目次

第 1 章

枠組み

統計モデル自体を形式的に定義したい．が，それを解析する手段は，幾何学的方法と解析学的方法の 2 つが存在し，互いに密接に，双対的に深く映りあっている．母数の空間は，統計モデルを多様体にする基礎空間となると同時に，それ自体位相線型空間である．あるいはこの考え方は，従来の数学という営み全体を，データの観点から捉え直す契機になることさえあるかもしれない．

記法 1.0.1. (1) Y を response variable とする．ベクトルである．

(2) X を covariates のなすベクトルとする．

(3) Z_i は i 番目の unit に対応する確率ベクトルとする．$Z_1, \cdots, Z_n$ は互いに独立同分布に従うとする．

(4) Z はある一回の観察に対応する確率ベクトルとする．

(5) 確率変数の realization には小文字 x, y を用いる．

(6) $h \in H$ の直交分解 $u_0 + u_1 \in U + U^\perp$ を，$u_0 = \Pi(h|U)$ と表す．

1.1 統計モデル

統計モデルという語の枠組みは明瞭になっているが，その選択に関しては，識別可能性程度しか概念が成熟していない．統計モデルの推論における有用さを，拡張可能性や自然性を圏論的に捉える試みがある．

1.1.1 統計モデルの枠組み

この枠組みに落とし込むまでに，ドメイン知識と mentalism が必要になる．[a]全てのモデルは間違っている．しかしその中でも，モデルは無限次元である方が表現力・応用可能性・頑健性がある．が，数理的な扱いは困難になる．そこで，セミパラという希望の光が差し込む．

[a] こういうところで，データ生成過程に対するドメイン知識が必要になる．統計学者のデイヴィッド・コックス卿は、「対象となる問題から統計モデルへの変換をどのように行うかが、分析の最も重要な部分であることが多い」と述べている。

定義 1.1.1 (sample space, parameter, experiment).

(1) 可測空間 X を**標本空間**という．

(2) (位相／可測) 線型空間 Θ の元を**母数**という．

(3) 可測な族 $(P_\theta)_{\theta\in\Theta} : \Theta \to \Delta(X)$ を**統計モデル**という．[†1]

(4) 組 $(X, (P_\theta))$ を**統計的実験**という．

定義 1.1.2 (identifiable, parametric, nonparametric, semiparametric).

(1) 族 $(P_\theta)_{\theta\in\Theta} : \Theta \to \Delta(X)$ が単射であることを，**識別可能**という．

[†1] $\Delta(X) \subset P(X)$ は単体である．$P(X) \subset M(X) := (C(X))^*$ は w^*-コンパクトな凸集合であることに注意．

(2) モデル $\mathcal{F}:\Theta\to\mathrm{Meas}(\mathcal{X},[0,1])$ の添字集合 Θ について，$\exists_{k\in\mathbb{N}}\;\Theta\subset\mathbb{R}^k$ を満たす（有限次元）パラメトリックモデルと，そうでないノンパラメトリックモデルに別れる．
(3) 母数が $\theta=(\beta,\eta)\in\Theta_1\times\Theta_2=\Theta$ という形に分解でき，Θ_1 が有限次元で Θ_2 が無限次元[†2]であるモデルを，**セミパラメトリックモデル**という．パラメトリックモデルとノンパラメトリックモデルは，セミパラメトリックモデルの退化と捉えたい．[†3]

要諦 1.1.3. ほとんどの場合，実用的には「平均治療効果」など，特定の有限個の母数にのみ興味がある．人間のリソースは有限であるからだ．しかし，モデルとしては無限である方が祈りが高い．実際，自然な関数空間は押し並べて無限次元である．そこで現れる形式がセミパラメトリックモデルである．

1.1.2　統計モデルを創るまで

What Is A Statistical Model?から作った節．

定義 1.1.4 (statistical unit, covariate / independent variable, response)**.** 統計的実験 $(X,(P_\theta))$ や観察実験は，次の 3 要素からなる．

(1) 統計的単位全体の集合 U.
(2) 独立変数全体の集合 Ω.[†4]
(3) 従属変数全体の集合 V.[†6]

独立変数や施策 (treatment) の配分 $\chi:U\to\Omega$ を実験計画といい，その全体 $D:=\Omega^U$ を計画空間という．あり得る結果の付値全体 $X:=V^U$ を標本空間という．確率分布は計画に依存することに注意．

1.1.3　統計という営み

定義 1.1.5 (statistic)**.**

(1) 標本空間 X 上の写像（特に X 上の確率測度の構造に依存しない写像）を**統計量**という．

1.2　分布という考え方

可測空間とは，汎関数が備わり得る空間である．確率密度関数などが考えられるが，一般の空間に拡張する際には，確率密度なる多様体的な幾何学的な概念にとって代わられる．統計モデルを幾何学的にみるか，解析的にみるかの双対性が成り立つはずである．
この視点からすると，測度＝分布とは，極めて幾何学的な設定でもある．「積分されるべきもの」という観点からは共通である．幾何と解析の双対性の基本がここに観て取れる．

超関数が distribution というのと同様，分布とは／測度とは関数の一般化であり，汎関数に他ならない．

定義 1.2.1 (volume, density, distributional density, generalized smooth function)**.**

(1) 可微分多様体上の，至る所消えない最高次元の微分形式 (top-dimentional form) を**体積形式**という．
(2) 可微分多様体上の**密度（形式）**とは，密度バンドルの切断をいう．作用素の視点からは，密度とは，座標変換に関して Jacobian の絶対値倍の変換を受ける関数の族である．向き付け可能な多様体について，密度は n-形式＝体積形式 dvol と標

[†2] 関数や測度など
[†3] むしろいくつかの特徴量で特徴付けられるような統計モデルの方が特殊なのである．
[†4] 共変量ともいう．説明変数，予測因子，独立変数ともいう．「より狭義の意味では、covariate は従属変数と最も関心のある独立変数との関係に影響を与える二次的な変数である」[†5]
[†6] 応答変数 (response variable) ともいう．is a concept, idea, or quantity that someone wants to measure.

準的に同一視できる．密度は体積形式の一般化である．

(3) 隆起関数の空間 $C_c^\infty(X)$ 上の線型汎函数を**密度分布**という．これは多様体上の密度の概念の一般化と捉えられる．

(4) 密度分布は可微分写像の極限と考えられ，この観点に立った際には**超関数**という．

例 1.2.2.

(1) 確率密度とは密度の一種であり，積分の結果確率測度を生むものをいう．

(2) 確率密度関数は，空間を可測空間に限定する．

1.3 回帰分析

推定論の基本として，目的変数と説明変数の間の関数関係に fitting するという指針の分析を，開基分析という．

歴史 1.3.1. 歴史上初めに行われた回帰分析は，Legendre と Gauss による最小二乗法である．回帰という用語は Francis Galton 1822-11（従兄に Darwin を持つ）による[†7]．Galton は Darwin に感化され，自身は統計学を武器に生物学へ切り込んだ．Galton Board を用いて中心極限定理を実感しながら，「平均への回帰」や「凡庸性の原理」を議論した．平均への回帰は相関係数が低い際に起こる普遍的現象であった．

1.4 生存分析

[†7] Cedric Villani の TED Talk "It keeps on occuring again and again for many theories and many experiments as a great example of the universality which is so dear to us mathematicians." このガウス分布についての Francis Galton 1822-11 の言葉が象徴的すぎる．“It would have been deified by the Greeks if they had known it. It is the supreme law of unreason.” Galton’s Board なんてあるのか．

第 2 章

パラメトリックモデル

2.1 正則なモデルの定義と十分条件

記法 2.1.1. $(\mathcal{X},\mathcal{B})$ 上のモデル $\mathcal{P}\subset P(\mathcal{X},\mathcal{B})$ を考える．

定義 2.1.2. $\mathcal{P}=(P_\theta)_{\theta\in\Theta}$ が次を満たすとき，パラメトリックモデルであるという：

(1) $\exists_{k\in\mathbb{R}}\ \Theta\subset\mathbb{R}^k$.
(2) (識別可能性) $\Theta\to\mathcal{P}$ は全単射.

さらに次の条件を課す.

(3) $\Theta\overset{\text{open}}{\subset}\mathbb{R}^k$.
(4) 各 P_θ は σ-有限な参照測度 ν に対して絶対連続.

このとき，

$$p_\theta:=\frac{dP_\theta}{d\nu},\quad l_\theta:=\log p_\theta,\quad s_\theta:=\sqrt{p_\theta}$$

とおく．P_θ を p_θ,s_θ と同一視することで，Banach 空間 $L_1(\mathcal{X},\mathcal{B},\nu)$ または Hilbert 空間 $L_2(\mathcal{X},\mathcal{B},\nu)$ の部分集合とパラメトリックモデル $\mathcal{P}$ とを同一視出来る．いずれの場合も $\mathcal{P}$ に同値な位相を定める．Hellinger の距離 $d_H(P_{\theta_1},P_{\theta_2}):=\|s_{\theta_1}-s_{\theta_2}\|_{\nu,2}$ によって距離化出来る．

この上さらに次の 3 条件を満たすとき，パラメトリックモデルは正則であるとする．

(5) パラメータ付け $\Theta\to\mathcal{P}\hookrightarrow L_2(\nu);\theta\mapsto s_\theta$ は $L_2(\nu)$-微分可能：$\exists_{\dot{s}_\theta\in L^2(\nu;\mathbb{R}^k)}\ \|s_{\theta+h}-s_\theta-h^\top\dot{s}_\theta\|_{\nu,2}=o(\|h\|)\ (h\to\infty)$．以降，微分を $\dot{s}_\theta$ で表す.
(6) Fisher 行列は可逆：$\displaystyle\int\dot{s}_\theta\dot{s}_\theta^\top d\nu\in\mathrm{GL}_k(\mathbb{R})$.
(7) 合成 $\Theta\to\mathcal{P}\to\mathbb{R};\theta\mapsto s_\theta\mapsto\dot{s}_{\theta_i}\ (i\in[k])$ は $L_2(\nu)$-連続：$\forall_{\theta\in\Theta}\ E_\nu[\|\dot{s}_{\theta+h}-\dot{s}_\theta\|^2]\to0\ (h\to0)$.

なおこのとき，$\mathcal{P}$ は k-次元多様体をなしている．

定義 2.1.3. $I_\theta:=4E_\nu[\dot{s}_\theta\dot{s}_\theta^\top]\in\mathrm{GL}_k(\mathbb{R})$ を Fisher 情報行列という．

補題 2.1.4 (正則性の十分条件)**.** 任意の $\theta\in\Theta\overset{\text{open}}{\subset}\mathbb{R}^k$ が (4) までを満たす上に，次が成り立つとき，$(P_\theta)_{\theta\in\Theta}$ は正則である：

(1) ν-a.e.x に対して p_θ は微分可能で連続な勾配 $\dot{p}_\theta$ を持つ．
(2) $\dot{l}_\theta(x):=2\dfrac{\dot{p}_\theta(x)}{p_\theta(x)}1_{p_\theta>0}(x)$ について，$\|\dot{l}_\theta\|\in L_2(P_\theta)$.
(3) I_θ は可逆で θ に関して連続.

補題 2.1.5 (lemma for Cramer-Rao)**.** 次の 2 条件は同値であり，パラメトリックモデルが正則ならば成り立つ：

(1) $P_\theta\dot{l}_\theta=0$.
(2) $\forall_{i\in[k]}\ \langle\dot{s}_{\theta_i},s_\theta\rangle=0$.

2.2　正則推定量

2.2.1　正則推定量

正則推定量は，Fisher 情報行列の逆を分散とする正規分布に従う X と，ある分布 M_θ に従う Y という独立な 2 つの確率変数の和に漸近同等である．この分解により，$M_\theta = 0$ のときに漸近分散が最小になることが従う．

定義 2.2.1 (locally normal)**.** 汎関数 $\nu:\Theta\to\mathbb{R}^m$ の推定量 (T_n) が $\theta\in\Theta$ において **(局所) 正則**であるとは，$\sqrt{n}\|\theta_n-\theta\| = O(1)$ を満たす任意のパラメータ列 (θ_n) に対して，弱収束 $\sqrt{n}(T_n-\nu(\theta_n)) \xrightarrow{P_{\theta_n}} L_\theta$ が成り立つことをいう．

要諦 2.2.2. 任意の方向ベクトル $h\in\mathbb{R}^k$ に対して，

$$\sqrt{n}\left(T_n-\psi\left(\theta+\frac{h}{\sqrt{n}}\right)\right) \xrightarrow{\theta+\frac{h}{\sqrt{n}}} L_{\theta,h}$$

なる弱収束極限が存在するとする．正則性とは，$L_{\theta,h}$ が十分近傍では θ にも依らないことをいう．あるいは，$\dot{\psi}_\theta h$ と漸近同等な推定量列のことをいう．A more informative name for "regular" is asymptotically equivalent-in-law.[11]

例 2.2.3 (Stein shrinkage estimator)**.** Stein の縮小推定量は正則でない．

定理 2.2.4 (Hajek-LeCam の畳み込み定理)**.** $\mathcal{P}$ の正則点 $\theta\in\Theta$ において，$\nu(\theta)$ の推定量列 (T_n) は正則で，極限分布 L_θ を持つとする．$\nu:\Theta\to\mathbb{R}^m$ が $\theta\in\Theta$ において微分可能ならば，ある確率分布 M_θ が存在して

$$L_\theta = N(0,\dot{\nu}(\theta)I_\theta^{-1}\dot{\nu}(\theta)^\top) * M_\theta$$

が成り立つ．

定義 2.2.5 (asymptotically efficient)**.** 推定量列 (T_n) が正則で，その漸近共分散行列 Σ_θ が $\dot{\nu}(\theta)I_\theta^{-1}\dot{\nu}(\theta)^\top$ と表せるならば，(T_n) は**漸近有効**であるという．

2.2.2　漸近有効推定量

漸近有効推定量は影響関数のことばによって特徴付けられる．そして影響関数とは，微分というよりも，漸近的な線型関係をいう．

記法 2.2.6. $|f|^2 = f^\top(X)f(X)$ が可積分な空間 $L_2(P)$ の，平均 0 のものがなす部分空間 $L_2^0(P) := \left\{f\in L^2(P) \mid Pf = 0\right\}$ を考える．絶対値の二乗が 2 次形式として特別な意味を持つのである．

定義 2.2.7 (asymptotically linear, influence function)**.** 一般のモデル $\mathcal{P}$ と汎関数 $\psi:\mathcal{P}\to\mathbb{R}^m$ について，次が成り立つならば，**漸近線型**であるという：

$$\forall_{P\in\mathcal{P}}\ \exists_{\check{\psi}_P\in\mathcal{L}_2^0(\mathcal{X},P;\mathbb{R}^m)}\ T_n = \psi(P)+\mathbb{P}_n\check{\psi}_P + o_P(n^{-1/2}).$$

このときの漸近的な線型主要部 $\check{\psi}_P$ を T_n の**影響関数**という．

要諦 2.2.8. 漸近線型であるときは，正規近似 $\sqrt{n}(T_n-\psi(P))\to N(0,P[\check{\psi}_P\check{\psi}_P^\top])$ が成功する．また，推定誤差 $T_n-\psi(P)$ に対する観測 X_i-x の影響は，1 次近似として $n^{-1}\check{\psi}_P(x)$ で与えられる．

定理 2.2.9 (パラメトリックモデルにおける漸近有効性の特徴付け)**.** パラメトリックモデル $\mathcal{P} = (P_\theta)_{\theta\in\Theta}$ の正則点 $\theta\in\Theta$ において，$\nu(\theta)$ は θ で微分可能であるとする．次の 2 条件は同値：

(1) $\nu(\theta)$ の推定量列 (T_n) が

$$T_n = \nu(\theta)+\mathbb{P}_n\widetilde{l}_{\theta|\nu,\mathcal{P}} + o_P(n^{-1/2}),\quad \widetilde{l}_{\theta|\nu,\mathcal{P}}(x) := \dot{\nu}(\theta)I_\theta^{-1}\dot{l}_\theta(x)$$

と展開される．

(2) (T_n) は $\theta \in \Theta$ において $\nu(\theta)$ の漸近有効推定量列であり，その漸近分散は $I^{-1}_{\theta|\nu,\mathcal{P}} := \dot{\nu}(\theta) I_\theta^{-1} \dot{\nu}(\theta)^\top$ である．

定義 2.2.10 (efficient influence function, information bound, efficient information).

(1) $\widetilde{l}_{\theta|\nu,\mathcal{P}}(x) := \dot{\nu}(\theta) I_\theta^{-1} \dot{l}_\theta(x)$ を $\mathcal{P}$ の ν における**有効影響関数**という．
(2) $I^{-1}_{\theta|\nu,\mathcal{P}} := \dot{\nu}(\theta) I_\theta^{-1} \dot{\nu}(\theta)^\top$ とその逆行列を，それぞれ**情報量限界**と**有効情報量**という．

2.3　分離度

正則なパラメトリックモデルは多様体をなす．そこでまず，多様体上の分離度の一般論を調べる．分離度とは，2 点の分離性を検出する関数であって，Taylor 展開の 2 次項が Riemann 距離を定めるものである．

定義 2.3.1 (divergence). $(M^p, (\xi_x))$ を可微分多様体とする．関数 $D : M \times M \to \mathbb{R}_+$ が**分離度**であるとは，

(1) 非退化性：$D[P, Q] = 0 \Leftrightarrow P = Q$.
(2) 点 ξ と同じ座標系で表せる近傍の点 $\xi + d\xi$ について，$D[\xi, \xi + d\xi] = \frac{1}{2} g_x(dx, dx) + O(|dx|^3) = \frac{1}{2} \sum_{i,j \in [p]} g_{ij}(\xi) d\xi_i \otimes d\xi_j + O(|d\xi|^3)$ $(|d\xi| \to 0)$ が定める行列 $G(\xi) = (g_{ij}(\xi))$ は正定値対称である．[†1]
(3) 点 $P \in M$ の r-近傍 $N(P, r) := \{Q \in M \mid D[P, Q] < r\}$ は r について単調に増大する．

$D^*(p, q) := D(q, p)$ を双対分離度という．

要諦 2.3.2. 対称性も三角不等式も成り立たず，距離の概念とは違い，むしろ距離の二乗のような，分散のような概念である（次元解析の考え方）．(1) より D が半正定値であることがわかるが，これは，D は定数部分と線型部分を持たないことを意味する．定数部分は $D[P, P] = 0$ に矛盾し，線型部分は非負性に矛盾する．これは，D の二次部分が半正定値な二次形式であることを意味する．こうして，特に (2) で正定値性が成り立てば，自然に Riemann 計量を定める．このとき非対称性は受け継がれ，これが双対性という方向を持った新たな構造を授ける．

注 2.3.3. 情報理論では，非対称性を強調して $D(p||q)$ などと表す．また条件付き確率 $P(A|B)$ の記法に似ているが，これと同様に分離度も相対測度として見做す態度がある．この見方をするとき KL 分離度を相対エントロピーと呼ぶ．

定義 2.3.4 (Kullback-Leibler information / divergence / relative entropy). $P, Q \in \mathcal{P}$ について，

$$I[Q, P] := \int_{\mathcal{X}} q(x) \log \frac{q(x)}{p(x)} \nu(dx)$$

によって定まる関数 $I : \mathcal{P} \times \mathcal{P} \to \mathbb{R}_+$ を **Kullback-Leibler 情報量**という．

例 2.3.5 (正規分布族の KL 分離度). $M_n(\mathbb{R})$ の中の半正定値行列のなす部分多様体 $\mathrm{GL}_n(\mathbb{R})_+$ は $n(n+1)/2$ 次元である．この多様体に

$$D[P, Q] = \mathrm{Tr}(PQ^{-1}) - \log|PQ^{-1}| - n$$

なるダイバージェンスが定まる．これは，P, Q を分散共分散行列とする平均 0 の正規分布の間の KL ダイバージェンスになっている．

[†1] $dx \in T^*(\mathcal{P})$ と見るのである！

2.4　Riemann 計量

2.4.1　計量の定義

$T(a,b) := T(M)^{\otimes a} \otimes T^*(M)^{\otimes b}$ の元を (a,b)-テンソルバンドルという（順に共変・反変）．ベクトル場は $(1,0)$-テンソル束の切断である．

定義 2.4.1 (metric)**.** 滑らかなベクトル束 $\pi : E \to M$ を考える．

(1) 全てのファイバー E_x に内積 $(-,-)_x$ が与えられており，内積空間である．
(2) 任意の滑らかな切断面 $\sigma_1, \sigma_2 \in \Gamma(E)$ に対して，合成写像 $(\sigma_1(-), \sigma_2(-))_{(-)} : M \to K$ は滑らかである．

E が $\mathbb{R}$ 上のベクトル束ならば Riemann 計量，$\mathcal{C}$ 上のベクトル束ならば Hermite 計量という．

要諦 2.4.2. $(\sigma_1(-), \sigma_2(-))_x \in \Gamma(E) \otimes_{C(X)} \Gamma(E)$ は確かに共変 2-テンソルである．任意の 2 つのベクトル場 $\sigma_1, \sigma_2 \in \Gamma(E)$ に対して，$x \in M$ をまずベクトル束 E_x 上の 2 つのベクトル $\sigma_1(x), \sigma_2(x)$ に対応させ，そこで内積を取る対応．

定理 2.4.3. 任意のベクトル束には計量が存在する．

2.4.2　双対バンドル

滑らかなベクトル束には計量が入り，各ファイバー V_x には内積が存在する．V_x の基底 $\{e_1, \cdots, e_p\}$ に対して $g_{\mu\nu} := (e_\mu | e_\nu)$ とおくと，対称性と基底であることより $(g_{\mu\nu})$ は対称で正 $\det(g_{\mu\nu}) > 0$ な行列となる．この行列の単位行列からの乖離は正規直交系からの乖離を意味する．当然この表示を用いると

$$(a|x) = \sum_{\mu,\nu=1}^{p} g_{\mu\nu} a^\mu x^\nu$$

が成り立ち，内積が定める同型 $V_x \sim V_x^*$ は

$$a = \sum_{\mu=1}^{p} a^\mu e_\mu \mapsto \sum_{\nu=1}^{p} \left(\sum_{\mu=1}^{p} g_{\mu\nu} a^\mu \right) e^\nu = \varphi_a = (a|-)$$

と表せる．双対基底を用いると，無理やり Fourier 係数を抽出出来る．
ここで大域に目を戻すと，計量はバンドル全体に同型対応 $E \simeq E^*$ を導く．

定義 2.4.4.

