)$ は平均 0，分散 1 のガウス確率場で，共分散関数は $\mathrm{Cov}(X(s), X(t)) = \langle \psi(s), \psi(t) \rangle_{\mathbb{R}^N}$．像 $\psi(M)$ の $\mathbb{R}^N$ の部分多様体としての計量は $g_{ij}(t) = \langle \partial\psi(t)/\partial t_i, \partial\psi(t)/\partial t_j \rangle_{\mathbb{R}^N}$ であり，これは (2) に一致する．

より一般に平均 0 のガウス確率場は，関数族 $\{r(\cdot, t) \mid t \in M\}$ で張られる再生核ヒルベルト空間の正規直交基底 $\{\psi_k\}_{k \geq 1}$ を用いて

$$X(t) = \sum_{k=1}^{\infty} \xi_k \psi_k(t)$$

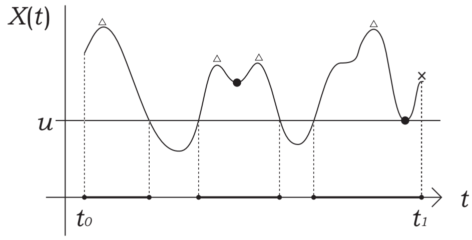
と表される (Karhunen-Loéve 展開)．ただし $\xi_1, \xi_2, \ldots$ は独立に標準正規分布に従う確率変数列である．$\psi(t) = (\psi_k(t))_{k \geq 1} \in \ell^2$ とおく．(ℓ^2 は 2 乗和が有限な数列の全体．）いま写像 $t \mapsto \psi(t)$ が M から ℓ^2 への滑らかな単射である場合，ℓ^2 の標準的内積によって，M に計量が

$$g_{ij}(t) = \sum_{k=1}^{\infty} \frac{\partial \psi_k(t)}{\partial t_i} \frac{\partial \psi_k(t)}{\partial t_i}$$

の形で誘導されるが，これは (2) に他ならない．$\mathrm{Var}(X(t)) = \|\psi\|_{\ell^2}^2 = 1$ であるので，像 $\psi(M)$ は ℓ^2 の単位球面の部分多様体である．

一般にリーマン多様体 M は，ナッシュの埋め込み定理によって，計量を保ったまま M を十分次元の高いユークリッド空間 $\mathbb{R}^N$ に埋め込むことができる．このとき，チューブ体積を考えるとスタイナーの公式 (1) がなりたつが，係数 $\mathcal{L}_j(M)$ は埋め込み方法に依存しないで定義される．(1) に対応して，エクスカーション集合のオイラー標数 $\chi(M_u)$ の期待値公式がなりたつ：

$$\mathbb{E}[\chi(M_u)] = \sum_{j=0}^{n} \frac{1}{(2\pi)^{\frac{j+1}{2}}} \mathcal{L}_j(M) H_{j-1}(u) e^{-\frac{u^2}{2}}, \qquad (3)$$



$\chi(M_u) = 4\,(\triangle) - 2\,(\bullet) + 1\,(\times) = 3$

図 6　エクスカーション集合とそのオイラー標数

$n = \dim M$．ここで $H_k(x)$ はエルミート多項式である．この証明は付録で行う．

なお前節と本節では，M をユークリッド空間 $\mathbb{R}^N$ に埋め込みチューブ領域を考えた．一方で分散 1 の確率場の添字集合 M は単位球面 $\mathbb{S}^{N-1}$ へ埋め込めるため，単位球面上でのチューブ (球面チューブ) を定義し，その体積から出発しても平行的な議論ができる．その方法論はチューブ法 (tube method) とよばれる[5,7,8]．

5. 最大値分布の近似とその応用

本節では式 (3) の統計的な意味を，図 6 の 1 次元の確率場に即して説明する．確率場の最大値の裾確率は (3) によって近似できるという関係式

$$\Pr\left(\max_{t \in M} X(t) \geq u\right) \approx \mathbb{E}[\chi(M_u)] \qquad (u \text{ が大きいとき}) \qquad (4)$$