(1) 内積空間 V について，同型 $V \to V^* : a \mapsto (a,-)$ が定まる．これを**内積が定める標準的な同型対応**という．
(2) 各ファイバー $E_x \simeq E_x^*$ にも標準同型が存在し，ここからベクトルバンドルとしての同型対応 $E \simeq E^*$ が引き起こされる．これを**計量が定める標準的な同型対応**という．

議論 2.4.5. バンドルとしての対応 $E \simeq E^*$ を局所的に調べる．
$U \overset{\text{open}}{\subset} M$ を座標近傍とし，E のこの上での自明性 $E|_U \simeq U \times \mathbb{R}^p$ を局所基底 $(e_1, \cdots, e_p) : U \to E$ が与えるとする．このとき，E_x $(x \in U)$ の基底 $\{e_1(x), \cdots, e_p(x)\}$ の双対基底 $\{e^1(x), \cdots, e^p(x)\}$ は，E^* の U 上での自明性 $E^*|_U \simeq U \times \mathbb{R}^p$ を与える．
ここから，E, E^* が同型であることを，対応を具体的構成することで示す．E の計量 $(-|-)_-$ に対して，$g_{\mu\nu}(x) := (e_\mu(x), e_\nu(x))_x \in \mathbb{K}$ と定める．$(g_{\mu\nu}(x)) \in \mathrm{GL}_p(\mathbb{K})$ だから，逆行列を持つが，これを $(g^{\mu\nu}(x))$ で表すこととする．次の 2 つの写像は互いに逆になっ

ている：

$$\begin{array}{ccc}
\Gamma(E|_U) & \longrightarrow & \Gamma(E^*|_U) \\
\cup & & \cup \\
\alpha(x) = \displaystyle\sum_{\mu=1}^{p} \alpha^\mu(x) e_\mu(x) & \longmapsto & \displaystyle\sum_{\mu=1}^{p} \left(\sum_{\nu=1}^{p} g_{\mu\nu}(x) \alpha^\nu(x) \right) e^\mu(x) = \beta(x)
\end{array}$$

$$\begin{array}{ccc}
\Gamma(E^*|_U) & \longrightarrow & \Gamma(E|_U) \\
\cup & & \cup \\
\beta(x) = \displaystyle\sum_{\mu=1}^{p} \beta_\mu(x) e^\mu(x) & \longmapsto & \displaystyle\sum_{\mu=1}^{p} \left(\sum_{\nu=1}^{p} g^{\mu\nu}(x) \beta_\nu(x) \right) e_{\mu(x)}
\end{array}$$

さらに，E^* に計量が押し出され，それは局所的には $g^{\mu\nu}(x) = (e^\mu(x), e^\nu(x))_x$ と表せる．

要諦 2.4.6. よって，Fisher 情報行列とその逆とは双対的な関係にある．

2.4.3　Riemann 計量

Riemann 計量とは，接束という特別具体的なベクトル束に与えられた計量をいう．こう見ると，これは $(0,2)$-テンソル場 $g \in \Gamma(T^*(M) \otimes T^*(M))$ である．局所的には正定値な対称行列であり，$(0,2)$-テンソル場としての変換則を満たす大域的な存在が Riemann 計量である．

要諦 2.4.7. 接バンドル $T(M)$ に計量が与えられたとき，M に Riemann 計量が与えられたという．計量の定義 2.4.1 を局所的な「各点ごとに見ると内積」とは違って大域的に定義すると，滑らかな $(0,2)$-テンソル場 $g \in \Gamma(T^*(M) \otimes T^*(M))$ であって，(Hermite) 対称かつ正定値なものをいう．すなわち，$g_x : T_x(M) \times T_x(M) \to \mathbb{R}$ を満たす双線型写像で，$g_x(X,Y) = g_x(Y,X)$ かつ $g_x(X,X) > 0\ (X \neq 0)$ として扱える．$g : M \to T^*(M) \otimes T^*(M)$ は，各 $(U, (x^1, \cdots, x^n))$ 上での C^∞ 関数 $g_{ij} := \left(\frac{\partial}{\partial x^i}, \frac{\partial}{\partial x^j} \right)$ の貼り合わせである．これに対して内積は $\langle X, Y \rangle_x := g_x(X,Y)$ で復元できる．

以上のことを逆から議論すると次のようになる．

記法 2.4.8. 各接空間 T_x の内積について，

$$g^\alpha_{\mu\nu}(x) := \left(\frac{\partial}{\partial x^\mu_\alpha}, \frac{\partial}{\partial x^\nu_\alpha} \right)_x$$

とおくと，$g^\alpha_{\mu\nu} : U_\alpha \to \mathbb{R}$ は滑らかな関数であり，各点で $\det(g^\alpha_{\mu\nu}) > 0$ を満たす．（ベクトル場の成分の 2 次元版）．さらに，内積は，ξ, η を数ベクトル $(\xi^\mu), (\eta^\nu)$ と Euclid 空間での内積 $\langle -, - \rangle$ を用いて

$$(\xi, \eta) = \langle (\xi^\mu), G(\eta^\nu) \rangle$$

と表せる．

$x \in U_\alpha \cap U_\beta$ のとき，$(g^\alpha_{\mu\nu}(x))$ と $(g^\beta_{\mu\nu}(x))$ との間の変換則は，

$$g^\beta_{\mu\nu}(x) = \sum_{i,j=1}^{n} \frac{\partial x^i_\alpha}{\partial x^\mu_\beta} \frac{\partial x^j_\alpha}{\partial x^\nu_\beta} g^\alpha_{ij}(x)$$

すなわち $(g^\beta_{\mu\nu}) = \left(\frac{\partial x^i_\alpha}{\partial x^\mu_\beta} \right)^\top (g^\alpha_{ij}) \left(\frac{\partial x^j_\alpha}{\partial x^\nu_\beta} \right)$ で与えられる．ベクトル場の変換則は $(\xi^\mu_\beta) = \left(\frac{\partial x_\beta}{\partial x_\alpha} \right) (\xi^\nu_\alpha)$ で与えられることに注意．たしかに 2 階の反変テンソルである．

逆に，各局所座標 U_α 上で定義された正値行列 $(g^\alpha_{\mu\nu}(x))$ で，$U_\alpha \cap U_\beta$ 上で上述の $(2,0)$-テンソルの貼り合わせの条件を満たすものを与えることで，計量が定まる．これを

$$ds^2 = \sum g^\alpha_{\mu\nu}(x) dx^\mu_\alpha dx^\nu_\alpha$$

と表す．

要諦 2.4.9. こうして，貼り合わせの考え方からみると，Riemann 計量とは局所的には各 U_α 上の正定値対称行列 $(g^\alpha_{\mu\nu}(x))$ で，$(2,0)$-テンソルとしての貼り合わせの変換則を満たすもの，と捉えられる（そのようなものを定めると Riemann 計量が一意に定まる）．テンソルと言ってもビビらなくて良い，大域的に存在するための整合性を満たすというだけの消息である．

2.5　接続

2.5.1　微分作用素の定義と例

微分作用素とは，ベクトル束の切断の間の線型写像であって，$\sigma \in \Gamma(E)$ がある開集合 U 上で恒等的に 0 ならば，微分しても 0 であるという性質を抽出して，「台を縮小させる切断間の線型作用素」と定義する．

定義 2.5.1 (differential operator). E, F を M 上のベクトル束とする．線型写像 $\Phi : \Gamma(E) \to \Gamma(F)$ が，任意の $\sigma \in \Gamma(E)$ に対して $\operatorname{supp} \Phi(\sigma) \subset \operatorname{supp} \sigma$ を満たすとき，これを**微分作用素**という．

補題 2.5.2.

(1) 微分作用素の全体は線型空間をなす．これを $D(E, F) \subset \mathrm{Hom}_{\mathbb{R}}(E, F)$ で表す．
(2) $C(M)$-加群の構造も持つ．

例 2.5.3.

(1) 多重指数 $\alpha \in \mathbb{N}^n$ について，D^α は微分作用素である．これらの線型結合も微分作用素である．
(2) 局所有限な開被覆 $(V_i)_{i\in\mathbb{N}}$ について，これに属する 1 の分解 (φ_i) を用いて，D^α の線型結合で表される微分作用素の局所有限和 $\Phi := \sum_{i=1}^\infty \varphi_i \Phi_i$ も微分作用素である．この形の微分作用素は一般に階数を持つとは限らない．Peetre の定理より，$D(\epsilon^1, \epsilon^1)$ の元はすべてこの形で表される．

2.5.2　Peetre の定理

Φ の局所座標への制限 $\Phi|_U : C_c(U, \mathbb{R}^p) \to C_c(U, \mathbb{R}^p)$ を行列表示したこととなる．

定理 2.5.4 (微分作用素の局所表示). $\Phi \in D(E, F)$ を微分作用素とする．M の座標近傍 $\{U, (x^1, \cdots, x^n)\}$ について，$\overline{U}$ はコンパクト，$E|_U, F|_U$ は自明束となっているとする．$\sigma \in \Gamma(E)$ は $\operatorname{supp} \sigma \subset U$ を満たし，$\sigma(x) = (\sigma^1(x), \cdots, \sigma^p(x))$ $(x \in U)$ と表せるとし，$\Phi(\sigma)(x) = (\tau^1(x), \cdots, \tau^q(x))$ $(x \in U)$ と表す．このとき，次の関係が成り立つ：

$$\begin{pmatrix} \tau^1(x) \\ \vdots \\ \tau^q(x) \end{pmatrix} = (\Phi^i_j(x)) \begin{pmatrix} \sigma^1(x) \\ \vdots \\ \sigma^p(x) \end{pmatrix}.$$

ただし，

$$\Phi^i_j(x) = \sum_{|\alpha|\le k} a^i_{j\alpha}(x) D^\alpha \quad (\alpha \in \mathbb{N}^n, a^i_{j\alpha} \in C^\infty(U)).$$

定義 2.5.5. $\sigma \in C(U, \mathbb{R}^p)$ に対して，α が点 $x_0 \in U$ で $D^\alpha \sigma(x_0) = 0$ $(|\alpha| \le k)$ を満たすとき，σ は x_0 で **k-平旦**であるという．

2.5.3　微分作用素の階数

定義 2.5.6. 微分作用素 Φ が局所的に常に

$$\Phi^i_j(x) = \sum_{|\alpha|\le k} a^i_{j\alpha}(x) D^\alpha \quad (\alpha \in \mathbb{N}^n, a^i_{j\alpha} \in C^\infty(U))$$

と表せるとき，これを **k-階**という．k-階微分作用素のなす部分空間を $D_k(E, F)$ で表す．

系 2.5.7. $\Phi \in D(E,F)$ について，

(1) k-階である．

(2) $\forall_{x\in M}\, j^k(\sigma)(x) = 0 \Rightarrow \Phi(\sigma)(x) = 0$．すなわち，局所座標を用いて表せば，$D^\alpha\sigma(x) = 0\ (|\alpha| \le k)$．

定理 2.5.8.

(1) $D_0(E,F) \subset D_1(E,F) \subset \cdots$ が成り立つ．

(2) $\cup_{k\in\mathbb{N}} D_k(E,F) = D(E,F)$ が成り立つことと，M がコンパクトであることとは同値．

2.6　局外母数と情報量の幾何

記法 2.6.1. 正則なパラメトリックモデル $\mathcal{P} = (P_\theta)_{\theta\in\Theta=N\times H}$ について，

(1) $\dot{l} := \dot{l}_{\theta_0} = \left.\dfrac{d}{d\theta}\log p_\theta(x)\right|_{\theta=\theta_0} = \dfrac{\dot{p}_{\theta_0}(x)}{p_{\theta_0}(x)} 1_{\{p_\theta\ne 0\}}(x)$ をスコア関数とする．

(2) $I := I_{\theta_0} = P_{\theta_0}[\dot{l}\dot{l}^\top] = E_{\theta_0}[|\dot{l}|^2]$ はスコアの 2 次のモーメントであるが，これを Fisher 情報行列という．

(3) $\widetilde{l} := \widetilde{l}_{\theta_0} = I_{\theta_0}^{-1}\dot{l}_{\theta_0}$ を有効影響関数とする．

第 3 章

セミパラメトリックモデル情報限界

一部の特徴はパラメータ付けすることが可能なほど知見があり，一方で残りは生の関数のまんま残しておくモデルをセミパラメトリックモデルという．このようなモデルは複雑な構造ゆえ，（解析的な解決は無理で）アルゴリズム的な実装のみが可能な，非線形な推定過程を必要とする．そこで，**最も重要な手続きは漸近理論**となり，実際のサンプルサイズに対するパフォーマンスは，シミュレーションと数値実験によって検証されることとなる．

観察研究などで，欠測データへの配慮を行わないこと（"complete case analysis"と呼ばれて，解析ソフトはデフォルトでこの挙動を行うものも多い）は深刻なバイアスをもたらしかねない．欠測データと言っても，"censored data"を含む．これは生存解析でよく直面する．[9]

一致で漸近正規な統計量を semiparametric estimator という．これの存在や探し方は，影響関数の幾何学から重大な示唆がある．

(1) 最尤推定法は，共変量 x と結果変数 y の間のパラメトリックなモデルを立て，その中では「最適」かもしれない．
(2) 傾向スコア・マッチングでは，傾向スコア $e(x) = \mathrm{pr}(z = 1|x)$ の推定が大事になる．すなわち，共変量 x と割当 z の間の回帰関係をパラメトリックに解析するが，その後の結果変数と共変量の間には回帰モデルを仮定する必要がない．
(3) 割当変数 z は大抵一次元で，モデルの誤設定の可能性は低いので，ロバストであると言える．[†1]が，もっと祈りを高く，二重にロバストであることを考えたい．

3.1 例

ほとんどの場合，無限次元空間は関数空間であり，有限次元空間は何かしらの特徴量である．

3.1.1 Restricted Moment Models

$\mu \in C(\Theta_1 \times \Theta_2), \beta \in \Theta_2 \subset \mathbb{R}^q$ に関して，$E(Y|X) = \mu(X, \beta)$ という回帰関係を仮定する．これ以上の仮定を置かない場合，これはセミパラメトリックモデルで，Chamberlain and Newey によって計量経済学から研究され，Liang and Zeger により広められた．

3.1.2 Proportional Hazards Model

We assume the treatment effect can be modeled multiplicatively (parametrically) on some completely unknown scale. Such models have arisen in a wide variety of contexts in recent years, particularly in economics, epidemiology, and astronomy.[6]

Cox により 72 年に導入された生存解析におけるモデル．

[†1] Drake 1993

$$\lambda(t|X) = \lim_{h\to 0}\left(\frac{P(t \le T < t+h | T \ge t, X)}{h}\right) \\ = \lambda(t)\exp(\beta^T X).$$

3.1.3　Nonparametric Model

セミパラメトリックモデルの解析に於ても，$\beta(\theta)$ なる汎関数を推定する際などは，θ の分割は特に数理的な示唆を与えない．

3.2　Hilbert 空間

例 3.2.1 (space of mean-zero and bounded second moments random functions)**.** 標本空間 (X, A, P) 上の q 次元確率関数 $h: X \to \mathbb{R}^q$ であって，$E[h(Z)] = 0, E[h^T(Z)h(Z)] < \infty$ を満たすもの全体は Hilbert 空間 H_{mb} をなす．なお，確率関数とは，可測関数 $h \circ Z$ を表す．内積 $(h_1|h_2) := E[h_1^T h_2]$ を共分散内積と呼ぶ．

3.3　接集合と情報量

記法 3.3.1. 観測量を $X_1, \cdots$，その標本空間を $\mathcal{X}$ とする．

3.3.1　スコア関数と接集合

LeCun による概念.

記法 3.3.2.

$$\mathcal{L}^2_*(\mathcal{X}, P) := \left\{g \in \mathcal{L}^2(\mathcal{X}, P) \mid E_P[g] = 0\right\}.$$

定義 3.3.3 (P-path)**.** $\mathbb{P} \in \mathcal{P}$ に対して，$\mathbb{P}$-道とは，$(P_t)_{t\in\mathbb{R}_+} : [0, \epsilon) \to \mathcal{P}; P_0 = \mathbb{P}$ をいう．

定義 3.3.4 (differentiability of path, score function)**.** 微分可能な P-道とは，P-道 $(P_t)_{t\in[0,\epsilon)}$ であって，

(1) ある可測関数 $g \in \mathcal{L}(\mathcal{X}, P)$ に対して

$$\int \left(\frac{dP_t^{1/2} - dP^{1/2}}{t} - \frac{1}{2} g dP^{1/2}\right)^2 \xrightarrow{t\to 0} 0.$$

を満たすことをいう．このとき，$dP^{1/2}$ は測度が定める線型作用素の二乗根作用素とする．すると，g は中心化された P-2 乗可積分関数 $g \in \mathcal{L}^2_*(\mathcal{X}, P)$ となることが必要．[13]

(2) 参照測度の族 (μ_t) に対して $\dfrac{dP}{d\mu_t} = p_{tt}$ とし，

$$\int \left(\frac{p_{tt}^2 - p_t^2}{t} - \frac{1}{2} g p_t^{1/2}\right) d\mu_t \xrightarrow{t\to 0} 0.$$

と理解しても良い．

このときの g を**スコア関数**や**道毎の微分**という．

定義 3.3.5 (tangent set / space)**.** P を通る全ての可微分な道 $(P_t)_{t\in\mathbb{R}_+}$ に関するスコア関数 g 全体の集合を，P における**接集合**といい，$\dot{\mathcal{P}}_P$[13] または $T_P(\mathcal{P})$[7] で表す．ほとんどの場合，これは閉部分空間となる．

補題 3.3.6. 任意のスコア関数は 2 乗可積分で，$E_P[g] = 0$ を満たす [13]．特に，$\dot{\mathcal{P}}_P \subset \mathcal{L}^2(\mathcal{X}, P)$ が成り立つ．

3.3.2　可微分性

定義 3.3.7 (differentiable, efficient influence function).

(1) 汎関数 $\psi:\mathcal{P}\to\mathbb{R}$ が $P\in\mathcal{P}$ において接空間 $\dot{\mathcal{P}}_P$ について**可微分**であるとは，

$$\exists_{\dot{\psi}_P\in L^2(P)^*}\forall_{g\in\dot{\mathcal{P}}_P}\ \frac{\psi(P_t)-\psi(P)}{t}\xrightarrow{t\to 0}\dot{\psi}_P(g)$$

を満たすことをいう．

(2) このときの有界線型汎函数 $\dot{\psi}_P:L^2(P)\to\mathbb{R}$ は，$\dot{\mathcal{P}}_P$ の生成する閉部分空間上では一意に定まる．これを**有効影響関数**という．$\dot{\mathcal{P}}_P$ 上への射影として有効影響関数を定めるような，有界線型汎函数 $\dot{\psi}_P:L^2(P)\to\mathbb{R}$ を**影響関数**という．

3.4　傾向スコア・マッチング

Rosenbaum and Rubin 1983．強く無視できる割当条件を満たすときの，平均処置効果の不偏推定量の標準的構成法である．また，多次元な共変量を 1 つのスカラー変数に集約することができれば，その 1 変数の上で層別化などを行うことができるため，従来のマッチングや層別での問題が回避される．なおこれは，まず傾向スコアを推定する必要が生じるため，二段階推定法となる．が，「推定した傾向スコアを用いても，推定の偏りは真の傾向スコアを用いた場合と同様である」ということが明らかになっている (Drake, 1993).[a]

[a] If treatment assignment is strongly ignorable given x, then the difference between treatment and control means at each value, and consequently pair matching, subclassification and covariance adjustment on a balancing score can produce unbiased estimates of the avarage treatment effect.

3.4.1　強く無視できる割当条件

傾向スコア $e(x)$ を事前に定めている実験が無作為割当実験である．一方で，割当がランダムでない場合は，まず傾向スコア $e(x)$ の推定が問題になる．[†2]

定義 3.4.1. 割当が無作為に近いことを，**強く無視できる**という．無作為割当では，結果変数は，x の下では割当と独立である：$(r_1,r_0)\perp\!\!\!\perp z|x$．また，$\mathrm{pr}(z_i|x)\in(0,1)$ である．割当が無作為ではないながらも，次の条件が保証されることをいう：

$$(r_1,r_0)\perp\!\!\!\perp z|v,\quad 0<\mathrm{pr}(z=1|v)<1$$

補題 3.4.2. 割当が強く無視できるとき，Rubin 1978 の意味でも無視できる（すなわち，「無視できる欠測」である）．が，逆は成り立たない．

3.4.2　定義と特徴付け

バランシングスコアは，共変量の値域の空間に定める同値類の細かさの順序について有界束をなし，最大元が $b(x)=x$ で，最小元が $b(x)=e(x)$ である．

定義 3.4.3 (balancing score, propensity score). 共変量 x の関数を考える．

(1) バランシングスコア $b(x)$ とは，これによって共変量 x を条件付けると，割付 z と独立になるような関数を言う：$z\perp\!\!\!\perp x|b(x)$.[†3] すなわち，$b(x)$ を与えたときの x の条件付き分布が，$z=0,1$ に依らず同一になるような，共変量の対応付けを言う．

[†2] logit モデルなどのパラメトリックモデルが使われることが多い．
[†3] David's (1979) notation

(2) $e(x) := \mathrm{pr}(z = 1|x)$ によって定まる共変量の関数を，**傾向スコア**という．これは，共変量 x を仮定したときに，処置を受ける傾向を表している．RCT では $e = 1/2$ という定数関数となる．傾向スコアは，バランシングスコア $b(x)$ の中で，最も粗い同値関係を定めるものとして特徴付けられる．

要諦 3.4.4. バランシングスコアは共変量の関数と定義されているが，これは共変量の値域に同値類を作るための装置に他ならない．傾向スコアとは，そのようにして関数概念を運用するという巧妙な形式科学である．

例 3.4.5. $b(x) = x$ はバランシングスコアである．この $b(x)$ で条件づけたとき，x の分布はデルタ分布になる．特に，バランシングスコアは存在する．

補題 3.4.6 (バランシングスコアと傾向スコアの特徴付け)**.** b を共変量の関数とする．次の 2 条件は同値.

(1) b はバランシングスコアである．
(2) $b(x)$ は $e(x)$ より細かい同値類を定める：関数 f が存在して，$e(x) = f(z(x))$ と表せる．

系 3.4.7 (Cochran and Rubin (1973))**.** 特に，割当と共変量とは，傾向スコアが与えられたとき，条件付き独立である：$x \perp\!\!\!\perp z | e(x)$.

3.4.3　平均処置効果の推定

割当が強く無視できるとき，$b(x)$ が取る各値において，平均処置効果は，それぞれの群の平均の差によって不偏推定可能である．

定理 3.4.8. 割当が x について強く無視できるとする．このとき，割当は任意のバランシングスコア $b(x)$ についても強く無視できる：

$$[\forall_x\ (r_1, r_0) \perp\!\!\!\perp z | x \wedge \mathrm{pr}(z = 1|x) \in (0, 1)] \Longrightarrow [\forall_b\ \forall_x\ (r_1, r_0) \perp\!\!\!\perp z | b(x) \wedge \mathrm{pr}(z = 1|b(x)) \in (0, 1)].$$

定理 3.4.9. 割当が x について強く無視でき，$b(x)$ をバランシングスコアとする．このとき，

$$E[r_1|b(x), z = 1] - E[r_0|b(x), z = 0] = E[r_1 - r_0|b(x)].$$

系 3.4.10 (pair matching on balancing scores)**.** 割当は強く無視できるとする．また，母集団から，$b(x)$ の値と，それに対応する単位であって，それぞれの群の中から $z_i = 1$ を満たす単位 i と $z_j = 0$ を満たす単位 j が抽出されたとする．このとき，$E[r_1 - r_0|b(x)] = E[y_i - y_j]$ が成り立つ．

系 3.4.11 (subclassification(層別解析))**.**

系 3.4.12 (covariance adjustment(共分散分析))**.**

注 3.4.13. 共分散分析は分散分析 (ANOVA) と回帰分析を組み合わせた，一般化線形モデルである．

3.4.4　傾向スコアの推定

傾向スコアの経験測度から構成される推定量から，x 上の sample balance を構成できる．

定義 3.4.14. P_N を標本から得た確率測度とする．$\widehat{e}(a) := P_N(z = 1|x = a)$ によって，傾向スコア $e(x)$ を推定することを考える．

定理 3.4.15. $\widehat{e}(a) \in (0, 1)$ ならば，$P_N(z = 0, x = a|\widehat{e}(x) = \widehat{e}(a)) \Rightarrow P_N(z = 0|\widehat{e}(x) = \widehat{e}(x))P_N(x = a|\widehat{e}(z) = \widehat{e}(a))$.

3.4.5　IPW 推定法

Rubin 1985 による，層別標本抽出における Horovitz and Thompson 1952 の拡張.

定義 3.4.16 (IPW: inverse probability weighting). y_0, y_1 の周辺平均の推定を考える.
y の値の，なんらかの確率測度についての期待値ではなく，傾向スコアの逆数が定める積分を考える：

$$\widehat{E}(y_1) := \frac{\sum \frac{z_i}{e_i} y_i}{\sum \frac{z_i}{e_i}}, \quad \widehat{E}(y_0) := \frac{\sum \frac{1-z_i}{1-e_i} y_i}{\sum \frac{1-z_i}{1-e_i}}$$

強く無視できる割当条件が成立するとき，真の傾向スコアがわかっているなら不偏推定量となり，傾向スコアを推定値で代用しても一致推定量となる.

注 3.4.17. IPW 推定量は，母数ベクトル $(E(y_1), E(y_0))$，推定関数 $\left(\frac{z}{e}(y - E[y_1]), \frac{1-z}{1-e}(y - E[y_0])\right)$ についての m-推定量でもある.

3.4.6　kernel マッチング

3.5　傾向スコア解析の拡張

3.5.1　一般化推定方程式における IPW 推定量

Liang and Zeger 1986

3.5.2　一般化傾向スコア

Imbens 2000 は，割当変数が 2 値でない場合にも傾向スコアが利用できるように拡張した.

3.6　一般的な周辺パラメトリックモデルの推定

第 4 章

汎関数推定

まだ名前が定まっていないが，Semiparametric Doubly Robust Targeted Double Machine Learning と呼ばれている．

おそらく，こういう話である．ノンパラメトリックなパラメータとは汎関数になる (これを targeted parameter ともいう)．汎関数の推定は，モデル $\mathbb{P} \in \mathcal{P}$ の密度自体の推定より簡単で，さらに汎関数に，影響関数の手法が使える良い構造を仮定すると，fast $\sqrt{n}$-収束が実現できる．

記法 4.0.1.

(1) $A = a$ とした場合の (反実仮想的) 観測値を Y^a で表す．

(2) $Z \in \mathbb{R}^m$ は観測量とする．

4.1　汎関数の例

本質的に，共変量上で平均を取った回帰関数の形を取る．回帰関数とは何の仮定も置かない場合はただの条件付き期待値であるから，条件付き期待値の平均，という形が標準的である．このような設定に対する意味論は，因果推論とより一般に欠損データである．

4.1.1　セミパラメトリック回帰について

例 4.1.1 (回帰分析とは，条件付き期待値なる汎関数の推定問題である)**.** $Z = (X, Y) \in \mathbb{R}^2$ が観測されたとする．回帰関数 $\psi(x; P) = E[Y|X = x]$ は汎関数 $\mathbb{R} \times \mathcal{P} \to \mathbb{R}$ となる．

そもそも，任意の観測量 $Z = (X, Y)$ に対して，$Y_i = E[Y_i|X_i] + \epsilon_i, E[\epsilon_i|X_i] = 0$ が成り立つ．これは X_i に依存する成分が，条件付き期待値によって可予測に表されていることによる．したがって，二乗誤差について，$E[Y_i|X_i]$ は X_i から Y_i を推定する最適推定量である．$E[Y_i|X_i]$ の形を指定するのが回帰モデルである．

4.1.2　処置効果

例 4.1.2 (平均処置効果)**.** $Z = (X, A, Y)$ とする．

positivity　処置割当がなされる：$0 < P[A = a] < 1$.

consistency

no unmeasured confounding

母集団がすべて $A = a$ の処置を受けたときの結果の期待値

$$\psi(P) := E[Y^a] = E[E[Y|X, A = a]] = E[\mu_a(X)]$$

は汎関数となる．これは，回帰関数の，X(の周辺分布) に関する平均値，という形をしている．回帰関数を $\mu_a(X) := E[Y|X, A = a]$ とすることも多い．

これに対して，平均処置効果は $E[Y^1 - Y^0]$ と定まる．

例 4.1.3 (平均欠損結果)**.** $Z = (X, A, AY)$ とする．$A = 0$ のとき AY とは欠損とする．

positivity

missing at random

母集団の平均結果は，回帰関数の X に関する平均

$$\psi(P) := E[Y] = E[E[Y|X, A = 1]]$$

となる．これは，数学的には平均処置結果の特別な場合となる．

例 4.1.4 (平均で重みづけた処置効果)**.** $Z = (X, A, Y)$ とする．平均処置効果と同様の仮定の下で，

$$\psi(P) = E[w(X)E[Y^1 - Y^0|X]] = \frac{E[\mathrm{Cov}[A, Y|X]]}{E[\mathrm{Var}[A|X]]} \qquad w(X) := \frac{\mathrm{Var}[A|X]}{E[\mathrm{Var}[A|X]]}$$

例 4.1.5 (確率的介入効果)**.** $Z = (X, A, Y)$ とする．平均処置効果と同様の仮定の下で，$A \sim dG(a|x)$ なる確率分布で割り当てられているとき，平均処置結果は

$$\psi(P) = E[E[Y|X, A^*]] = \iint E[Y|X = x, A = a]dG(a|x)dP(x).$$

例 4.1.6 (操作変数効果 / 局所平均処置効果)**.** $Z = (X, R, A, Y)$ とする．

positivity

consistency

IV-unconfoundedness

exclusion

instrumentation

monotonicity

コンプライアー $A^{r=1} > A^{r=0}$ の平均処置効果は

$$\psi(P) = E[Y^{a=1} - Y^{a=0}|A^{r=1} > A^{r=0}] = \frac{E[E[Y|X, R = 1]] - E[Y|X, R = 0]}{E[E[A|X, R = 1]] - E[A|X, R = 0]}$$

monotonicity の仮定を effect homogeneity に置き換えることで，同じ汎関数は，処置された集団に対する効果となる．

例 4.1.7 (時間変化する処置効果)**.**

4.1.3　その他

例 4.1.8 (密度の期待値)**.** Z が真の密度 p を持つとする．密度の期待値は汎関数になる：

$$\psi(P) := E[p(Z)] = \int p(z)^2 dz.$$

例 4.1.9 (エントロピー)**.** Z が真の密度 p を持つとする．エントロピーは汎関数になる：

$$\psi(P) = -\int p(z)\log p(z)dz = -E[\log p(Z)]$$

例 4.1.10 (f-分離度)**.** $Z = (A, Y)$ は，割当 $A \in 2$ と条件付き密度 $p(y|a)$ を持つとする．このとき，$p(y|a = 1), p(y|a = 0)$ の分離度は汎関数となる：

$$\psi(P) = \int f\left(\frac{p(y|a = 1)}{p(y|a = 0)}\right) p(y|a = 0)dy$$

4.2　ノンパラメトリック効率限界

target parameter ψ とモデル $\mathcal{P}$ を設定して後，下界を得ることで，効率性についてのベンチマークを考える．ちょうど，滑らかなパラメトリックモデルに対する Cramer-Rao の不等式のような，情報理論的議論である．このとき，どのような効率限界が，どのようなモデルに対して成り立つかは，一つ課題である．

記法 4.2.1. モデル $\mathcal{P}$ を，標本空間 $(\mathbb{R}^n,\mathcal{B}(\mathbb{R}^n))$ 上のすべての確率測度とする．

4.2.1　パラメトリックモデルの扱い

4.2.1.1　Cramer-Rao の不等式

滑らかなパラメトリックモデル $\mathcal{P}=(P_\theta)_{\theta\in\mathbb{R}}$ について，

定理 4.2.2 (Cramer-Rao の不等式)**.** 滑らかなパラメトリックモデル $\mathcal{P}=(P_\theta)_{\theta\in\mathbb{R}}$ と滑らかな汎関数 $\psi:\mathcal{P}\to\mathbb{R}$ について，任意の不偏推定量 $\widehat{\psi}$ は次を満たす：

$$\forall_{\theta\in\Theta}\quad \mathrm{Var}_\theta[\widehat{\psi}]\ge\frac{\psi'(\theta)^2}{\mathrm{Var}_\theta[s_\theta(Z)]}\qquad s_\theta(z)=\frac{\partial}{\partial\theta}\log p_\theta(z).$$

4.2.1.2　minimax 危険関数の下界

より広いクラスについては，次のように，危険関数に言及することによるベンチマークの与え方がある．

定義 4.2.3 (Holder class)**.** $T\subset\mathbb{R}$ を区間，$\beta,L>0$ とする．T 上の $\Sigma(\beta,L)$-Holder クラスとは，$l:=\lfloor\beta\rfloor$ に対して，

$$\Sigma(\beta,L):=\left\{f\in C^l(T)\mid f^{(l)}\text{は}\beta-l\text{ 次 Holder 連続}\right\}$$

定理 4.2.4 (Theorem 2.8 [12])**.** $\mathcal{P}$ が $\Sigma(\beta,L)$-Holder クラスであるとき，推定量 $\widehat{\psi}$ のミニマックス危険関数の下界は

$$\inf_{\widehat{\psi}}\sup_{P\in\mathcal{P}}E_P\left[(\widehat{\psi}-\psi(P))^2\right]\ge Cn^{-\frac{1}{1+d/2s}}$$

で与えられる．

4.2.1.3　滑らかなパラメトリックモデルの危険関数下界

Cramer-Rao のように，滑らかなパラメトリックモデルに対しては，より繊細な下界を危険関数に対して取れる．

定理 4.2.5 (Theorem 8.11 [11])**.** パラメトリックモデル P_θ は θ について L_2-微分可能で，非特異な Fisher 情報行列 $I_\theta=\mathrm{Var}_\theta[s_\theta(Z)]$ を持つとする．$\psi(\theta)$ は θ 上微分可能で微分係数 $\psi'(\theta)$ を持つとき，任意の推定量 $\widehat{\psi}$ は次を満たす：

$$\inf_{\delta>0}\liminf_{n\to\infty}\sup_{\|\theta'-\theta\|<\delta}nE_{\theta'}\left[(\widehat{\psi}-\psi(\theta'))^2\right]\ge\psi'(\theta)\mathrm{Var}_\theta[s_\theta(Z)]^{-1}\psi(\theta)^\top.$$

4.2.2　部分パラメトリックモデル

Stein[a]による．とは言っても，1 次元のものしか考えない場合は，真の分布 $\mathbb{P}$ を $\epsilon=0$ にて通る道に他ならない．

[a] C. Stein. Efficient nonparametric testing and estimation. Proceedings of the Third Berkeley Symposium on Mathematical Statistics and Probability, 1:187 - 195, 1956.

定義 4.2.6 (parametric submodel)**.** モデル $\mathcal{P}$ 内の部分パラメトリックモデルとは，滑らかな写像 $(P_\epsilon)_{\epsilon\in\mathbb{R}}:\mathbb{R}\to\mathcal{P}$ であって $P_0=\mathbb{P}$ を満たすものをいう．

例 4.2.7 (効率限界を与える 1 次元部分モデル). 任意の中心化された確率変数 $h:\mathcal{Z}\to\mathbb{R}$ に対して，$\exists_{M\in\mathbb{R}}\ \|h\|_\infty \le M \land \exists_{\epsilon<1/M}\ p_\epsilon(z)\ge 0$ が成り立つとき，

$$p_\epsilon(z) := dP(z)(1+\epsilon h(z))$$

とおくと，これは 1 次元の部分パラメトリックモデルであり，次の形の影響関数を持つ：

$$\frac{\partial}{\partial\epsilon}\log p_\epsilon(z)|_{\epsilon=0} = \frac{\partial}{\partial\epsilon}\log(1+\epsilon h(z))|_{\epsilon=0} = h(z).$$

したがって，この部分モデルの効率限界は，任意の $\epsilon\in\mathbb{R}$ について，

$$\frac{\psi'(P_\epsilon)^2}{\mathrm{Var}_{P_\epsilon}[s_\epsilon(Z)]} = \frac{\frac{\partial}{\partial\epsilon}\psi(P_\epsilon)|_{\epsilon=0}}{E_{P_\epsilon}[h(Z)^2]}$$

で与えられる．

4.2.3　道毎の微分可能性

まずは分母の $\dfrac{\partial}{\partial\epsilon}\psi(P_\epsilon)$ について考えるために，Frechet 多様体上の関数 $\psi:\mathcal{P}\to\mathbb{R}$ が微分可能なクラスについて，その振る舞いを考える．基本的に $\mathcal{P}$ を無限次元 Frechet 多様体と見て，path smoothness と呼んでいるのであろう．

定義 4.2.8 (smooth / differentiable / asymptotic linear, pathwise differentiable).

(1) 関数 $\psi:\mathcal{P}\to\mathbb{R}$ が**可微分**であるとは，任意の $\overline{P}\in\mathcal{P}$ に対して，対応する $\varphi(z;\overline{P})\in\mathcal{L}^2(\mathcal{Z}), R_2(P,\overline{P})\in\mathbb{R}$ が存在して

$$\psi(\overline{P})-\psi(P) = \int\varphi(z;\overline{P})d(\overline{P}-P)(z)+R_2(P,\overline{P}),\qquad E_P[\varphi(z;P)]=0, \mathrm{Var}_P[\varphi(z;P)]<\infty.$$

と表せることをいう．

(2) 関数 $\psi:\mathcal{P}\to\mathbb{R}$ が**道毎に微分可能**であるとは，任意の道 $\mathbb{R}\to\mathcal{P}, P_0=\mathbb{P}$ について，

$$\frac{\partial}{\partial\epsilon}\psi(P_\epsilon)|_{\epsilon=0} = \int\varphi(z;\mathbb{P})s_\epsilon(z)d\mathbb{P}(z)$$

を満たすことをいう．

注 4.2.9 (influence curve). この式を von Mises 展開といい，このときの関数 φ を影響関数，道毎の微分，勾配，Neyman 直交スコアなどと呼ぶ．なお，母数についての影響関数とは別の概念であるから，この場合は**影響曲線**と呼び分けることとする．

Chernozhukov が Neyman 直交性と呼んでいるのは，この道毎の微分可能性に同値．

例 4.2.10 (道毎に微分可能な汎関数).