を主張したい．

まずエクスカーション集合の定義から，$X(t)$ が u 以上となることは M_u が空集合とならないことと同値で，さらにそれは M_u の連結成分の個数が 0 でないことと同値である．すなわち

$$\max_{t \in M} X(t) \geq u \;\Leftrightarrow\; M_u \neq \emptyset \;\Leftrightarrow\; \chi(M_u) \geq 1$$

である．ここで u の値が大きいとき，$X(t)$ が複数回 u を超える事象は稀であると考えられる．すなわち事象 $\chi(M_u) = k$ $(k \geq 2)$ が起こる確率は小

さいことが期待できる．この近似のもとでは，

$$\mathbb{1}\left\{\max_{t\in M} X(t) \geq u\right\} - \chi(M_u)$$
$$= -\sum_{k=2}^{\infty}(k-1)\mathbb{1}(\chi(M_u)=k) \qquad (5)$$

が無視できる量となり，両辺の期待値をとることにより (4) がなりたつことが期待される．

M が 2 次元以上の場合でも，u が大きいときは，M_u は空集合でない限り可縮集合であることが期待でき，その仮定のもとで $\mathbb{1}\{\max_{t\in M} X(t) \geq u\} \approx \chi(M_u)$ である．

近似式 (4) の相対誤差は，多くの場合

$$\frac{\Pr\left(\max_{t\in M} X(t) \geq u\right)}{\mathbb{E}[\chi(M_u)]} - 1 = O\left(u^K e^{-\rho_c^2 u^2/2}\right),$$

$u \to \infty$ と評価される．(K はある整数．) $\rho_c > 0$ はチューブの臨界半径で，この値が大きいと相対誤差は u が大きくなるにつれて急速に小さくなる．

近似法 (4) をオイラー標数法 (Euler characteristic heuristic) という．信号処理分野のライス公式の一般化でもある．(4) と (3) を組み合わせて変化点問題に用いることができる．その例を示そう．

6. 変化点問題への応用

図 1 の素粒子のカウントデータに戻ろう．横軸 (素粒子の質量) を多くの幅の狭い区間に細分し，その m 番目の区間内のカウント数を N_m とおく．ここで N_m はポアソン分布に従うとし，その平均を

$$\mathbb{E}[N_m] = \beta_m(\theta) + \kappa_m(\phi)\mu_m$$

とする．ここで $\beta_m(\theta)$ はバックグラウンドのカウント数の平均値で，$\kappa_m(\phi)$ は目的の素粒子のカウント数である．これらの関数に含まれるパラメータは事前に推定するものとする．また $\mu_m\ (\geq 0)$ は感受性 (sensitivity) を表すパラメータで，特に $\mu_m = 0$ の場合は，質量 m の素粒子が存在しないというモデルに対応する．

このモデルにおいて，すべての m について μ_m の最尤推定量 $\widehat{\mu}_m$ と，モデルのフィッシャー情報行列から計算される標準誤差の推定量 s.e.$(\widehat{\mu}_m)$ を求める．このとき $T(m) = \widehat{\mu}_m/\text{s.e.}(\widehat{\mu}_m)$ は帰無仮説 $H_m : \mu_m = 0$ を対立仮説 $A_m : \mu_m > 0$ に対して検定する検定統計量であり，その値が大きいときにはその質量の素粒子の存在が示唆される．ここで問題となるのはその閾値である．m は離散的であるが多数であるので連続パラメータとみなすと，確率場の幾何の方法で，$\mu_m \equiv 0, \forall m$ の仮定のもとで裾確率が

$$\overline{F}(c) = \Pr\left(\max_{m\in M} T(m) \geq c\right)$$
$$\approx \frac{1}{2\pi}\mathrm{Vol}_1(M)e^{-c^2/2} + \frac{1}{\sqrt{2\pi}}\int_c^{\infty} e^{-u^2/2}\mathrm{d}u$$