(1)

補題 4.2.11. 可微分ならば，道毎に微分可能である．

$\mathcal{P}$ は $\mathcal{Z}=\mathcal{X}\times\mathcal{Y}$ 上の確率測度の空間で，従って各 P は $L^1(\mathcal{Z})$ 上の汎関数であり，この汎関数 $P:L^1(\mathcal{Z})\to\mathbb{R}$ の"微分"$dP:T_P(L^1(\mathcal{Z}))=:\dot{\mathcal{P}}_P\to\mathbb{R}$ が接空間上の有界線型汎函数として定まる．接空間は $L^2(P)$ の部分空間として Hilbert 空間の構造も持つから，この微分 dP は積分核を持つ：$\exists_{\varphi\in\mathcal{L}^2(P)}\ dP(g)=(g|\varphi)_P=\displaystyle\int g\varphi dP$. この積分核 $\varphi(z;P)$ を影響関数という．だから幾何学的なのか．そして各接空間の内積構造も少しずつ違う，Riemann 多様体のようになっている．$T_P(L^1(\mathcal{Z}))$ は，各移動可能な方向 $\overline{P}\in\mathcal{O}(P)$ によって張られる無限次元空間である．これに直接言及は難しいので，現在のセミパラメトリック理論では接集合という概念が登場する．

4.2.4　下界

汎関数 $\psi : \mathcal{P} \to \mathbb{R}$ に対して，$\mathbb{P}$ での微分の表現元 φ のノルムの二乗が効率限界である．

議論 4.2.12. 部分モデルの族

$$p_\epsilon(z) := dP(z)(1 + \epsilon h(z)) \qquad \exists_{M\in\mathbb{R}}\ \|h\|_\infty \le M \wedge \exists_{\epsilon<1/M}\ p_\epsilon(z) \ge 0$$

のスコアは $h(z)$ 自身だから，下界は Cauchy-Schwartz の不等式によって

$$\sup_{P_\epsilon} \frac{\psi'(P)^2}{\mathrm{Var}[s_\epsilon(Z)]} = \sup_h \frac{\mathbb{E}[\varphi Z; \mathbb{P}h(Z)]^2}{E[h(Z)^2]} \le \mathbb{E}[\varphi(Z;\mathbb{P})^2] = \mathrm{Var}[\varphi(Z)].$$

そしてこの最右辺は達成可能であるから，この，$\varphi \in \dot{\mathcal{P}}_P$ のノルムの二乗が効率限界であることがわかった．

4.3　影響関数に基づく推定量

実際に効率限界を達成する推定量を，影響関数を用いて構成する．

記法 4.3.1. モデル $\mathcal{P}$ に滑らかさ，または，疎性の仮定をおく．

4.3.1　plug-in 推定量の修正

4.4　二重にロバストな推定

Robins et al. 1994 は，推定方程式に $z = 0$ 群の共変量データを用いた項を付加することでデータ利用効率を向上し，推定量の分散を減少させる **augmented inverse probability weighted estimator** を開発した．これは実は局所有効なセミパラメトリック推定量になる．現在では「共変量で結果変数を説明する回帰関数の形で表された項を付加することで局所有効なセミパラメトリック推定量を構成する」手続きは一般化されている．

第 5 章

影響関数の幾何

セミパラメトリックモデルでは，最尤推定が必ずしも良くない (Neyman-Scott 問題 1948)．これを解決に導いたのが情報幾何であった．

記法 5.0.1. Z の密度は，ある測度 ν_Z に関して $(p_Z(z;\theta))_{\theta\in\Omega}$ に属すると仮定する．母数は $\theta = (\beta^T, \eta^T)^T \in \mathbb{R}^q \times \mathbb{R}^r = \mathbb{R}^p$ とする．真値は $\theta_0 = (\beta_0^T, \eta_0^T)$ で表す．

5.1 影響関数

ロバスト解析の文脈から影響の語が取られた (Hampel 74)．良い統計量の例として漸近線型統計量が考えられる．漸近線形統計量は影響関数によって特徴付けられる．M-推定量は漸近線型推定量の例である．

定義 5.1.1 (asymptotic linear estimator, influence function).

(1) 推定量 $\widetilde{\beta_n}$ が次を満たすとき，**漸近線型**であるという：

$$\exists_{\varphi\in H_{\mathrm{mb}}} \quad \sqrt{n}(\widehat{\beta_n} - \beta_0) = \frac{1}{\sqrt{n}}\sum_{i=1}^{n}\varphi(Z_i) + o_p(1)$$

(2) このときの q-次元ベクトル値関数 φ を，推定量 $\widetilde{\beta_n}$ の**影響関数**という．影響関数は実は一意に定まる．

(3) $\mathrm{Avar}(\widetilde{\beta_n}) := E[\varphi(Z,\theta)\varphi(Z,\theta)^T]$ を**漸近分散共分散行列**という．

注 5.1.2. 影響関数 $\varphi(Z)$ は実は真の分布 $p(z,\theta_0)$ の関数である．$\varphi(Z,\theta)$ とも書く．よって，$E[\varphi(Z)]$ は $E_{\theta_0}[\varphi(Z,\theta_0)]$ の略記である．

定理 5.1.3. 任意の漸近線型統計量について影響関数が存在し，ほとんど確実にただ一つに定まる．

第 6 章

経験過程論

大数の法則や中心極限定理などの，推定量の大標本漸近理論を Banach 空間値の確率変数=確率過程にまで拡張する試みを，経験過程論という．ここで退化していた「一様」の概念が出現し，Banach 空間のノルムと関連する．

そもそも確率変数・分布の収束とは，本質的に 2-射的な概念で，圏論的な意味での極限にすごく近いものがあるな．圏論に「図式の極限」という例があるので，関数空間上の点列の収束の概念はそこまで怖くない．

測度の収束には，全変動ノルムによる収束と，各点（各集合）収束とがまず思い付くが，これらよりも広い概念である弱収束が同様な論法で定義できる．

[5] によると，代数幾何 ⇔ 特異点論 ⇔ 超関数論 ⇔ 経験過程 ⇔ 学習理論の途があるらしい．「この架け橋によって，実世界の中にある学習システムの挙動を，対応する数学の概念を用いることにより予言することができる．」

6.1　まとめ

経験過程論

経験過程論とはなにか．標本 $X_1,\cdots,X_n$ を分布と見ることは，線型汎関数とみるということである．すると，中心極限定理は線型汎関数に関して一般化される．測度の空間には作用素ノルムの変種，特に全変動ノルムの一般化と見れるノルムを入れる．この $\mathcal{F}$ が定めるノルムが完備になるか，特に $\mathbb{P}_n$ について収束するかを考える．こちらのほうが幾何学的で，中心極限定理が直感的に捉えやすいのではないか？

定義 6.1.1 (stochastic process, empirical process, sample function / realization / sample path).

(1) 確率空間 $(\mathcal{X},\mathcal{A},P)$ と任意の可測空間 T に関して，可測写像の族 $T\to\mathrm{Meas}(\mathcal{X},\mathbb{R})$ を確率過程という．

(2) 経験測度 $\mathbb{P}_n:=n^{-1}\sum_{i=1}^n\delta_{X_i}$ と任意の可測関数族 $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ について，確率過程 $(\mathbb{P}_nf)_{f\in\mathcal{F}}:\mathcal{F}\to\mathrm{Meas}(\mathcal{X}^n,\mathbb{R})$ が定まる．これを**一般化された経験過程**という．

(3) 経験過程 $\mathcal{F}\to\mathrm{Meas}(\mathcal{X}^n,\mathbb{R})$ が定める確率要素 $\mathcal{X}^n\to\mathrm{Map}(\mathcal{F},\mathbb{R})$ の値 $(X_t(x))_{t\in T}$ を**見本関数**，**実現**または**標本道**という．これは時系列値の確率変数と見れるため，この道を理解することが目的の一つとなる．[†1]

要諦 6.1.2. $\mathcal{X}=\mathbb{R}$ であるとき，経験分布関数 $\mathbb{F}_n:\mathbb{R}\to[0,1]$ は，定義関数のクラス $\mathcal{F}:=\{1_{x\le t}\mid t\in\mathbb{R}\}$ が定める経験過程 $(\mathbb{P}_nf)_{f\in\mathcal{F}}$ と同一視できる．添字集合 T の同型類は，必要に応じて自由に取るが，主な手法は $T\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ の場合について与えられる．経験測度が $\mathrm{Meas}(\mathcal{X},\mathbb{R})$ 上に線型汎関数を定めるためであろう．

(1) 経験過程には弱収束の概念が定義できる．経験過程 $(\mathcal{X}^n\to\mathrm{Map}(\mathcal{F},\mathbb{R}))_{n\in\mathbb{N}}$ は Banach 空間 $l^\infty(\mathbb{R})$-値確率変数の列であり，始域こそ違うが，終域上に定まる測度に関する収束である弱収束は問題なく定義できる．さらに $X_n:\Omega_n\to\mathcal{X}$ 始域が違うことは，外積分の定義に含まれて効いてくるが，$\Omega_n:=\coprod_{n\in\mathbb{N}}\mathcal{X}^n$ に埋め込んで定義することとする．[†2]なお，経験過程の値域は，双対空間 $\mathbb{P}_{(X_1,\cdots,X_n)}:\mathcal{F}\to\mathbb{R}\in\mathcal{X}^{**}$ の元であることに注意．これが距離空間というのは部分的な見方である．

[†1] $w_1\in\mathcal{X}$ を銘柄として，$X(w_1)$ は時系列を返す，など．
[†2] この技術的な議論が完全性 (perfect map) の議論である．

(2) Glivenko-Cantelli 1933 の古典的な結果 $\sup_{t\in\mathbb{R}}|\mathbb{F}_n(t)-F(t)| \xrightarrow[n\to\infty]{\text{a.s.}} 0$ は，任意の経験分布関数 $\mathbb{F}_n:\mathcal{X}^n\to\mathrm{Meas}(\mathbb{R},[0,1])$ は一様に真の分布関数 F に概収束することを保証する．これを，経験過程= Banach 空間値確率変数の列に関する「一様大数の法則」の特殊な場合だと見て，一般に，$\sup_{f\in\mathcal{F}}|\mathbb{P}_nf-Pf| =: \|\mathbb{P}_n-P\|_{\mathcal{F}} \xrightarrow{\text{a.s.}^*} 0$ を満たす確率過程 $(\mathbb{P}_nf)_{f\in\mathcal{F}}$ を，P-Glivenko-Cantelli であるという．前述の結果は，$\mathcal{F}=\{1_{(-\infty,t]}\in\mathrm{Meas}(\mathcal{X},\mathbb{R})\mid t\in\mathbb{R}\}$ に関する特別な場合に相対化される．

(3) これを中心化して，$\mathbb{G}_n:=\sqrt{n}(\mathbb{P}_n-P)$ という符号付測度の w^*-収束を考える．これは標本 $X_1,\cdots,X_n$ と真の分布との誤差に関する主張であり，古典的な中心極限定理は $\mathcal{X}=[0,1]$ について，$\mathbb{G}_n(t):=\sqrt{n}(\mathbb{F}_n(t)-F(t))$ が平均 0 で分散 $F(t)(1-F(t))$ を持つ Gauss 確率変数 $G(t):[0,1]\to[0,1]$ に各点 $t\in[0,1]$ で分布収束するという主張に相当する．Donsker 1952 の古典的な結果は，Dudley による整理と併せて，Brown 橋 $\mathbb{G}_n$ に対して $\|\mathbb{G}_n-\mathbb{G}\|_\infty$ が 0 に確率収束すること，すなわち，$\mathbb{G}_n$ が $\mathbb{G}$ に $l^\infty(\mathcal{F})$ 上の一様ノルムについて弱収束することを示した．[†3]

(4) すなわち，標本 $X_1,\cdots,X_n$ は測度を定め，真の分布との差を考えることは符号付き測度を定めるが，これを線型汎関数と見て，これが（ノルム収束するだけでなく）w^*-収束することを示すことが，誤差論の中心的な議論となる．

議論 6.1.3 (Brownian bridge について). 極限過程 G は Gaussian 過程（任意の有限部分集合 $T_k\subset T$ に関して多変量正規分布を周辺分布にもつ）であり，標準 Brownian bridge $\mathbb{B}$ を用いて $\mathbb{G}=\mathbb{B}(F(t))$ と表せる．よって $\mathbb{G}$ 自体も Brownian bridge と呼ぶ．なお，標準 Brownian bridge は標準ブラウン運動 $\mathbb{W}$ を用いて，$\mathbb{B}(t)=\mathbb{W}(t)-t\mathbb{W}(1)$ $(t\in[0,1])$ と表せる．標準ブラウン運動は $[0,\infty)$ 上の Gaussian 過程の代表であり，$\mathbb{W}(0)=0$ で共分散が $s\wedge t$ で特徴付けられる．

議論 6.1.4 (中心極限定理の翻訳). 経験過程 $\mathbb{G}_n\dfrac{1}{\sqrt{n}}(\mathbb{P}_n-P)$ の $l^\infty(\mathcal{F})$ 上での分布収束は，独立同分布に従う Banach 空間値確率要素 $\delta_{X_1}-P,\cdots,\delta_{X_n}-P$ の Banach 空間 $l^\infty(\mathcal{F})$ 上での中心極限定理とみなせる．逆に，次が成り立つ．

命題 6.1.5. 任意の独立同分布に従う Banach 空間値確率変数に関する中心極限定理は，経験過程に関する結果によって等価に表現できる．

したがって，確率論に対する関数解析の視点からの壮大な方針転換と言える．[†4]

6.2　一般化された弱収束の理論

統計量は一般に一様ノルムについて可測ではないので，必ずしも可測ではない写像列の「弱収束」=「終域上に定める測度の弱収束」の概念を定めたい．そのためには，一般の写像について，それが終域上に引き起こす測度を考えたいから，外確率・外積分の概念を導入する．**そもそも確率変数の収束の定義の仕方は，可測性に関連がない**．このことを明らかにした．これは，Billingsley が 68 に集大成したように，50s の確率変数の収束の議論を，経験過程論との相互作用の中で Hoffmann-Jorgensen と Dudley によって整備された．
このアプローチは非常に自然で，経験過程論の中心的な道具となる（なお，確率過程は自然に Banach 空間値の確率変数と見れるから，一般の距離空間を値に取る確率要素について議論する）．一般の写像の弱収束は伊藤清の確率積分と同様かもしれないが，この弱収束の理論の外積分による拡張は極めて自然で，まさに正道という感じがする．実際，cadlag 過程に対する Skorohod 位相などの不思議な位相ではなく，一様ノルムの位相についての収束が外測度上の収束に対応する．

6.2.1　問題意識：非可測な確率過程の存在

自体はユニタリ表現などのときのように，一様ノルムが定める σ-代数に関して非可測な確率過程が自然に出現する．ノルム位相が強すぎるのである．位相を弱める方針もありえるかもしれないが，ここでは従来の弱収束の定義を拡張したい．

定義 6.2.1 (pseudometric).

[†3] Donsker のオリジナルな結果は，標本道が Skorokhod 空間 $D(-\infty,\infty)$ 上で分布収束することを保証する．

[†4] 一部の中心極限定理は経験過程に翻訳するにあたって不自然な結果をうむので，あまり執着し過ぎるべきではない，とのこと．特に $l^\infty(\mathcal{F})$ という形の Banach 空間について考える．これは $\mathbb{R}^n$ を含む．$\mathcal{F}$ の選び方によって様々な Banach 空間を走るが，これが $\mathcal{F}$ の条件のみによって表せる．

(1) $d(x,y)=0 \Rightarrow x=y$ が必ずしも成り立たない距離関数 $d:D\times D\to\mathbb{R}_{\ge 0}$ を，**擬距離**という．ノルム空間は距離空間を定めるように，半ノルム空間 ($\|u\|=0\Rightarrow u=0$ が成り立たないノルム空間) は擬距離空間を定める．
(2) 擬距離が生成する位相とは，開球 $B_r(x):=\{y\in X \mid d(x,y)<r\}$ を開基として生成する位相をいう．擬距離が距離であるための必要十分条件は，生成する位相が Kolmogorov(T_0) であることに同値．この位相が生成する Borel σ-集合体を $\mathcal{D}$ と表すこととする．

注 6.2.2. 擬距離と半ノルムは関数空間では極めて自然な概念である．例えば実数値関数空間 $f,g\in\mathcal{F}(X)$ について，$d(f,g):=|f(x_0)-g(x_0)|$ $(x_0\in X)$ と定めると，これは擬距離である．すると，位相が識別可能であるかどうか (T_0) が焦点になり，これに Kolmogorov の名前がついている．

記法 6.2.3 (関数空間の記法)**.**

(1) 実数値有界関数の全体を $l^\infty(D)$ と表す．一般に $l^\infty(J)$ は直和 $\oplus_{j\in J}\mathbb{R}=\left\{(j,x_j)\in J\times\mathbb{R}\;\middle|\;\sup_{j\in J}x_j<\infty\right\}$ に一様ノルムを入れたもので，$J=\mathbb{N}$ の場合が数列空間であるが，その一般化と捉えるのである．
(2) その上の一様ノルムを $\|f\|_\infty$ または $\|f\|_D$ などと表す．
(3) 実数値有界連続関数の空間を $C_b(X)$ と表す．

定義 6.2.4 (random element (Fréchet 48))**.** 確率空間 $(\Omega,\mathcal{A},P)$ から距離空間 (D,d) への $\mathcal{A}/\mathcal{D}$-可測関数 $X:\Omega\to D$ を**確率要素**という．$D=\mathbb{R}$ の場合を**確率変数**という．

定義 6.2.5 (weak convergence of measure)**.** 測度を測度空間 $(D,\mathcal{D})$ 上の有界連続関数の空間 $C_b(D)$ 上に定まる線型汎関数とみなす．このときの $\sigma(C_b(D)^*,C_b(D))$ 位相に関する収束，すなわち弱 $*$-収束を，測度の弱収束という．詳しくは，次の通り：

(1) Borel 可測空間 $(D,\mathcal{D})$ 上の Borel 確率測度の列 (L_n) が L に弱収束する $L_n\rightsquigarrow L$ とは，次が成り立つことをいう：$\forall_{f\in C_b(D)}\ \lim_{n\to\infty}\int_D fdL_b=\int_D fdL.$
(2) L_n,L を分布として引き起こす確率要素 $X_n:\Omega_n\to D, X:\Omega\to D$ を用いると，この条件は次のように特徴付けられる：$\forall_{f\in C_b(D)}\ \lim_{n\to\infty}E[f(X_n)]=E[f(X)]$．この条件を，$X_n$ が X あるいは L に**弱収束**するといい，$X_n\rightsquigarrow X$ や $X_n\rightsquigarrow L$ と表す．

注 6.2.6. (2) で確率過程 (X_n) の弱収束の概念が登場しているが，重要な確率過程で Borel 可測でないものが存在する．

例 6.2.7 (Borel 可測でない確率過程：D-値過程)**.** $B([0,1])$ は一様ノルムに関して Banach 空間であり，一様距離を定める．これの部分空間 $D([0,1])\subset B([0,1])$ として Skorokhod 空間を考えると，これは Skorokhod 距離とは異なる．この位相が定める Borel σ-加法族を $\mathcal{D}$ とし，$\mathcal{B}$ を $[0,1]$ の Borel σ-加法族，λ を $[0,1]$ の Lebesgue 測度とする．一様分布に独立に従う確率変数 $\xi_i:([0,1],\mathcal{B},\lambda)^n\to([0,1],\mathcal{B},\lambda)$ $(i\in[n])$ を射影として定める．このとき，

- 経験分布過程 $F_n(t)=\dfrac{1}{n}\sum_{i=1}^n\chi_{[0,t]}(\xi_i)$
- 一様経験過程 $X_n(t)=\sqrt{n}(F_n(t)-t)$

は D 空間への写像 $[0,1]^n\to D([0,1])$ を定めるが，いずれも $\mathcal{D}/\mathcal{B}^n$-可測ではない．特に，$\sigma$-加法族を $\mathcal{D}$ が大きすぎるので，$\mathcal{B}^n$ では要素数が足りない．ユニタリ表現における $B(H)$ に入れる位相で，ノルム位相では強すぎるという状況に似ている．$n=1$ の場合で実際に見てみると，確率変数 $\xi(\omega)=\omega\in[0,1]$ に対して，経験分布関数はただの特性関数 χ_ω となるから，関数 $z:[0,1]\to D([0,1])$ を $z_x(t):=\chi_{[x,1]}(t)-t$ と定めるとこれは $z_x\in D([0,1])$ で，一様経験過程は $X_1(t)=z_{\xi(\omega)}(t)=z_\omega(t)$ と表せる．$X_1^{-1}(A)\notin\mathcal{B}$ を満たす $A\in D([0,1])$ を構成する．$[0,1]$ の非 Borel 集合 $H\notin B$ を一つとり，$A:=\cup_{x\in H}B_{1/2}(z_x)$ と定めれば良い．これは $D([0,1])$ の開集合であるから $A\in\mathcal{D}$ であるが，$D([0,1])$ の距離は一様距離としたから，$x_1\ne x_2$ ならば，$d(z_{x_1},z_{x_2})=\sup_{t\in[0,1]}|\chi_{x_1,1}(t)-\chi_{x_2,1}(t)|=1$ となるので，$\forall_{x\in[0,1]}\ B_{1/2}(z_x)\cap A=\{z_x\}$ であり，$X_1^{-1}(A)=\{x\in[0,1]\mid z_x\in A\}=H\notin B$ である．

定義 6.2.8 (RCLL / càdlàg, càglàd, Skorokhod space)**.** 関数 $f:[a,b]\to\mathbb{R}$ について，

(1) $[a,b)$ 上右連続 $\lim_{x\searrow x_0}f(x)=f(x_0)$ で，$(a,b]$ 上左極限を持つ $\lim_{x\nearrow x_0}f(x)\in\mathbb{R}$ とき，**右連続左極限関数**または càdlàg (continue à droite, limites à gauche) 関数または RCLL (Right Continuous with Left Limits) という．
(2) $(a,b]$ 上左連続で，$[a,b)$ 上右極限を持つとき，f を**左連続右極限関数**という．

(3) càdlàg 関数 $f : E \to M$ 全体からなる集合を $D(E;M)$ または $D(E)$ と表し，**スコロホッド空間**という．E が閉区間であるとき，$D(E) \subset B(E)$ である．

これらは Weiner 過程のように連続ではなく，ジャンプをもつ過程を捉える基本言語となる．

例 6.2.9.