の形でえられる (式 (8))．ただし $\mathrm{Vol}_1(M)$ は M の長さで，$\mathrm{Cov}(T(s), T(t))$ より数値的に計算する．この上側確率 $\overline{F}$ の近似公式を用いて，実際のデータから得られる $T(m)$ の値の有意性を

$$\text{多重性調整}\ p\ \text{値} = \overline{F}\big(T(m)\ \text{の観測値}\big)$$

と評価することができる．この p 値が，統計的発見 “質量 m の素粒子が存在する” が偽陽性である確率である．

7. ミンコフスキー汎関数によるガウス性の検定

いままでは，確率場 $X(t)$ のエクスカーション集合の期待オイラー標数を用いた最大値分布近似について論じてきた．最後に確率場のミンコフスキー汎関数・オイラー標数の別の応用を紹介する．

図 7 は，シミュレーションにより生成した 2 次元ガウス確率場のエクスカーション集合のオイラー標数のプロット (オイラー曲線) $\chi(E_v)$ (左) と非ガウス確率場のオイラー曲線 (右) である．ここで非ガウス確率場として自由度 2 のカイ 2 乗確率場を用いている．図中の破線は，ガウス性を仮定したときの期待値 $\mathbb{E}[\chi(E_v)]$ である．破線からの乖離は非ガウスのときに大きく，オイラー曲線が，確率場のガウス性の検定に用いることができることが分かる．

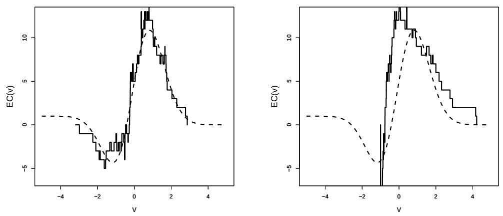


図 7　ガウス確率場 (左) と非ガウス確率場 (右) のオイラー曲線
実線：$\chi(E_v)$ (統計量), 破線：$\mathbb{E}[\chi(E_v)]$ (ガウス性のもとでの期待値)

このようなガウス性の検定は，宇宙論において重要である．宇宙マイクロ波放射 (CMB) は，宇宙の温度ゆらぎを観測しているものであるが，いままでの研究で，この信号は等方的なガウス確率場に非常に近いものであることが分かっている．それが完全にガウスであるか，もしそうでない場合非ガウス性はどのようなものであるかは初期宇宙モデルで決定される．そこで用いられているガウス性の検定統計量の一つが，エクスカーション集合のミンコフスキー汎関数，特にオイラー標数 (ジーナス統計量) である．

平均 0 のガウス確率場は共分散関数 (2 点相関関数) だけで決定される．Matsubara[4] は，非ガウスの下でのオイラー標数の挙動の解析のために，ガウス確率場ならば 0 となる 3 次相関関数の存在を仮定したときの期待オイラー標数を摂動展開の形で導出し，初期宇宙モデルとの対応を調べている．さらに高次の相関関数の存在を仮定した摂動展開も可能である (投稿準備中)．

付録：モースの定理とカッツ・ライス公式

ここでは，M が 1 次元の場合と 2 次元以上の場合について (3) を証明する．

図 6 は 1 次元の区間 $M=[t_0,t_1]$ を添字集合にする確率場の例である．$\chi(M_u)$ は M_u の連結成分の個数で，図 6 では 3 である．また連結成分の個数は，(i) 内点 $t\in(t_0,t_1)$ で $X(t)\geq u$ かつ $X(t)$ の極大点である t の個数 N_+，(i') 内点 $t\in(t_0,t_1)$ で $X(t)\geq u$ かつ $X(t)$ の極小点である t の個数 N_-，(ii) $t=t_0$ または t_1 で $X(t)\geq u$ かつ外向き勾配 $-\dot X(t_0)$ または $\dot X(t_1)$ が正である点の個数 N'_+ を用いて，

$$\chi(M_u)=N_+-N_-+N'_+ \tag{6}$$

である．図 6 の場合は，$N_+=4$, $N_-=2$, $N'_+=1$, $\chi(M_u)=4-2+1=3$ となる．(6) はモースの定理の 1 次元版である．(i), (i'), (ii) で数え上げている点は，モース関数 $-X(t)$ に対する臨界点である．