(1) 任意の累積分布関数は càdlàg である．考えている区間 $(-\infty, a]$ が右閉であるからである．
(2) 基本的には，右から左へ遡及してみて行ったときに，左端が閉じていれば良い．

補題 6.2.10 (càdlàg 関数は有界)**.** 集合 $T \subset [a,b]$ 上の関数 f の連続度 $w_f : P([a,b]) \to \mathbb{R}$ を $w_f(T) := \sup_{u,v \in T} |f(u) - f(v)|$ で定める．

(1) 任意の $\epsilon > 0$ に対して，$[a,b]$ の分割 $a = t_0 < t_1 < \cdots < t_N = b$ であって，$\forall_{i \in m}\ w_f([t_{i-1}, t_i)) < \epsilon$ を満たすものが存在する．

$D([a,b]) \subset B([a,b])$．すなわち，区間 $[a,b]$ 上の càdlàg 関数 f は有界である．

［証明］.

(1)
$$\delta^* := \sup\left\{\delta \in (0, b-a] \mid [0,\delta) \text{ の分割 } 0 = t_0 < t_1 < \cdots < t_N \leq \delta \text{であって，条件を満たすものが存在する}\right\}$$
と定めると，$\delta^* = b - a$ を示せば良い．f の右連続性から，0 に十分小さい点では $\forall_{\epsilon>0} \sup_{u,v \in [0,\delta)} |f(u) - f(v)| < \epsilon$ を満たすから，$\delta^* > 0$．$0 < \delta^* < b - a$ と仮定して矛盾を示せば良い．このとき，左極限が存在するから，$[0, \delta^*]$ の分割で条件を満たすものが存在する．また，δ^* における右極限が存在するから，$r > 0$ を十分小さくすることで $w_f([\delta^*, \delta^* + r)) < \epsilon$ かつ $[\delta^*, \delta^* + r) \subset [a,b]$ を満たすようにできる．したがって，δ^* の最大性に矛盾．
(2) $\epsilon = 1$ などとして適用すれば良い．

■

要諦 6.2.11. ジャンプを無限個持つことがあり得ることも含めて議論が成り立っている！例え無限個ジャンプを持っても，そのジャンプの幅の合計が無限に飛ぶことはない．よって，連続度のように，y 軸で切る見方がうまくいく．これは見事だ．

6.2.2　外積分

Kosorok[8] の 6.2 そのままである．外測度とほとんど同様の議論を関数空間上で再展開しただけである．可測な写像 $U : \Omega \to \overline{\mathbb{R}}$ を用いて，その $E : \mathrm{Meas}(\Omega, \overline{\mathbb{R}}) \to \overline{\mathbb{R}}$ による像の下限と定めるのだが，使うときは列をとって議論する．

6.2.2.1　定義と特徴付け

非可測な対象を扱うためには，Caratheodory の頃から outer な対象を使うと決まっている．こうして，$E : \mathrm{Meas}(\Omega, \overline{\mathbb{R}}) \to \overline{\mathbb{R}}$ の延長 $E^* : \mathrm{Map}(\Omega, \overline{\mathbb{R}}) \to \overline{\mathbb{R}}$ が定まり，実は $\mathrm{Im}\, E = \mathrm{Im}\, E^*$ が成り立つ．
いや，寧ろ積分とは，外積分の制限であるという Carathéodory の理論も拡張できないか？その見方で，E^* が基本的に非負・単調・σ-劣加法的であると特徴付け 6.2.13 の結論を見た方がスッキリするようだ．$-^* : \mathrm{Map}(\Omega, \overline{\mathbb{R}}) \to \mathrm{Meas}(\Omega, \overline{\mathbb{R}})$ も同様の単調写像と思える．

定義 6.2.12 (outer integral / expectation, outer probability)**.** $(\Omega, \mathcal{A}, P)$ を確率空間とする．

(1) 写像 $T \in \mathrm{Map}(\Omega, \overline{\mathbb{R}})$ に対して，T の P による**外積分**とは，
$$E^*T := \inf\left\{EU \in \overline{\mathbb{R}} \mid U \geq T, U \in \mathrm{Meas}(\Omega, \overline{\mathbb{R}}), EU \in \overline{\mathbb{R}}\text{が積分確定}\right\}$$

をいう．[†5]

(2) 部分集合 $B \in P(\Omega)$ に対して，**外確率**とは，$P^*(B) := \inf\{P(A) \in [0,1] \mid A \in \mathcal{A}, B \subset A\}$ をいう．すなわち，$P^*(B) = E^*[1_B]$ であり，外積分の特殊な場合と考えられる．

T の上限とは $-T$ の下限であるから，双対概念「内積分」「内確率」は次のように定めれば良い：$E_*[T] := -E^*[-T]$，$P_*(B) := 1 - P^*(\Omega \setminus B)$.

補題 6.2.13 (外積分の特徴付け：minimal measurable majorant, maximal measurable minorant)**.** 任意の写像 $T \in \mathrm{Map}(\Omega, \overline{\mathbb{R}})$ に対して，可測関数 $T^* \in \mathrm{Meas}(\Omega, \overline{\mathbb{R}})$ が存在して，次の 2 条件と (3) を満たす：

(1) $T^* \geq T$,
(2) $\forall_{U \in \mathrm{Meas}(U, \overline{\mathbb{R}})}\ U \geq T$ a.s. $\Rightarrow U \geq T^*$ a.s.（本質的に極小）.
(3) ET^* が存在するならば $E^*T = ET^*$ である．

この T^* を**極小可測優関数**という．

［証明］．

T^* の構成　$f \in B(\mathbb{R})$ を狭義単調増加な有界連続関数とし，$\overline{\mathbb{R}}$ への連続拡張 $\overline{f} : \overline{\mathbb{R}} \to \mathbb{R}$ を考える．

合成写像 $f \circ T : \Omega \to \overline{\mathbb{R}}$ についての外積分 $E^*[f(T)]$ について，定義から，積分確定な $\mathrm{Meas}(\Omega, \overline{\mathbb{R}})$ の列 (U_n) が存在して，$f \circ T \leq U_n \leq f(\infty)$ かつ $E[U_n] \searrow E^*[f(T)]$ を満たすものが存在する．ここで，$V_n := f^* U_n : \Omega \to \overline{\mathbb{R}}$ を考えると，条件 $E[U_n] \searrow E^*[f(T)]$ は $E[f(V_n)] \searrow E^*[f(T)]$ と表せる．このとき，$T^* := \lim_{n\to\infty} \inf_{1 \leq k \leq n} V_k$ と定めれば良い．

$$\begin{tikzcd}
\Omega \arrow[d, bend right, "T"'] \arrow[d, dotted, "V_n"] \arrow[dr, "U_n"] & \\
\overline{\mathbb{R}} \arrow[r, "\overline{f}"'] & \overline{\mathbb{R}}
\end{tikzcd}$$

検証　このとき，$T \leq \inf_{1\leq k\leq n} V_k \leq V_n$ で，第二項と第三項は可測関数だから，

$$E^*[f(T)] \leq E[f(\inf_{1\leq k\leq n} V_k)] \leq E[f(V_n)]$$

が成り立つ．$n \to \infty$ の場合を考えると，$E^*[f(T)] = E[f(T^*)]$.

(1) 構成から，$f \circ T \leq U \iff f \circ T \leq f \circ V_n$ より，$T \leq V_n$ であるから，$T \leq T^* = \lim_{n\to\infty} \inf_{1\leq k\leq n} V_k$ である．

(2) 可測写像 $V \in \mathrm{Meas}(\Omega, \overline{\mathbb{R}})$ が $T \leq V$ a.s. を満たすとする．このとき，(1) から $f \circ T \leq f \circ (T^* \wedge V)$ a.s. でもあるから，$T^* \wedge V$ が可測であることに注意して，

$$E[f(T^*)] = E^*[f(T)] \leq E[f(T^* \wedge V)]$$

より，$f(T^*) = f(T^* \wedge V)$ a.s. $\iff T^* \leq V$ a.s..

(3) T^* が積分確定のとき，(1) より $E^*[T] \leq E[T^*]$ であるから，逆向きの不等号を示せば良い．任意の $V \in \mathrm{Meas}(\Omega, \overline{\mathbb{R}})$ について，(2) より $T \leq V \Rightarrow T^* \leq V \Rightarrow E[T^*] \leq E[V]$ であるから，V に関しての下限を考えると，$E[T^*] \leq E^*[T]$.

■

要諦 6.2.14.

(1) なんだこの構成は．
(2) 結局外積分は，Lebesgue 積分のときの議論と同様，$\mathrm{Im}\ E = \mathrm{Im}\ E^*$．そのときのファイバーは極小可測優関数 T^* で $E^*T = ET^*$ と与えられる．これは Lebesgue 可測性の位相的特徴づけよりも結論が強い．外測度の場合はこちらが先に定義され，測度はその制限とされた．なるほど，外積分も絶対公理化できるな．そして Lebesgue 積分とはその制限なのかもしれない！！
(3) (2) の主張を，* の単調性 $U \geq T \Rightarrow U^* = U \geq T^*$ と読むと理解しやすい．

[†5] 統計量 T は可測とは限らない＝確率変数とは限らないので，期待値とは呼ばずに積分と読んだ方が良いだろう．

補題 6.2.15 (外確率の特徴付け)**.** 任意の集合 $B\subset\Omega$ に対して，可測集合 $B^*\in\mathcal{A}$ が存在して，次の 2 条件と (3) を満たす：

(1) $B^*\supset B$,
(2) $\forall_{A\in\mathcal{A}}\ B\subset A\Rightarrow B^*\subset A$.
(3) $1_{B^*}=(1_B)^*$ a.s. 及び $P(B^*)=E(1_B)^*=P^*(B)$ が成り立つ.

［証明］.

構成　B に対して，1_B の極小可測優関数 $(1_B)^*$ を用いて $B^*:=\{\omega\in\Omega\mid(1_B)^*(\omega)\ge 1\}$ と定めれば良い．$B^*\in\mathcal{A}$ かつ $(1_B)^*=1_{B^*}$ である．

検証
(1) 外積分の特徴付け 6.2.13(1) より，$1_{B^*}=(1_B)^*\ge 1_B$ より，$B\subset B^*$.
(2) $B\subset A$ を満たす任意の可測集合 $A\in\mathcal{A}$ について，$1_B\le 1_A$ より，$(1_B)^*=1_{B^*}\le 1_A=(1_A)^*$．よって，$B^*\subset A$.
(3) $P^*(B)=E^*[1_B]$ を示せば，外積分の特徴付け 6.2.13(3) より，$E^*[1_B]=E[(1_B)^*]=E[1_{B^*}]=P(B^*)$ が従う．$E^*[1_B]\le P^*(B)$ は，任意の可測集合 $A\in\mathcal{A}$ について，$B\subset A\Leftrightarrow 1_B\le 1_A\Rightarrow E^*[1_B]\le E[1_A]=P(A)$．$A$ についての下限を考えて，$E^*[1_B]\le P^*(B)$ がわかる．逆の $E^*[1_B]\ge P^*(B)$ も，任意の積分確定な可測関数 $U\ge 1_B$ について，$B\subset\{U\ge 1\}$ より，$P^*(B)\le P(U\ge 1)\le E[U]$．$U$ に関しての下限を考えて，$P^*(B)\le E^*[1_B]$ を得る．

■

要諦 6.2.16. 特性関数と可測性の交錯があまりにも見事だ.

6.2.2.2　外積分の性質

外積分・外確率の計算は結局，写像 T や集合 A について，極小可測優関数 T^* と極小可測集合 A^* を考えれば良いのであった．そこで，この 2 つについての計算規則をまとめる．それにしても，今まで見たことのない計算規則．いや，こういう対称性破れは，Borel クラスの位相的正則性のときから見慣れているものと同根かもしれない．

補題 6.2.17. 任意の写像 $S,T\in\mathrm{Map}(\Omega,\overline{\mathbb{R}})$ に対して，次の関係式が意味を持つならば，確率 1 で成り立つ.

(1) $S_*+T^*\le(S+T)^*\le S^*+T^*$：等号成立は S が可測であるとき.
(2) $S_*+T_*\le(S+T)_*\le S_*+T^*$：等号成立は T が可測であるとき.
(3) $(S-T)^*\ge S^*-T^*$.
(4) $|S^*-T^*|\le|S-T|^*$.
(5) $\forall_{c\in\mathbb{R}}\ (1_{\{T>c\}})^*=1_{\{T^*>c\}}$.
(6) $\forall_{c\in\mathbb{R}}\ (1_{\{T\ge c\}})^*=1_{\{T_*\ge c\}}$.
(7) $(S\vee T)^*=S^*\vee T^*$.
(8) $(S\wedge T)^*\le S^*\wedge T^*$：等号成立は S が可測であるとき.

補題 6.2.18. 任意の集合 $A,B\subset\Omega$ に対して，

(1) $(A\cup B)^*=A^*\cup B^*$，$(A\cap B)_*=A_*\cap B_*$.
(2) $(A\cap B)^*\subset A^*\cap B^*$，$(A\cup B)_*\supset A_*\cup B_*$．$A,B$ のいずれか一方が可測である場合は逆も成立.
(3) $A\cap B=\emptyset\Rightarrow P_*(A)+P_*(B)\le P_*(A\cup B)\le P^*(A\cup B)\le P^*(A)+P^*(B)$．特に，最右辺の結論には前提 $A\cap B=\emptyset$ が必要ないことに注意.
(4) $P_*(A\cap B)\ge P_*(A)+P_*(B)-1$.

［証明］.

(1) 1
(2) 2
(3) 3

(4) (3) で $A = \overline{A}, B = \overline{B}$ を考えると，

$$\begin{aligned} & P^*(\overline{A} \cup \overline{B}) \le P^*(\overline{A}) + P^*(\overline{B}) \\ \Leftrightarrow\quad & 1 - P_*(A \cap B) \le 1 - P_*(A) + 1 - P_*(B) \\ \Leftrightarrow\quad & P_*(A \cap B) \ge P_*(A) + P_*(B) - 1 \end{aligned}$$

より，$A \cup B = \Omega \Rightarrow P_*(A \cap B) \ge P_*(A) + P_*(B) - 1$ がわかる．$A \cup B = \Omega$ ではない場合も，部分空間 $A \cup B \subset \Omega$ で考えれば良い．

■

補題 6.2.19. $T : \Omega \to \mathbb{R}$ を写像，$\phi : \mathbb{R} \to \mathbb{R}$ は広義単調で $\overline{\mathbb{R}}$ への延長をもつとする．次の関係式が意味をもつならば，それは確率 1 で成り立つ．

(1) ϕ が広義単調増加ならば，
  (a) $\phi(T^*) \ge (\phi(T))^*$: ϕ が $[-\infty, \infty)$ 上左連続のとき等号成立．
  (b) $\phi(T_*) \le (\phi(T))_*$: ϕ が $(-\infty, \infty]$ 上右連続のとき等号成立．
(2) ϕ が広義単調減少ならば，
  (a) $\phi(T^*) \le (\phi(T))_*$: ϕ が $[-\infty, \infty)$ 上左連続のとき等号成立．
  (b) $\phi(T_*) \ge (\phi(T))^*$: ϕ が $(-\infty, \infty]$ 上右連続のとき等号成立．

6.2.2.3　Chebyshev の不等式

> Jensen の不等式など，種々の不等式が引き継がれる．

補題 6.2.20 (外積分における Chebyshev の不等式)**.** 任意の写像 $T : \Omega \to \mathbb{R}$ と $(0, \infty)$ 上正値な単調増加関数 $\phi : [0, \infty) \to [0, \infty)$ について，次が成り立つ：

$$\forall_{u>0}\ P^*(|T| \ge u) \le \frac{E^*\phi(|T|)}{\phi(u)}.$$

［証明］．任意の $u > 0$ について，

$$\begin{aligned} \phi(u)P^*(|T| \ge u) &= \phi(u)E^*[1_{\{|T|\ge u\}}] \\ &= E^*[\phi(u)1_{\{|T|\ge u\}}] && \because \phi(u) > 0 \\ &\le E^*[\phi(|T|)]. \end{aligned}$$

■

6.2.2.4　完全性と Fubini の定理

> Fubini の定理が，一般化された Lebesgue の優収束定理と同じような消息になる．もうだめだ，もはやこれが本来的だと感じる．

定義 6.2.21 (perfect)**.** 可測関数 ϕ と写像 T に関して，$T^* \circ \phi$ は $T \circ \phi \le T^* \circ \phi$ を満たす可測関数だから，外積分の特徴付け 6.2.13(2) より，$(T \circ \phi)^* \le T^* \circ \phi$ であるが，逆も成り立つかはわからない．

$$\begin{tikzcd} (\widetilde{\Omega}, \widetilde{\mathcal{A}}, \widetilde{P}) \arrow[r, "\phi"] \arrow[dr, "T\circ\phi"'] & (\Omega, A, \phi_*\widetilde{P} = P) \arrow[d, "T"] \arrow[d, bend right, "T^*"'] \\ & \mathbb{R} \end{tikzcd}$$

そこで，$\forall_{T \in B(\Omega)}\ (T \circ \phi)^* = T^* \circ \phi$ を満たす可測関数 $\phi : \widetilde{\Omega} \to \Omega$ を完全という．したがって，任意の $T : \Omega \to \mathbb{R}$ に対して，$\widetilde{E}^*(T \circ \phi) = \widetilde{E}(T^* \circ \phi) = E[T^*] = \int T^* dP$ と計算できる．特に，任意の $B \subset \Omega$ について $T = \chi_B$ とすれば，$\widetilde{P}^*(\phi \in B) = P^*(B)$ である．

補題 6.2.22. 直積確率空間からの射影は完全である．

定義 6.2.23 (累次外積分)**.** $T:(\Omega_1\times\Omega_2,\mathcal{A}_1\times\mathcal{A}_2,P_1\times P_2)\to\mathbb{R}$ を写像とする．

(1) $(E_2^*[T])(\omega_1):=\inf\left\{E_2[U]\in\overline{\mathbb{R}}\mid\forall_{\omega_2\in\Omega_2}\;U(\omega_2)\ge T(\omega_1,\omega_2),U\in\mathrm{Meas}(\Omega_2,\overline{\mathbb{R}}),E_2U\text{ が存在}\right\}$ とする．「$\omega_1\in\Omega_1$ について止めて一変数と見たときの外積分」と定める．
(2) $E_1^*(E_2^*T)$ を，関数 $E_2^*T:\Omega_1\to\overline{\mathbb{R}}$ の外積分とする．

定理 6.2.24 (Fubini)**.** $T:(\Omega_1\times\Omega_2,\mathcal{A}_1\times\mathcal{A}_2,P_1\times P_2)\to\mathbb{R}$ を写像とする．このとき，

$$E_*T\le E_{1*}E_{2*}T\le E_1^*E_2^*T\le E^*T.$$

［証明］．

方針 $E^*[T]<\infty$ の場合に $E_1^*[E_2^*[T]]\le E^*[T]$ を示せば良い．このとき，$T\le U$ かつ $E^*[T]\le E[U]<\infty$ を満たす可測関数 $U\in\mathrm{Meas}(\Omega_1\times\Omega_2,\overline{\mathbb{R}})$ が取れる．すると，$T^*\le U$ であるから，$E[(T^*)^+]\le E[U^+]<\infty$ である．

証明 いま $T\le T^*$ だから，$\forall_{\omega_1\in\Omega_1}\;T(\omega_1,-)\le T^*(\omega_1,-)$ より，

$$\begin{aligned}E_2^*[T]&\le E_2[T^*]\quad P_1\text{- a.s.}\\&=E_2[(T^*)^+]-E_2[(T^*)^-]\end{aligned}$$

であるが，$E[(T^*)^+]<\infty$ であったから，右辺は積分確定．よって，これについての外積分を考えて，

$$\begin{aligned}E_1^*[E_2^*[T]]&\le E_1[E_2[(T^*)^+]-E_2[(T^*)^-]]\\&=E[(T^*)^+]-E[(T^*)^-]&&\because\text{ Fubini の定理}\\&=E[T^*]=E^*[T].\end{aligned}$$

■

6.2.3　緊密性と可分性

可分・緊密な確率変数

可分確率変数＝ Radon 確率変数というクラスを定義する．これは「確率変数の像が可分」ということを確率論的に緩めて，「確率 1 で可分な空間に含まれる」ことと定義する．σ-コンパクト性と σ-有限性の混合のように，「確率 1 に至るコンパクト集合の増大列が取れる」ことを緊密という．コンパクト集合の増大列が取れなくても全有界集合の増大列が取れる場合，プレタイトという．

定義 6.2.25 (tight, separable, pretight)**.** 距離空間上の Borel 確率空間 $(D,\mathcal{B}(D),L)$ 上の Borel 可測写像 $X:\Omega\to D$ について，

(1) Borel 可測写像 $X:\Omega\to D$ が**緊密**であるとは，内部正則であることをいう：$\forall_{\epsilon>0}\;\exists_{K\overset{\mathrm{cpt}}{\subset}D}\;L^X(K)\ge1-\epsilon.$[†6]
(2) 可算な稠密部分集合が存在する位相空間 X，すなわち X の任意の空でない開集合が交わりを持つような X の点列が存在するような空間を**可分**という．
(3) Borel 可測写像 $X:\Omega\to D$ が**可分**であるとは，可分な可測集合 $A\in\mathcal{B}(D)$ が存在して $L^X(A)=1$ を満たすことをいう．
(4) 測度 L が**プレタイト**であるとは，任意の $\epsilon>0$ に対して，全有界な可測集合 $B\in\mathcal{B}(D)$ であって，$L^X(B)\ge1-\epsilon$ を満たすものが存在することをいう．