さらに関係式 (6) は次のようにデルタ関数で記述できる．

$$\begin{aligned}\chi(M_u)=&\int_{t_0}^{t_1}\mathbb{1}(X(t)\geq u)(-\ddot X(t))\delta(\dot X(t))\mathrm{d}t\\&+\mathbb{1}(X(t_0)\geq u)\mathbb{1}(\dot X(t_0)<0)\\&+\mathbb{1}(X(t_1)\geq u)\mathbb{1}(\dot X(t_1)>0).\end{aligned} \tag{7}$$

これが (1 次元版) カッツ・ライス公式である．

ガウス確率場の微分確率場はガウス確率場であること，また $\mathbb{E}[X(t)]=0, \mathbb{E}[X(t)^2]=1$ の両辺を t で繰り返し微分すると $\mathbb{E}[\dot X(t)]=0, \mathbb{E}[\ddot X(t)]=0,$ $\mathbb{E}[\dot X(t)X(t)]=0, \mathbb{E}[\ddot X(t)X(t)]+\mathbb{E}[\dot X(t)\dot X(t)]=0$ となるから，$g(t)=\mathbb{E}[\dot X(t)\dot X(t)]$, $R(t)=\ddot X(t)+g(t)X(t)$ とおくと，(固定した t に対して) $X(t)$, $\dot X(t)$, $R(t)$ は独立な平均 0 のガウス変量となる．これより期待値は

$$\begin{aligned}\mathbb{E}[\chi(M_u)]=&\int_{t_0}^{t_1}\mathbb{E}[\mathbb{1}(X(t)\geq u)(-R(t)\\&\qquad+g(t)X(t))]\mathbb{E}[\delta(\dot X(t))]\mathrm{d}t\\&+\Pr(X(t_0)\geq u)\Pr(\dot X(t_0)<0)\\&+\Pr(X(t_1)\geq u)\Pr(\dot X(t_1)>0)\end{aligned}$$

と書けることが分かる．$X(t)\sim N(0,1)$, $\dot X(t)\sim N(0,g(t))$, $\mathbb{E}[R(t)]=0$ に注意し，また $\mathbb{E}[\delta(\dot X(t))]$ は $\dot X(t)$ の密度関数の原点での値 $(2\pi g(t))^{-1/2}$ となるので

$$\begin{aligned}\mathbb{E}[\chi(M_u)]=&\frac{1}{2\pi}\int_u^\infty ue^{-u^2/2}\mathrm{d}u\,\mathcal{L}_1(M)\\&+\frac{1}{\sqrt{2\pi}}\int_u^\infty e^{-u^2/2}\mathrm{d}u\,\mathcal{L}_0(M),\end{aligned} \tag{8}$$

ただし

$$\mathcal{L}_1(M) = \mathrm{Vol}_1(M) = \int_{t_0}^{t_1} g(t)^{1/2}\mathrm{d}t,$$
$$\mathcal{L}_0(M) = 1$$

を得る．これは $n=1$ のときの (3) に一致する．

M の次元が 2 以上の場合は，(6), (7) を一般化する必要がある．図 8 は 2 次元確率場を示している．M として，区分的に滑らかな境界を持った多様体 (層化多様体) $M = \bigsqcup_{j=0}^{d} \partial_j M$ ($\dim \partial_j M = j$) を考える[2, 5]．ここで $\partial_d M$ は M の内点集合 $\mathrm{int}M$，M の境界は $\partial M = \bigsqcup_{j=0}^{d-1} \partial_j M$ である．例えば $\mathbb{R}^3$ の閉直方体は層化多様体で，$\partial_3 M$ は長方体の内点集合，$\partial_2 M$ は 6 つの面 (辺は含まない)，$\partial_1 M$ は 12 本の辺 (頂点は含まない)，$\partial_0 M$ は 8 つの頂点となる．

(i) $t \in \mathrm{int}M$ のときは，$X(t) \geq u$ かつ $\nabla X(t) = 0$ のとき点 t を臨界点と定義する．そのような臨界点で，ヘッセ行列 $\nabla^2 X(t)$ の正の固有値が偶数個のものの個数を N_+，(i') 奇数個のものの個数を N_- とする．

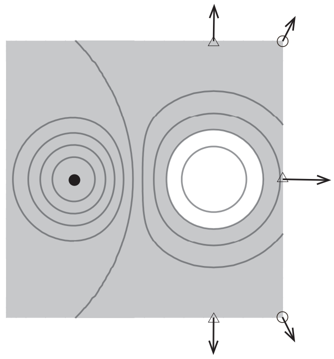
(ii) $t \in \partial_j M$ のときは，まず M の点 t における法錐 (接錐の双対錐) を $N_t(M)$ とおく．$X(t) \geq u$ かつ，$\partial_j M$ に制限された勾配 (接空間 $T_t \partial_j M$ 方向の勾配) $\nabla|_{\partial_j M} X(t)$ を 0 とし，また勾配ベクトル $\nabla X(t)$ が錐 $N_t(M)$ に含まれる場合，t を臨界点と定義する．制限されたヘッセ行列 $\nabla|^2_{\partial_j M} X(t)$ の正の固有値が偶数個のものの個数を N'_+，(ii') 奇数個のものの個数を N'_- とする．

このとき (6) に対応するモースの定理は

$$\chi(M_u) = N_+ - N_- + N'_+ - N'_-$$

である．またカッツ・ライス公式は

$$\begin{aligned}\chi(M_u) &= \int_{\mathrm{int}M} \mathbb{1}(X(t) \geq u) \\ &\quad \times \det(-\nabla^2 X(t))\delta(\nabla X(t))\mathrm{d}t \\ &\quad + \sum_{j=0}^{d} \int_{\partial_j M} \mathbb{1}(X(t) \geq u, \nabla X(t) \in N_t(M)) \\ &\quad \times \det\bigl(-\nabla|^2_{\partial_j M} X(t)\bigr)\delta\bigl(\nabla|_{\partial_j M} X(t)\bigr)\mathrm{d}t'\end{aligned}$$



$$\chi(M_u) = 1\,(\bullet) - 3\,(\triangle) + 2\,(\circ) = 0$$

図 8　エクスカーション集合とそのオイラー標数 (2 次元) (曲線は確率場の等高線，$\bullet$ は最大点，矢印は $\nabla X(t)$)

($\mathrm{d}t'$ は点 t における $\partial_j M$ の体積要素) と一般化される．この期待値をとることにより (3) を導くことができる．導出の詳細は 8) を参照．

本稿に有益なコメントをいただきました下平英寿先生，松原隆彦先生に感謝いたします．

参考文献

1) R. J. Adler. *The geometry of random fields.* John Wiley & Sons, Ltd., Chichester, 1981.
2) R. J. Adler and J. E. Taylor. *Random fields and geometry.* Springer, New York, 2007.
3) S. Kuriki, Y. Harushima, H. Fujisawa, and N. Kurata. Approximate tail probabilities of the maximum of a chi-square field on multi-dimensional lattice points and their applications to detection of loci interactions. *Ann. Inst. Statist. Math.*, Vol. 66, No. 4, pp. 725–757, 2014.
4) T. Matsubara. Statistics of smoothed cosmic fields in perturbation theory. I. Formulation and useful formulae in second-order perturbation theory. *Astrophysical J.*, Vol. 584, pp. 1–33, 2003.
5) A. Takemura and S. Kuriki. On the equivalence of the tube and Euler characteristic methods for the distribution of the maximum of Gaussian fields over piecewise smooth domains. *Ann. Appl. Probab.*, Vol. 12, No. 2, pp. 768–796, 2002.
6) D. A. van Dyk. The role of statistics in the discovery of a higgs boson. *Annu. Rev. Stat. Appl.*, Vol. 1, pp. 41–59, 2014.
7) 栗木哲. チューブ法の理論・応用とその周辺. 統計数理, Vol. 67, No. 2, pp. 229–240, 2019.
8) 栗木哲, 竹村彰通. チューブの体積と正規確率場の最大値の分布. 数学, Vol. 60, No. 2, pp. 134–155, 2008.

（くりき・さとし，統計数理研究所）