補題 6.2.26. 距離空間上の Borel 確率空間 $(D,\mathcal{B}(D),L)$ について，(1)⇔(2)⇒(3) であり，D が完備であるとき (3) も同値である．

(1) L は緊密である．
(2) σ-コンパクト集合 $\widetilde{K}\in\mathcal{B}(D)$ が存在して，$L(\widetilde{K})=1$ を満たす．
(3) L は可分である．

［証明］．

[†6] ある種の σ-compect 性や σ-有限性の拡張に思える．

(1)⇒(2)　任意の $n\in\mathbb{N}$ について，$L(K_n)\ge 1-1/n$ を満たす $K_n \overset{\text{cpt}}{\subset} D$ が取れる．これについて $\widetilde{K}:=\cup_{n=1}^\infty K_n$ と置くとこれは σ-コンパクトで，$\forall_{n\in\mathbb{N}}\ \widetilde{K}\supset K_n$ より $\forall_{n\in\mathbb{N}}\ L(\widetilde{K})\ge 1-1/n$ だから $L(\widetilde{K})=1$．

(2)⇒(1)　σ-コンパクトな集合 $\widetilde{K}\subset D$ が存在して，$L(\widetilde{K})=1$ を満たす．仮に $\widetilde{K}=\cup_{n=1}^\infty K_n$ とし，コンパクト集合の列 $M_i:=\cup_{n=1}^i K_n$ を考えると，これは $\widetilde{K}$ に収束する．よって，$\forall_{\epsilon>0}\ \exists_{N>0}\ \forall_{n\ge N}\ L(M_n)\ge 1-\epsilon$．よって，$K:=\cup_{n=1}^N K_n$ と定めれば良い．コンパクト集合の有限合併はコンパクトである．

(2)⇒(3)　距離空間において σ-コンパクトな集合は可分であることから従う．

距離空間のコンパクト性は全有界かつ任意の被覆にルベーグ数が存在することで特徴付けられるが，距離空間が全有界ならば可分である．

(3)⇒(2)　？

■

補題 6.2.27. Borel 確率変数 $L:\Omega\to D$ が可分であるならば，プレタイトである．

例 6.2.28 (標準的な確率空間の場合). $D=\mathbb{R}^n$ のとき，$\mathbb{R}^n$ の全ての部分集合は σ-コンパクトだから，全ての Borel 確率測度は緊密である．

6.2.4　弱収束の定義と特徴付け

任意の（可測とは限らない）写像列の，Borel 確率変数への弱収束を議論する（したがって弱収束とは純測度論的な概念ではない）．これには，$B(D)$ を，そして Riesz 空間の関数を媒介に用いる．

定義 6.2.29 (weak convergence of mappings). 確率空間 $(\Omega_n,\mathcal{A}_n,P_n)$ からの D-値写像の列を (X_n) とする．

(1) (X_n) が $(D,\mathcal{D})$ 上の Borel 確率測度 $L:\mathcal{D}\to[0,1]$ に**弱収束**するとは，次が成り立つことをいう：$\forall_{f\in C_b(D)}\ E^*[f(X_n)]\xrightarrow{n\to\infty} E[f]$.

(2) (X_n) が Borel 可測写像 $X:\Omega\to D$ に**弱収束**するとは，次が成り立つことをいう：$\forall_{f\in C_b(D)}\ E^*[f(X_n)]\xrightarrow{n\to\infty} E[f(X)]$.

ただし，各 $E^*[f(X_n)]$ は P_n の定める外積分で，$E[f(X)]$ は L の定める期待値である．

定義 6.2.30 (Riesz space / vector lattice).

(1) $\mathcal{F}\subset l^\infty(D)$ が実線型空間かつ $\wedge,\vee$ について束をなすとき，**Riesz 空間**または**線型束**という．

(2) $\mathcal{F}$ が D の各点を分離するとは，$\forall_{x,y\in D}\ x\ne y\Rightarrow[\exists_{f\in\mathcal{F}}\ f(x)\ne f(y)]$.

定義 6.2.31 (Lipschitz norm, Lipschitz (continuous) map).

(1) Lipschitz ノルムを $\|f\|_{\text{lip}}:=\inf\left\{L\in\overline{R}\mid\forall_{x,y\in D}\ |f(x)-f(y)|\le Ld(x,y)\right\}=\sup_{x\ne y\in D}\dfrac{|f(x)-f(y)|}{d(x,y)}$ と定める．

(2) Lipschitz ノルムが有限となる関数を **Lipschitz（連続）関数**という．

(3) 上限が 1 以下または short な[†7]Lipschitz 連続関数の空間の部分空間 $\text{BL}_1(D):=\{f\in C_b(D)\mid\|f\|_\infty\vee\|f\|_{\text{lip}}\le 1\}\subset\text{BL}(D)$ に名前をつける．

補題 6.2.32 (Lipschitz 連続性).

(1) $\|f\|_{\text{lip}}$ はノルムになる．

(2) Lipschitz 連続関数 $f:\mathbb{R}\to\mathbb{R}$ は絶対連続であり，したがって殆ど至る所微分可能である．

(3) 絶対連続関数は一様連続である．

(4) 微分可能な Lipschitz 連続関数 $f:\mathbb{R}\to\mathbb{R}$ について，$\|f\|_{\text{lip}}=\|f'\|_\infty$ である．

［証明］.

[†7] 非拡大な写像を short という．縮小写像は short map である．

(1) (a) $\|f\|_{\mathrm{lip}} = 0$ とすると，$\forall_{x,y\in D}\ |f(x) - f(y)| = 0$ であるから，$f = 0$.

(b) 任意に $a \in \mathbb{R}$ を取ると，任意の $x, y \in D$ について，$|af(x) - af(y)| = |a||f(x) - f(y)| \le \|af\|_{\mathrm{lip}} d(x,y)$ より，$\|f\|_{\mathrm{lip}} = \dfrac{\|af\|_{\mathrm{lip}}}{|a|}$.

(c) 任意の $x, y \in D$ について，$|(f+g)(x) - (f+g)(y)| \le |f(x) - f(y)| + |g(x) - g(y)| \le (\|f\|_{\mathrm{lip}} + \|g\|_{\mathrm{lip}})d(x,y)$ より，$\|f+g\|_{\mathrm{lip}} \le \|f\|_{\mathrm{lip}} + \|g\|_{\mathrm{lip}}$.

■

補題 6.2.33 (弱収束の概念の安定性)**.** 距離空間 D 上の Borel 確率測度 L_1, L_2 について，次の (1) と (2) は同値である．L_1, L_2 が可分ならば (3) も同値で，緊密ならば (4) も同値である．

(1) $L_1 = L_2$.

(2) $\forall_{f\in C_b(D)}\ \displaystyle\int_D f dL_1 = \int_D f dL_2$.

(3) $\forall_{f\in \mathrm{BL}_1(D)}\ \displaystyle\int_D f dL_1 = \int_D f dL_2$.

(4) $\mathcal{F} \subset C_b(D)$ が定数関数を含み，D の各点を分離するベクトル束ならば，$\forall_{f\in\mathcal{F}}\ \displaystyle\int_D f dL_1 = \int_D f dL_2$.

［証明］.

(1)⇒(2)　自明.

(2)⇒(1)　Borel クラスの上で $L_1 = L_2$ を示せば良いから，任意の $G \overset{\text{open}}{\subset} D$ について $L_1(G) = L_2(G)$ を導けば良い．任意の $G \overset{\text{open}}{\subset} D$ に対して $f_m(x) := (m \cdot d(x, D\setminus G)) \wedge 1$ と定めると，f_m は非負値で $\overline{G}$ を台とする，有界な Lipschitz 関数であり ($\|f\|_{\mathrm{lip}} \le m$ より)，$0 \le f_m \nearrow 1_G\ (m \to \infty)$ を満たす．よって，$\forall_{m\in\mathbb{N}}\ \displaystyle\int_D f_m dL_1 = \int_D f_m dL_2$ と単調収束定理より，$G(L_1) = G(L_2)$.

■

要諦 6.2.34. $f_m(x) := (m \cdot d(x, D\setminus G)) \wedge 1$ という非負値で $\overline{G}$ を台とする有界な Lipschitz 関数は，微分可能性こそないものの，随分普遍的に使えそうなテクニックである．(4) は Stone-Weierstrass の定理による．

定理 6.2.35 (portmanteau (仏:旅行鞄)：弱収束の特徴付け)**.** 確率空間 $(\Omega_n, \mathcal{A}_n, P_n)$ 上の任意の写像列 $(X_n : \Omega_n \to D)$ と D 上の Borel 確率測度 L について，次の 7 条件は同値．また，L が可分で Borel 可測写像 X が L を分布に持つとき，(8) も同値．

(1) X_n は L に弱収束する．

(2) 任意の開集合 $G \subset D$ に対して，$\displaystyle\liminf_{n\to\infty} P_*(X_n \in G) \ge L(G)$.

(3) 任意の閉集合 $F \subset D$ に対して，$\displaystyle\limsup_{n\to\infty} P^*(X_n \in F) \le L(F)$.

(4) 下に有界な全ての下半連続関数 f に対して，$\displaystyle\liminf_{n\to\infty} E_* f(X_n) \ge \int_D f dL$.

(5) 上に有界な全ての上半連続関数 f に対して，$\displaystyle\limsup_{n\to\infty} E^* f(X_n) \le \int_D f dL$.

(6) $L(\partial B) = 0$ を満たす任意の Borel 集合 B に対して，$\displaystyle\lim_{n\to\infty} P^*(X_n \in B) = \lim_{n\to\infty} P_*(X_n \in B) = L(B)$.

(7) 全ての有界で非負の Lipschitz 連続関数 f に対して，$\displaystyle\liminf_{n\to\infty} E_* f(X_n) \ge \int_D f dL$.

(8) $\displaystyle\sup_{f\in \mathrm{BL}_1(D)} |E^* f(X_n) - Ef(X)| \to 0$.

また，$D = \mathbb{R}^k$ のとき，(9) と同値で，(X_n) が Borel 可測であるならば，(10) とも同値.

(9) L の累積分布関数 F の任意の連続点において，$\displaystyle\lim_{n\to\infty} P^*(X_n \le x) = \lim_{n\to\infty} P_*(X_n \le x) = \lim_{n\to\infty} F_n(x) = F(x)$.

(10) $\forall_{t\in\mathbb{R}^k}\ \lim Ee^{it^T X_n} = \displaystyle\int e^{it^T x} dL(x)$.

［証明］.

(1)⇒(7)　$E^*[f(X_n)] \to \displaystyle\int_D f dL$ のとき，$E_*[f(X_n)] \to \displaystyle\int_D f dL$ でもあるから，$\displaystyle\lim_{n\to\infty} \inf_{n\le k} E_*[f(X_k)] = \int_D f dL$.

(2)⇔(3)　補集合を考えることにより，

$$\liminf P_*(X_n \in D \setminus F) \geq L(D \setminus F) \quad \Leftrightarrow \quad \liminf(1 - P^*(X_n \in F)) \geq 1 - L(F)$$

とわかる．

(4)⇔(5)　f を $-f$ と入れ替えることによりわかる．

(7)⇒(2)　任意の $G \overset{\text{open}}{\subset} D$ を取る．$0 \leq f_m \nearrow 1_G \ (m \to \infty)$ を満たす有界な Lipschitz 関数列が取れる．(7) より，$1_G \geq f_m$ について，$\forall_{m\in\mathbb{N}} \ \liminf P_*(X_n \in G) \geq \liminf E_*[f(X_n)] \geq \int f_m dL$ が成り立つ．ここで $m \to \infty$ を考えると，単調収束定理より，(2) を得る．

$P_*[X_n \in G] = P_*(X_n^{-1}(G)) = E_*[1_G(X_n)]$ の読み替えが肝要であった．

(5)⇒(1)　(4)⇔(5) より，任意の上に有界な上半連続関数 f について，

$$E[f] \geq \limsup E^*[f(X_n)] \geq \liminf E_*[f(X_n)] \geq E[f]$$

(2)⇒(4)　f を非負の下半連続関数とすると，これは下に有界で，$G_i := \left\{x \in D \,\middle|\, f(x) > \frac{i}{m}\right\}$ は任意の $i \in \mathbb{N}$ について開集合であり，$D = \cup_{i\in\mathbb{N}} G_i$ である．$f_m(x) := \sum_{i=1}^{m^2} \frac{1}{m} 1_{G_i}(x)$ と定めるとこれは単関数で，

$$f_m(x) = \begin{cases} \frac{i}{m}, & x \in G_i \cap (D \setminus G_{i+1}) = \left\{x \in D \,\middle|\, \frac{i}{m} < f(x) \leq \frac{i+1}{m} \leq m\right\}, \\ m, & x \in G_{m^2} = \{x \in D \mid f(x) > m\}. \end{cases}$$

なので，$0 \leq f_m \leq f \wedge m \leq f$ であり，$\forall_{x\in D\setminus G_{m^2}} \ |f_m(x) - f(x)| \leq \left|\frac{i}{m} - \frac{i+1}{m}\right| = \frac{1}{m}$．よって，$(f_m)$ は f に収束する単関数列である．このとき，

$$\begin{aligned} \liminf_{n\to\infty} E_*[f(X_n)] \geq \liminf_{n\to\infty} E_*[f_m(X_n)] &= \liminf_{n\to\infty} E_*\left[\sum_{i=1}^{m^2} \frac{1}{m} 1_{G_i}(X_n)\right] \\ &\geq \sum_{i=1}^{m^2} \frac{1}{m}\left[\liminf_{n\to\infty} P_*(X_n \in G_i)\right] && \liminf \text{ の性質というかなんというか} \\ &\geq \sum_{i=1}^{m^2} \frac{1}{m} L(G_i) = \int f_m dL && \text{(2) の仮定} \end{aligned}$$

と評価できる．ただし，最後の等式は単関数 f_m に関する Lebesgue 積分の定義による．$m \to \infty$ を考えると，非負関数の列 (f_m) の極限 f の可積分性は不明だが，(f_m) が f に収束する限り Fatou の補題により，

$$\int_D f dL \leq \liminf_{m\to\infty} \int_D f_m dL \leq \limsup_{m\to\infty} \int_D f_m dL \leq \int_D f dL$$

であるから，$\liminf_{m\to\infty} \int_D f_m dL = \limsup_{m\to\infty} = \int_D f dL$ が従い，(4) の主張を得る．[†8]一般の値を取る下に有界な f についても同様.

(2)⇒(6)　(2),(3) より，任意の Borel 集合について，

$$L(B^\circ) \leq \liminf_{n\to\infty} P_*(X_n \in B^\circ) \leq \limsup_{n\to\infty} P^*(X_n \in \overline{B}) \leq L(\overline{B})$$

が成り立つ．$L(\partial B) = 0$ のとき，等号成立.

(6)⇒(3)　任意の閉集合 $F \subset D$ をとる．$F^\epsilon := \{x \in D \mid d(x, F) < \epsilon\} \ (\epsilon > 0)$ と定めると，$\epsilon_1 \neq \epsilon_2 \Rightarrow \partial F^{\epsilon_1} \cap \partial F^{\epsilon_2} = \emptyset$ だから，$E := \{\epsilon \in \mathbb{R}_{>0} \mid L(\partial F^\epsilon) > 0\}$ は可算集合．実際，これが非可算だとすれば，$F := \mathrm{Im}\, L(\partial F^\epsilon) \subset (0, 1] \ (\epsilon \in E)$ を有限に分割した区間 $(i/n, (i+1)/n]$ のいずれかに無限個の点を持つ（さもなくば E が非可算無限であることに矛盾）．その部分の和を取ると，∞ に発散するため，$L(D) = 1$ に矛盾．よって，$L(\partial F^{\epsilon_m}) = 0$ を満たす $\mathbb{R}_{>0}$ の単調減少列 $(\epsilon_m)_{m\in\mathbb{R}}$ が取れる．このとき，(6) より，

$$\limsup_{n\to\infty} P^*(X_n \in F) \leq \limsup_{n\to\infty} P^*(X_n \in \overline{F^{\epsilon_m}}) = L(\overline{F^{\epsilon_m}})$$

が成り立つ．ここで，$m \to \infty$ を考えると，$\cap_{m\in\mathbb{N}} \overline{F^{\epsilon_m}} = F$ より，(3) が従う．

[†8] f の可積分性がわからないので，Lebesgue の優収束定理が使えない．

(6)⇒(9)　特別な場合．

■

要諦 6.2.36. 全て Fatou の補題ちっくな特徴づけ．

(2),(3)　やはり Borel クラスを議論しているので，位相の言葉で特徴付けると良い．

(4),(5)　下に有界な下半連続関数について $\int_D f dL \le \liminf E_*[f(X_n)]$，上に有界な上半連続関数について $\limsup E^*[f(X_n)] \le \int_D f dL$．確かに Fatou の補題の弱化に見える．半連続関数に注目するのは，その場合に $D = \cup_{i\in\mathbb{N}} G_i$ という標準的な開集合の合併による表現が使え，その上に定義された単関数 f_m は．$D \setminus G_{m^2}$ 上で $f - f_m$ は $1/m$ で一様に抑えられ，$m \to \infty$ に対して $G_{m^2} \to D$ となる．

(6)　これも (2),(3) の特徴付け．

(7)　これは弱収束の定義の「$f \in \mathcal{B}(D)$」を「$f \in \mathrm{BL}(D)$」に強めている．開集合 G の特性関数 1_G（ちなみに下半連続）に下から収束する非負値で $\overline{G}$ を台とする有界な Lipschitz 関数の列 (f_m) を用いた，(2) へ帰着させる補題 6.2.33 の証明抽出．

(8)

(9)　Prokhorov の定理による．

(10)　特性関数は分布を特徴付ける．Glibenko の定理という確率論の話．

補題 6.2.37. 部分距離空間 $D_0 \subset D$ について，$X, X_n : \mathcal{X} \to D$ を写像とする．この時，(X_n) が D_0 への写像として X に弱収束することと，D への写像として弱収束することとは同値．

［証明］. D_0 は位相空間として D の部分空間であるから，D_0 の開集合と D の部分空間としての開集合は一致するため，portmanteau 定理 6.2.35(2) からわかる． ■

例 6.2.38. $\mathbb{R}$ 上の Dirac 測度 $\epsilon_{1/n}$ は，ϵ_0 に弱収束するが，各点収束はしない．例えば $A := (0,1] \in \mathcal{B}_1$ 上では，$\forall_{n\in\mathbb{N}}\ \epsilon_{1/n}(A) = 1$ である．

6.2.5　弱収束の例：経験過程は Brownian bridge に収束する

Glivenko-Cantelli の定理は関数 F の推定量としての一致性 (consistency) を示している．ではどこに収束するか？経験分布関数 F_n の収束先は F-Brownian bridge と呼ばれ，このときの収束は「一般の関数の可測関数への弱収束」として定義できる．「経験分布関数の理論では，この収束と連続写像定理を組み合わせることで，種々の統計量の漸近分布が調べられている．」[2]

当然であるが，経験分布関数 $\mathbb{F}_n$ は真の分布 F の自然な推定量であるから，適合度検定で使われる．

例 6.2.39 (empirical distribution function, empirical process, Brownian bridge / pinned Brownian motion).

(1) 可測空間 $(\overline{\mathbb{R}}, \mathcal{B}(\overline{\mathbb{R}}))$ 上の Borel 確率測度 P に従う独立な $\overline{\mathbb{R}}$-値確率変数を $X_1, \cdots, X_n$ とする．この観測値が定める**経験分布関数** $\mathbb{F}_n : \overline{\mathbb{R}} \to [0,1]$ とは，$\mathbb{P}_n(Y) := \#|\{X_i \in (\{X_1,\cdots,X_n\}) \mid X_i \in Y\}|$ $(Y \in \mathcal{B}(\overline{\mathbb{R}}))$ と定めた経験測度 $\mathbb{P}_n : \mathcal{B}(\overline{\mathbb{R}}) \to [0,1]$ の，定義関数の族 $\mathcal{F} := \{\chi_{(-\infty,t]} \in \mathrm{Meas}(\overline{\mathbb{R}}, \overline{\mathbb{R}}) \mid t \in \overline{\mathbb{R}}\} \simeq_{\mathrm{Set}} \overline{\mathbb{R}}$[†9]への制限として定まる：$\mathbb{F}_n(t) := \int_{\overline{\mathbb{R}}} 1_{(-\infty,t]} d\mathbb{P}_n = \mathbb{P}_n((-\infty,t]) = \frac{1}{n}\#|\{i \in [n] \mid X_i \le t\}|$.

(2) 中心化・正規化された経験測度 $\mathbb{G}_n := \sqrt{n}(\mathbb{P}_n - P) : \mathcal{B}(\overline{\mathbb{R}}) \to \mathbb{R}$ についても同様のことを考える．

  (a) これは $A \in \mathcal{B}(\overline{\mathbb{R}})$ を固定したときの値 $\mathbb{G}_n(A) \in \mathbb{R}$ の $n \to \infty$ のときの弱収束極限は，中心極限定理により正規分布 $N(0, P(A)(1-P(A)))$ に従う．実際，二項分布 $B(n, P(A))$ に独立に従う確率変数 $Y_1, \cdots, Y_n$ について $E[Y_1 + \cdots + Y_n] =$

[†9] これは Borel クラスを生成する「基底」である．実はこれが Donsker クラスである，という理論が展開される．もしかしてこれは「σ-代数の基底」の理論なのか？また，Set 上の消息に引き戻す理論が経験過程論ってことかもしれない．

$$nP(A), \mathbb{P}_n(A) = \frac{Y_1 + \cdots + Y_n}{n}$$ であるから，

$$\mathbb{G}_n(A) = \sqrt{n}(\mathbb{P}_n(A) - P(A)) = \sqrt{n}\left(\frac{Y_1 + \cdots + Y_n}{n} - P(A)\right) = \frac{1}{\sqrt{n}}\sum_{i=1}^n (Y_i - P(A)).$$

このとき定まる族 $(\mathbb{G}_n(A)) : \mathcal{A} \times \mathbb{N} \to \mathbb{R}$ を，**可測集合の族 $\mathcal{A} \subset \mathcal{B}(\overline{\mathbb{R}})$ 上の経験過程**という．

(b) 一方で，この測度 $\mathbb{G}_n : \mathcal{B}(\overline{\mathbb{R}}) \to [0, 1]$ を，これが積分によって定める線型汎函数 $\mathbb{G}_n : \mathrm{Meas}(\overline{\mathbb{R}}, \mathbb{R}) \supset \mathcal{F} \to \overline{\mathbb{R}}$ と見る．すると，$f \in \mathcal{F}$ を固定したとき，$\mathbb{G}_n f = \sqrt{n}(\mathbb{P}_n - P)f \in \mathbb{R}$ の $n \to \infty$ のときの弱収束極限は，正規分布 $N(0, E(f - Ef)^2)$ に従う．このとき定まる族 $(\mathbb{G}_n f) : \mathcal{F} \times \mathbb{N} \to \mathbb{R}$ を，**可測関数の族 $\mathcal{F}$ 上の経験過程**という．

(c) 今回の例では $\mathcal{F}$ は Donsker クラスであり，F を（真の）確率測度 P の分布関数 $F(t) := P((-\infty, t])$ とすると，$\mathbb{G}_n : \mathcal{F} \simeq_{\mathrm{Set}} [-\infty, \infty] \to [0, 1]; t \mapsto \sqrt{n}(\mathbb{F}_n - F)(t)$ の極限過程 $\mathbb{G}_F : [-\infty, \infty] \to \mathbb{R}$ が存在して，**F-Brown 橋／固定端 Brown 運動**と呼ばれる．

(d) 2 つの見方は双対的なもので，大した違いはない．基本的に (b) を採用する．

(3) こうして，経験過程 $\mathbb{N} \to D(\overline{\mathbb{R}}); n \mapsto \mathbb{G}_n(t)$ が定まった．**このときの値が関数空間 $D(\overline{\mathbb{R}})$ の大きすぎる σ-加法族に関して可測とは限らないことに注意** 6.2.7．これについて，Brownian bridge への収束 $\mathbb{G}_n \to \mathbb{G}_F$ を弱収束として定義できたことになる．あとは，そのための $\mathcal{F}$ の十分条件を考えたい．

例 6.2.40 (適合度検定における応用)**.** $\mathbb{F}_n$ と F の差を評価する測度には

Kolmogorov-Smirnov 統計量　$\sqrt{n}\|\mathbb{F}_n - F\|_\infty = \|\mathbb{G}_n\|_\infty$
Cramér-von Mises 統計量　$n\int(\mathbb{F}_n - F)^2 dF = \int \mathbb{G}_n^2 dF$

などがある．$\|\cdot\|_\infty : D([-\infty, \infty]) \to [0, \infty]$ は連続であり，いずれも経験過程 $\mathbb{G}_n$ の連続関数になっているから，$\mathbb{G}_n$ の極限からこれらの統計量の極限分布が，連続写像定理 6.2.42 からわかる．

命題 6.2.41. $\lambda : \mathcal{B}([0, 1]) \to [0, 1]$ を一様分布とする．F-Brownian bridge $\mathbb{G}_F$ について，$\mathbb{G}_F = \mathbb{G}_\lambda \circ F$ である．

6.2.6　連続写像定理

連続写像は，確率変数の弱収束も，一般の写像の Borel 可測関数への弱収束も保つ．（したがって特に，一般の確率変数の弱収束，確率収束，概収束も保つ）．どうしてこんなものが作れたのか．図式の変換を定める射は「殆ど至る所連続な写像」が適格，ということだろうか．

定理 6.2.42 (continuous map theorem (Henry Mann, Abraham Wald 43))**.** D, E を距離空間とし，$g : D \to E$ を部分集合 $D_0 \subset D$ 上で連続で，不連続点 D_g が零な写像とする．この時，写像 $(X_n : \Omega_n \to D)$ が Borel 可測関数 $X : \Omega \to D$ に弱収束し，$\mathrm{Im}\, X \subset D_0$ ならば，$(g(X_n))$ も $g(X)$ に弱収束する．

$$\begin{tikzcd} D \arrow[r, "g"] & E \\ \Omega \arrow[u, "X_n"] \arrow[ur, dotted, "g \circ X_n"'] & \end{tikzcd}$$

［証明］.

方針　任意の閉集合 $F \subset E$ について，$\limsup_{n\to\infty} P^*(g(X_n) \in F) \le P(g(X) \in F)$ を示せば良い (Portmanteau 定理 6.2.35(3))．いま，$g^{-1}(F) \subset \overline{g^{-1}(F)}$ で，$\overline{g^{-1}(F)}$ は閉集合だから，Portmanteau 定理 6.2.35(3) より，

$$\limsup_{n\to\infty} P^*(g(X_n) \in F) \le \limsup_{n\to\infty} P^*(X_n \in \overline{g^{-1}(F)}) \le P(X \in \overline{g^{-1}(F)}).$$

実は，$P(X \in \overline{g^{-1}(F)}) = P(g(X) \in F)$ である．まず，$\overline{g^{-1}(F)} \subset g^{-1}(F) \cup D_g$ であるが，これと $L(D_g) = 0$ を併せると従う．任意の $g^{-1}(F)$ の列 $\{x_n\}$ の収束先 $x \in D$ が，g の連続点である場合は $x \in g^{-1}(F)$ で，連続点でない場合は $x \in D_g$ である．

■

注 6.2.43. 証明を簡単にするため，「不連続点 D_g が零集合である」という仮定を含めたが，この部分がなくとも，D_g が Borel 集合である（したがって可測である）ことと，測度が 0 であることが導ける．

6.2.7　相対コンパクト性の特徴付け：Prokhorov の定理

確率変数 $X:\Omega\to D$ がタイトとは，D のコンパクト集合の列 (K_n) であって，full set に収束するものが存在することをいう．一様にタイトとは，D のコンパクト集合の列 (K_n) であって，全ての写像 X_n について $P(X_n^{-1}(K))\ge 1-\epsilon$ となるような列が取れることをいう．一様にタイトな Borel 可測関数列は，タイトな Borel 可測関数に弱収束する部分列を持つ．しかしタイト性も Borel 可測性も，多くの応用においては期待できないが，一般の写像列についても，漸近的にさえタイト性と可測性が成り立てば十分であることがわかる．

「タイトな Borel 可測関数」を「広義一様収束」と読み替え，「漸近的にタイトかつ漸近的可測」は「正規族である」ことと読み替えれば，完全に Ascoli-Arzelà の定理と議論がパラレルである．可測性と複素微分可能性が対応して，広義一様収束がタイト性に対応する．いずれもコンパクト集合をうまく使った収束性の弱化が行われている．

Prokhorov の定理は，「収束する部分列が存在する」ことを特徴付けるが，これは，距離空間においてコンパクト性と点列コンパクト性は同値だから，相対コンパクト性として特徴付けられる．いわば，複素解析における Montel の定理のような結果である．

定義 6.2.44 (asymptotically measurable, asymptotically tight, uniformly tight)**.** 写像列 $(X_n:\Omega_n\to D)$ について，

(1) $\forall_{f\in B(D)}\ \lim_{n\to\infty}(E^*[f(X_n)]-E_*[f(X_n)])=0$ が成り立つ時，X_n は**漸近的可測**であるという．

(2) $\forall_{\epsilon>0}\ \exists_{K\overset{\mathrm{cpt}}{\subset}D}\ \forall_{\delta>0}\ \liminf_{n\to\infty}P_*(X_n\in K^\delta)\ge 1-\epsilon$ が成り立つ時，(X_n) は**漸近的緊密**という．
ただし，$K^\delta:=\{x\in D\mid d(x,K)<\delta\}$ とした．

Borel 可測関数の列 $(X_n:\Omega_n\to D)$ について，(X_n) が**一様に緊密**とは，$\forall_{\epsilon>0}\ \exists_{K\overset{\mathrm{cpt}}{\subset}D}\ \forall_{n\ge 1}\ P(X_n\in K)\ge 1-\epsilon$ が成り立つことをいう．

例 6.2.45 (一様にタイトな Borel 可測関数列)**.** $(X_n:D\to\mathbb{R})$ を確率変数列とする．

(1) $L^{X_n}:=N(n,1)$ ならば，X_n は一様にタイトではない．

(2) $L^{X_n}:=N(\mu_n,1)\ (-\infty<a\le\mu_n\le b<\infty)$ ならば，X_n は一様にタイトである．

補題 6.2.46. Borel 可測関数の列 $(X_n:\Omega_n\to D)$ について，

(1) (a)⇒(b) が成り立つ．D が Polish である時，(b)⇒(a) も成り立つ．
  (a) (X_n) は一様にタイトである．
  (b) (X_n) は漸近的にタイトである．

(2) (X_n) が一様にタイトである時，弱収束する部分列が存在する．

［証明］．

(1) (a)⇒(b) は $K\subset K^\delta$ より，$P(X_n\in K)\le P_*(X_n\in K^\delta)$ であるから明らか．

■

命題 6.2.47 (純方向は簡単)**.** $(X_n:\Omega_n\to D)$ を写像列とする．

(1) (X_n) が X に弱収束するならば，(X_n) は漸近的可測である．

(2) (X_n) が X に弱収束するとき，(X_n) が漸近的にタイトであることと X がタイトであることとは同値．

［証明］．

(1) 任意の $f\in B(D)$ について，X_n が X に弱収束するならば，定義より $E^*[f(X_n)]\to E[f(X)]$ である．このとき $E_*[f(X_n)]=-E^*[-f(X_n)]\to -E[-f(X)]=E[f(X)]$ であるから，$E^*[f(X_n)]-E_*[f(X_n)]\to E[f(X)]-E[f(X)]=0$.

(2) $\Rightarrow$ (X_n) を漸近的タイトとすると，任意の $\epsilon > 0$ に対して，コンパクト集合 $K \subset D$ が存在して，$\forall_{\delta>0} \liminf_{n\to\infty} P_*(X_n \in K^\delta) \ge 1-\epsilon$．portmanteau 定理 6.2.35(3) より，

$$P(X \in \overline{K^\delta}) \ge \limsup_{n\to\infty} P^*(X_n \in \overline{K^\delta}) \ge \liminf_{n\to\infty} P_*(X_n \in \overline{K^\delta}) \ge 1-\epsilon.$$

$\delta \searrow 0$ を考えて，$P(X \in K) \ge 1-\epsilon$ を得る．

$\Leftarrow$ X をタイトとすると，任意の $\epsilon > 0$ に対して，コンパクト集合 $K \subset D$ が存在して，$P(X \in K) \ge 1-\epsilon$．開集合 $K^\delta \overset{\text{open}}{\subset} D$ に対して，portmanteau 定理 6.2.35(2) より，

$$\forall_{\delta>0} \liminf_{n\to\infty} P_*(X_n \in K^\delta) \ge P(X \in K^\delta) \ge 1-\epsilon.$$

■

系 6.2.48. 写像の列 $(X_n : \Omega_n \to D)$ が可分な $X : \Omega \to D$ に弱収束するならば，X_n を完備拡大 $\widetilde{D}$ への写像列と見た時，漸近的にタイトである．

定理 6.2.49 (Prokhorov)**.** $(X_n : \Omega_n \to D)$ を写像列とする．(1)$\Rightarrow$(2) である．

(1) (X_n) は漸近的にタイトかつ漸近的に可測である．

(2) (X_n) の部分列で，タイトな Borel 確率測度 L に弱収束するものが存在する．

命題 6.2.50 (連続写像は漸近的性質を保つ)**.** $(X_n : \Omega \to D)$ を写像の列，$g : D \to E$ を連続写像とする．

(1) (X_n) が漸近的タイトならば，$(g(X_n))$ も漸近的タイトである．

(2) (X_n) が漸近的可測ならば，$(g(X_n))$ も漸近的可測である．

［証明］．

(1) (X_n) は漸近的にタイトだから，$\forall_{\epsilon>0} \exists_{K \overset{\text{cpt}}{\subset} D} \forall_{\eta>0} \liminf_{n\to\infty} P_*(X_n \in K^\eta) \le 1-\epsilon$．ここで，$g$ が連続であることより，任意の $\delta > 0$ について，$K^{\eta(\delta)} \subset g^{-1}(g(K)^\delta)$ を満たす $\eta(\delta) > 0$ が取れるから，

$$\liminf_{n\to\infty} P_*(g(X_n) \in g(K)^\delta) \ge \liminf_{n\to\infty} P_*(X_n \in K^{\eta(\delta)}) \ge 1-\epsilon$$

が従う．こうして，条件を満たすコンパクト集合の族 $(g(K))_{\delta>0}$ が見つかった．

(2) 任意の $f \in B(E)$ に対して，$f \circ g \in B(D)$ より，(X_n) の漸近的可測性から

$$E^*[f(g(X_n))] - E_*[f(g(X_n))] \xrightarrow{n\to\infty} 0.$$

これは $g(X_n)$ の漸近的可測性を表している．

■

補題 6.2.51 (漸近的可測性の十分条件)**.** $\mathcal{F} \subset B(D)$ を D の各点を分離する代数とする．写像列 (X_n) が次の条件 (#) を満たすならば，漸近的可測である．

(#) (X_n) は漸近的タイトで，$\forall_{f\in\mathcal{F}} E^*[f(X_n)] - E_*[f(X_n)] \to 0$.

要諦 6.2.52. これは Stone-Weierstrass の近似定理による．まさに複素関数での議論に同じである．

Prokhorov の定理まとめ

「相対コンパクト」と言った時など，収束先がタイトな Borel 確率測度であることを暗黙の了解とする．すると，Prokhorov の定理は「弱収束するならば漸近的にタイトかつ漸近的に可測」の逆の成立を主張していると思える．さらに一歩踏み込んで，写像の族 $\mathcal{F} \subset \mathrm{Map}(\Omega, D)$ が相対コンパクトであることの十分条件は，漸近的にタイトであることと，ある $B(D)$ の部分代数について「漸近的可測」であれば良い．

6.2.8　収束性の遺伝：Slutsky の定理

Prokhorov の定理で迂回できる

2 つの確率変数が弱収束し，片方が定数になるならば，それらの積写像も弱収束し，また連続写像定理を組み合わせると，弱収束する確率変数から標準的な構成を行なったものは基本的に全て弱収束することがわかる．これを Slutsky の定理といい，これは一般の写像の弱収束についても成り立つ．
片方が定数にならない場合は同様な結果は弱収束では成り立たないが，確率収束では成り立つ．なお，定数に弱収束するならば，定数に確率収束する（補題 6.2.60）．

補題 6.2.53 (直積距離空間)**.** $(D,d),(E,e)$ を距離空間とする．

(1) $\rho((x_1,y_1),(x_2,y_2))=d(x_1,x_2)\vee e(y_1,y_2),\sqrt{d(x_1,x_2)^2+e(y_1,y_2)^2},d(x_1,x_2)+e(y_1,y_2)$ はいずれも距離関数で，$D\times E$ に直積位相を誘導する．
(2) $D\times E$ を直積距離空間とする．一般に $\mathcal{B}(D)\times\mathcal{B}(E)\subset\mathcal{B}(D\times E)$ であり，D,E がいずれも可分であるならば逆も成り立つ．
(3) $(X,Y):\Omega\to D\times E$ は Borel σ-加法族の積について常に可測である．したがって，D,E が可分でない場合は，可測関数の積写像 $(X,Y):\Omega\to D\times E$ は直積距離位相の Borel σ-加法族 $\mathcal{B}(D\times E)$ については必ずしも可測とは限らない．
(4) コンパクト集合 $K_1\subset D,K_2\subset E$ と距離関数 ρ について，$\forall_{\delta>0}\;(K_1\times K_2)^\delta=K_1^\delta\times K_2^\delta$ である．

［証明］．

(1) a
(2) a
(3) a
(4)

$$\rho((x,y),K_1\times K_2)=d(x,K_1)\vee e(y,K_2)<\delta\quad\Leftrightarrow\quad d(x,K_1)<\delta\wedge e(y,K_2)<\delta$$

であるから．

■

補題 6.2.54 (結合写像の漸近的タイト性)**.** $(X_n:\Omega_n\to D),(Y_n:\Omega_n\to E)$ を写像列とする．

(1) $(X_n),(Y_n)$ が共に漸近的タイトであることと，積写像 (X_n,Y_n) が漸近的タイトであることは同値.
(2) $(X_n),(Y_n)$ が漸近的にタイトであるとき，共に漸近的可測であることと，積写像 (X_n,Y_n) が漸近的可測であることは同値.

［証明］．

(1) 直積距離空間の性質 6.2.53(4) より $(K_1\times K_2)^\delta=K_1^\delta\times K_2^\delta$ であるから，6.2.18(4) より，

$$P_*((X_n,Y_n)\in(K_1\times K_2)^\delta)=P_*((X_n,Y_n)\in K_1^\delta\times K_2^\delta)=P_*(X_n^{-1}(K_1^\delta)\cap Y_b^{-1}(K_2^\delta))\le P_*(X_n\in K_1^\delta)+P_*(Y_n\in K_2^\delta)-1.$$

また，$K^\delta\subset\mathrm{pr}_1(K)^\delta\times\mathrm{pr}_2(K)^\delta$ より，

$$P_*((X_n,Y_n)\in K^\delta)\le P_*(X_n^{-1}(\mathrm{pr}_1(K)^\delta)\cap Y_n^{-1}(\mathrm{pr}_2(K)^\delta))\le P_*(X_n\in\mathrm{pr}_1(K)^\delta)\wedge P_*(Y_n\in\mathrm{pr}_2(K)^\delta).$$

$(X_n),(Y_n)$ が共に漸近的タイトである時，一式目よりコンパクト集合の族 $(K_1\times K_2)_{\delta>0}$ が見つかる．(X_n,Y_n) が共に漸近的タイトである時，二式目よりコンパクト集合の族 $(\mathrm{pr}_1(K))_{\delta>0},(\mathrm{pr}_2(K))_{\delta>0}$ がそれぞれ見つかる．
(2) $\Leftarrow$　任意の $f\in B(D)$ について，$f\circ\mathrm{pr}_i\in B(D\times E)$ $(i=1,2)$ より，

$$E^*[f(\mathrm{pr}_i(X_n,Y_n))]-E_*[f(\mathrm{pr}_i(X_n,Y_n))]\xrightarrow{n\to\infty}0$$

だから，$(X_n),(Y_n)$ の漸近的可測性を得る．
$\Rightarrow$　?

■

定理 6.2.55 (Slutsky). 可分な Borel 可測写像 $X:\Omega\to D$ と，任意の $c\in E$ について，写像列 $(X_n:\Omega\to D),(Y_n:\Omega\to E)$ が $X_n\Rightarrow X, Y_n\Rightarrow c$ を満たすとする．

(1) $(X_n,Y_n)\Rightarrow(X,c)$ である．
(2) $D=E$ を線型空間とする．$X_n+Y_n\Rightarrow X+c$ である．
(3) Y_n が定数ならば，$Y_nX_n\Rightarrow cX$ である．また $c\ne 0$ ならば $X_n/Y_n\Rightarrow X/c$ である．

［証明］．

(1)　X の終域 D を完備化 $\overline{D}$ に埋め込むことにより，可分な X は緊密でもあるとして一般性を失わない 6.2.26．X,c はいずれもタイトだから，これに弱収束する $(X_n),(Y_n)$ はいずれも漸近的にタイトかつ漸近的に可測で (Prokhorov の逆 6.2.47)，したがって (X_n,Y_n) も漸近的にタイトかつ漸近的に可測 6.2.54．よって，Prokhorov の定理 6.2.49（を繰り返し適用することに）より，族 $\{X_n,Y_n\}$ は「相対コンパクト」（任意の $\{X_n,Y_n\}$ の列はタイトな確率要素に収束する部分列を持つ）．しかし，収束する部分列は要素ごとに見ると，結局収束先は (X,c) であることが必要．特に (1) が成り立つ．

(2),(3)　線型空間の演算 $+,\cdot,/$ などの連続写像について，連続写像定理 6.2.42 より．

■

系 6.2.56. Donsker クラスは，Glivenko-Cantelli in probability である．

要諦 6.2.57. 実際は，almost surely に成り立つ．

6.2.9　確率収束と概収束

弱収束と同様に，確率収束と概収束も非可測な場合に拡張できる．また，表現という概念を用いれば，弱収束するネットは，表現の違いを除いて（確率空間は違うかもしれないが法則としては一致する確率変数が存在して）概収束する．

定義 6.2.58 (convergence in outer probability, convergence outer almost surely). $(X_n:\Omega\to D), X:\Omega\to D$ を写像とする．

(1) (X_n) が X に**外確率収束**する ($X_n\xrightarrow{P^*}X$) とは，次が成り立つことをいう：

$$\forall_{\epsilon>0}\quad \lim_{n\to\infty}P(d(X_n,X)^*>\epsilon)=\lim_{n\to\infty}P^*(d(X_n,X)>\epsilon)=0.$$

(2) (X_n) が X に**外概収束**する ($X_n\xrightarrow{\text{a.s.}^*}X$) とは，次が成り立つことをいう：

$$P\left(\lim_{n\to\infty}d(X_n,X)^*=0\right)=1.$$

補題 6.2.59. $X:\Omega\to D$ を Borel 可測関数，(X_n) を写像の列とする．

(1) (X_n) が外概収束するならば，外確率収束する．
(2) (X_n) が確率収束することは，$\{X_n\}$ が相対コンパクト（$\{X_n\}$ の任意の部分列は X に概収束する部分列を持つ）ことに同値．

補題 6.2.60. $X:\Omega\to D$ を Borel 可測関数，$(X_n),(Y_n)$ を写像の列とする．

(1) (X_n) が X に弱収束し，かつ $d(X_n,Y_n)\xrightarrow{P^*}0$ ならば，(Y_n) も X に弱収束する．
(2) (X_n) が X に外確率収束するならば，X に弱収束する．
(3) (X_n) が定数 c に外確率収束することと，c に弱収束することは同値．

定理 6.2.61 (連続写像定理の拡張). D の部分集合の列 $(D_n)_{n\in\mathbb{N}}$ と写像 $g_n:D_n\to E$ の列は，$\forall_{n\in\mathbb{N}_+}\ \forall_{x\in D_0}\ x_n\in D_n\wedge x_n\to x\Rightarrow g_n(x_n)\to g(x)$ を満たすとする．このとき，写像の列 $(X_n:\Omega\to D_n)$ と可分な Borel 可測関数 $X:\Omega\to D_0$ について，次が成り立つ：

(1) (X_n) が X に弱収束するならば，$g_n(X_n)$ も $g(X)$ に弱収束する．
(2) (X_n) が X に外確率収束するならば，$g_n(X_n)$ も $g(X)$ に外確率収束する．
(3) (X_n) が X に外概収束するならば，$g_n(X_n)$ も $g(X)$ に外概収束する．

要諦 6.2.62. 準備が必要．

6.3　有界関数の空間値確率要素

$l^\infty(T)$-値確率変数がタイトな Gauss 過程に弱収束するための，f.f.d. が弱収束することに加えて必要な条件を割り出す．これは第一義的には「漸近的タイト性」によって明瞭に捉えられるが，これをさらに計算可能な概念に翻訳し，収束先がタイトな Gauss 過程である場合についてさらに特殊化する．

動機付け

$D = \widetilde{B}(T)$ $(T \in \mathrm{Set})$ とした各論（確率過程論）を展開する．確率過程 $X : T \to \mathrm{Meas}(\Omega, \mathbb{R})$ は，見本過程を実現値とする「確率変数」$X : \Omega \to \mathrm{Map}(T, \mathbb{R})$ とも見れる．始域の追加（２変数関数化）という意味でも，終域の関数空間化という意味でも，実確率変数の拡張でもある．この値域 $\mathrm{Im}\, X$ に種々の制限をつけて考えることになるのだが，特に $\mathrm{Im}\, X \subset \widetilde{B}(T)$ は，一様ノルムの構造が入るので，重要な仮定である．
$T = 2$ の場合が結合変数であるが，その場合でも失敗するのであるから，確率過程は扱いが難しい．

問題意識

タイトな Gauss 過程は，最低限の一様連続性を持つため，非常に大事なクラスとなる．このクラスに弱収束するための必要十分条件 6.3.13 を探ることを考える．条件 (2) の漸近的タイト性が厄介な課題となり，経験過程論の発展の重要な動機の一つであってきた．

6.3.1　有界関数の空間での弱収束

見本過程とは評価写像 ev_ω による確率過程の像である．これに注目するという技法，評価写像が連続であることを用いる論法も，複素解析と少し似ている．証明においては，「有限個の点で評価すると必ず有界連続であるようなクラス $\mathcal{F} \subset B(\widetilde{B}(T))$」が Riesz 空間になることに注目する．すると，$l^\infty(T)$ 上の確率分布は，任意の有限周辺分布が一致するならば，分布全体も一致する．

定義 6.3.1 (stochastic process, sample path).

(1) **確率過程**とは，確率変数 $\mathrm{Meas}(\Omega, \mathbb{R})$ の族 $X := (X_t)_{t\in T} : T \to \mathrm{Meas}(\Omega, \mathbb{R})$ をいう．
(2) 確率過程は，同じ確率空間上の Banach 空間値確率変数 $\Omega \to \mathrm{Map}(T, \mathbb{R})$ を定める．このときの実現値の各々を**見本過程**という．$X(\omega) : T \to \mathbb{R}$ は特に可測とも有界とも限らないことに注意．

例 6.3.2. 経験過程 $(\mathbb{G}_n f)_{f\in\mathcal{F}} : \mathcal{F} \to \mathrm{Meas}(\mathbb{N}, \mathbb{R})$ は，$\forall_{x\in\mathcal{X}}\ \sup_{f\in\mathcal{F}} |f(x) - Pf| < \infty$ を満たすとき，確率変数 $\mathbb{G}_n : \mathbb{N} \to l^\infty(\mathcal{F})$ を定める．

定義 6.3.3 (確率過程の周辺分布, version / expression). 評価写像 $\mathrm{ev}_{(t_1,\cdots,t_k)} : \widetilde{B}(T) \to \mathbb{R}^k$ を $f \in \widetilde{B}(T)$ に対して

$$\mathrm{ev}_{(t_1,\cdots,t_k)}(f) = (f(t_1), \cdots, f(t_k)) \in \mathbb{R}^k$$

と定めると，これは連続である．この連続写像に注目すると，

(1) 見本過程へのランダム関数とみた確率過程 $X : \Omega \to \widetilde{B}(T)$ が Borel 可測ならば，$\mathrm{ev}_{(t_1,\cdots,t_k)} \circ X : \Omega \to \mathbb{R}^k$ も確率変数であるから，有限な周辺分布 $L^{(X_{t_1},\cdots,X_{t_k})}$ が $L \circ (\mathrm{ev}_{(t_1,\cdots,t_k)} \circ X)^*$ と与えられる．
(2) ２つの確率過程 $X : T \to \mathrm{Meas}(\Omega, \mathbb{R}), Y : T \to \mathrm{Meas}(\Omega', \mathbb{R})$ について，一方が他方の**バージョン**または**表現**であるとは，対応する全ての有限次元周辺分布が一致することをいう：$\forall_{k\in\mathbb{N}}\ \forall_{\{t_1,\cdots,t_k\}\subset T}\ L^{(X_{t_1},\cdots,X_{t_k})} = L^{(Y_{t_1},\cdots,t_k)}$．これは２つの異なる確率空間 Ω, Ω' 上の確率変数の族を関係付け得る概念である．

補題 6.3.4 (見本過程への注目). 写像の列 $(X_n : \Omega_n \to \widetilde{B}(T))$ を漸近的タイトとする．このとき，次の２条件は同値．

(1) (X_n) は漸近的可測.

(2) 任意の $t \in T$ について，確率変数 $\mathrm{Meas}(\Omega, \mathbb{R})$ の列 $(X_n(t))$ は漸近的可測.

［証明］. 漸近的可測性の特徴付け 6.2.51 による.

$$\mathcal{F} := \left\{ f \in B(\widetilde{B}(T)) \mid \forall_{k \in \mathbb{N}} \forall_{\{t_1,\cdots,t_k\} \subset T} \exists_{g \in B(\mathbb{R}^k)} f = g \circ \mathrm{ev}_{(t_1,\cdots,t_k)} \right\}$$

と定めると，これは代数であり，定数関数を含み，$\widetilde{B}(T)$ の各点を分離するベクトル束である.

(1)⇒(2)　(X_n) を漸近的可測とする：$\forall_{g \in B(\mathbb{R})} E^*[g(X_n)] - E_*[g(X_n)] \xrightarrow{n \to \infty} 0$. $\forall_{g \in B(\mathbb{R})} \forall_{z \in \widetilde{B}(\mathbb{R})} \forall_{t \in T}\ g \circ z(t) = g \circ \mathrm{ev}_t(z)$ という関係が成り立つから，$z = X$ とすると $\mathrm{ev}_t = X(t)$ だから，$E^*[g(X_n(t))] - E_*[g(X_n(t))] \xrightarrow{n \to \infty} 0$ を得る.

(2)⇒(1)　$(X_n(t))$ を漸近的可測とする：$\forall_{t \in T} \forall_{g \in B(\mathbb{R})} E^*[g(X_n(t))] - E_*[g(X_n(t))] \xrightarrow{n \to \infty} 0$. ここで，$\mathrm{ev}_{(t_1,\cdots,t_k)} : \widetilde{B}(T) \to \mathbb{R}^k$ は連続なので，漸近的タイトな写像との合成 $\mathrm{ev}_{(t_1,\cdots,t_k)}(X_n)$ も漸近的タイトである 6.2.50. この下で，各 $(X_n(t_i))$ は漸近的可測との仮定より，その積であるから漸近的可測でもある 6.2.54. $k \in \mathbb{N}$ は任意としたから，結局全ての $f \in \mathcal{F}$ に対して $E^*[f(X_n)] - E_*[f(X_n)] \xrightarrow{n \to \infty} 0$. よって補題 6.2.51 より (X_n) は漸近的可測.

■

要諦 6.3.5. 結合写像の漸近的可測性の特徴付け 6.2.54 の一般化になっている.

補題 6.3.6 (有限な周辺分布への注目). $X : \Omega \to \widetilde{B}(T), Y : \Omega' \to \widetilde{B}(T)$ をタイトな Borel 可測関数とする．このとき，次の 2 条件は同値.

(1) $L^X = L^Y$.

(2) X は Y の表現である：$\forall_{k \in \mathbb{N}} \forall_{\{t_1,\cdots,t_k\} \subset T}\ L^{(X_{t_1},\cdots,X_{t_k})} = L^{(Y_{t_1},\cdots,t_k)}$.

［証明］. (2)⇒(1) を示す. (2) が成り立つ時，全ての $f \in \mathcal{F}$ に対しては,

$$\int_{\widetilde{B}(T)} f dL^X = \int_{\widetilde{B}(T)} g dL^{(X(t_1),\cdots,X(t_k))} = \int_{\widetilde{B}(T)} g dL^{(Y(t_1),\cdots,Y(t_k))} = \int_{\widetilde{B}(T)} f dL^Y$$

が成り立つ．よって，分布が一致することの特徴付け 6.2.33 の (4) より，$L^X = L^Y$. ■

要諦 6.3.7. 収束先の同値類は「バージョン」である，とわかった.

6.3.2　弱収束の特徴付け：漸近的タイト性

関数空間 $l^\infty(T)$ 上での弱収束の必要十分条件として，f.d.d. の収束に加えて必要な条件は「漸近的タイト性」だと判明した．そこでこの概念に集中するが，これはまだまだ複雑である.

定理 6.3.8 (タイトな確率過程に弱収束する条件の特徴付け 1). $(X_n : \Omega_n \to \widetilde{B}(T))$ を写像の列とする.

(1) 次の 2 条件は同値.
   (a) (X_n) は $\widetilde{B}(T)$ におけるタイトな Borel 確率測度に弱収束する.
   (b) (X_n) は漸近的タイトで，かつ，全ての有限な周辺写像 $(X_n(t_1), \cdots, X_n(t_k))$ が弱収束する.

(2) (X_n) は漸近的タイトで，任意の有限な周辺写像 $(X_n(t_1), \cdots, X_n(t_k))$ がある確率過程 $X : T \to \mathrm{Meas}(\Omega, \mathbb{R})$ の周辺確率ベクトル $(X(t_1), \cdots, X(t_k))$ に弱収束するとする．このとき，X のバージョンで，$\widetilde{B}(T)$ に属する見本過程を持つような $\widetilde{X}$ が存在し，$X_n \Rightarrow \widetilde{X}$ が成り立つ.

［証明］.

(1) (a)⇒(b)　漸近的タイト性は Prokhorov の定理の逆 6.2.47 から．有限な周辺写像の収束は $\mathrm{ev}_{(t_1,\cdots,t_k)}$ に関する連続写像定理 6.2.42 による.

(b)⇒(a)　$(X_n(t))$ が弱収束する時，Prokhorov の定理の逆 6.2.47 より，漸近的可測である．(X_n) は漸近的可測だから，補題 6.3.4 より，各 $(X_n(t))$ が漸近的可測であることは (X_n) が漸近的可測であることに同値．よって，Prokhorov の定理

6.2.49 より，$\{X_n\}$ は相対コンパクトである．これが弱収束すると示すには，任意の収束する部分列の極限が一致することを示せば良い．

$\{X_n\}$ の列 $(X_m),(X_l)$ の部分列がそれぞれ Borel 確率測度 L,G に弱収束するとする．連続写像定理から，$\{t_1,\cdots,t_k\}\subset T$ に関して，$(X_{m'}(t_1),\cdots,X_{m'}(t_k)),(X_{l'}(t_1),\cdots,X_{l'}(t_k))$ も弱収束し，その極限は必然的に L と G であるが，仮定から $(X_n(t_1),\cdots,X_n(t_k))$ が弱収束しているから，$L\circ \mathrm{ev}^*_{(t_1,\cdots,t_k)}=G\circ \mathrm{ev}^*_{(t_1,\cdots,t_k)}$ が必要．補題 6.3.6 より，これは $L=G$ と同値．

(2) (1) より，(X_n) は $\widetilde{B}(T)$ 上でタイトな Borel 確率測度 L に弱収束する．この時，$\widetilde{X}:=\mathrm{id}_{\widetilde{B}(T)}$ とおくと，$L^{\widetilde{X}}=L$ となり，$\widetilde{X}$ は $\widetilde{B}(T)$ 上に見本過程を持つ．この時，当然 $\widetilde{X}$ は X のバージョンである．

■

要諦 6.3.9. この構成だと $\widetilde{X}:\widetilde{B}(T)\to\widetilde{B}(T)$ は確率要素であって確率過程ではなくない？

6.3.3　漸近的タイト性の連続度による特徴付け

漸近的タイト性は「連続度が 0 に確率収束する」ことに同値

前節で，弱収束を示すには，漸近的タイト性と有限周辺分布の収束を示せば良いと解った．周辺写像の収束の議論は，Euclid 空間上の弱収束の証明のためのテクニックを用いて議論できるから，最終的に一般の確率過程の弱収束の特徴付け 6.3.8 は Ascoli-Arzelà の定理と似たステートメント 6.3.13 に変形できる．

T 上の確率場 X_n が漸近的タイトであることを特徴付けると，各成分 $X_n(t)$ が漸近的タイトであることと，場 (T,ρ) が全有界になることと，X_n の「連続度の 0 への確率収束」が必要になる．

定義 6.3.10 (asymptotically uniformly ρ-equicontinuous in probability)**.** ρ を T 上の準距離とする．写像列 $X_n:\Omega_n\to\widetilde{B}(T)$ が任意の $\epsilon,\eta>0$ に対して $\delta>0$ が存在して

$$\limsup_{n\to\infty} P^*\left(\sup_{\rho(s,t)<\delta}|X_n(s)-X_n(t)|>\epsilon\right)<\eta$$

を満たすとき，(X_n) は**漸近的 ρ-同程度一様確率連続**という．

定理 6.3.11 (漸近的タイト性の特徴付け)**.** $(X_n:\Omega_n\to\widetilde{B}(T))_{n\in\mathbb{N}}$ を任意の写像列とする．このとき，次の 3 条件は同値である．

(1) (X_n) は漸近的タイトである．

(2) 任意の $t\in T$ に対して $(X_n(t):\Omega_n\to\mathbb{R})$ は漸近的タイトで，任意の $\epsilon,\eta>0$ に対して T の有限分割 $T=\cup_{i=1}^k T_i$ が存在して

$$\limsup_{n\to\infty} P^*\left(\sup_{1\le i\le k}\sup_{s,t\in T_i}|X_n(s)-X_n(t)|>\epsilon\right)<\eta$$

が成り立つ．

(3) 任意の $t\in T$ に対して $(X_n(t):\Omega_n\to\mathbb{R})$ は漸近的タイトで，T 上の準距離 ρ が存在して (T,ρ) が全有界となり，X_n は漸近的 ρ-同程度一様確率連続となる．

［証明］．

(1)⇒(3)　(X_n) を漸近的タイトとすると，$X_n(t)=\mathrm{ev}_t(X_n)$ も漸近的タイトである 6.2.50.

構成 1　また，(X_n) は漸近的タイトだから，$\widetilde{B}(T)$ のコンパクト集合の列 $K_1\subset K_2\subset\cdots$ であって，$\forall_{\epsilon>0}\ \liminf_{n\to\infty} P_*(X_n\in K_m^\epsilon)\ge 1-1/m$ を満たすものが存在するから，これを用いて $\rho_m(s,t):=\sup_{z\in K_m}|z(s)-z(t)|\ (s,t\in T)$ と定めると，これは T 上の擬距離であり，(T,ρ_m) は全有界となる．

構成 2　これを用いて，T 上の準距離 ρ を $\rho(s,t):=\sum_{m=1}^\infty 2^{-m}(\rho_m(s,t)\wedge 1)$ と定めると，(T,ρ) は再び全有界となる．

検証　定め方から，$\forall_{z\in K_m}\ |z(s)-z(t)|\le\rho_m(s,t)$ より，$\forall_{\widetilde{z}\in K_m^\epsilon}\ \forall_{s,t\in T}\ |\widetilde{z}(s)-\widetilde{z}(t)|<2\epsilon+\rho_m(s,t)$．ここで，$\rho_m(s,t)\wedge 1\le$

$2^m\rho(s,t)$ だから，$\forall_{\epsilon\in(0,1)}\ K_m^\epsilon\subset\left\{z\in\widetilde{B}(T)\;\middle|\;\sup_{\rho(s,t)<\delta}|z(s)-z(t)|\le 3\epsilon\right\}$．よって，$\delta<2^{-m}\epsilon$ に対して，

$$\liminf_{n\to\infty}P_*\left(\sup_{\rho(s,t)<\delta}|X_n(s)-X_n(t)|\le 3\epsilon\right)\ge 1-\frac{1}{m}.$$

(3)⇒(2)　(T,ρ) は全有界だから，任意の ϵ,η に対して，T は直径 δ の有限個の球で被覆できる．これらを互いに素にしたものを $T_1,\cdots,T_k$ と定めれば，

$$\limsup_{n\to\infty}P^*\left(\sup_{1\le i\le k}\sup_{s,t\in T_i}|X_n(s)-X_n(t)|>\epsilon\right)<\eta$$

が成り立つ．

(2)⇒(1)　**準備**　$\|X_n\|_T$ は $\mathbb{R}$ の列として漸近的タイトである．実際，任意の有限分割 $(T_i)_{i\in[k]}$ を取ると，ある $t_i\in T_i$ に対して

$$\|X_n\|_T\le\sup_{i\in[k]}\sup_{t\in T_i}|X_n(t)-X_n(t_i)|+\max_{i\in[k]}|X_n(t_i)|$$

が成り立つから，少なくとも $1-\eta$ の内確率で $\|X_n\|_T\le\max_{i\in[k]}|X_n(t_i)|+\epsilon$ と抑えられる．

実写像 $(X_n(t_i))$ は漸近的タイトであるという仮定より，$\max_{i\in[k]}|X_n(t_i)|+\epsilon$ も漸近的タイトであり（連続写像との合成 6.2.50），従って $\|X_n\|_T$ は $\mathbb{R}$ で漸近的タイトである．

K_m の構成　任意の $\zeta>0$ を取る．(ϵ_m) を 0 に収束する任意の単調減少な正実数列とする．$\|X_n\|_T$ の漸近的タイト性を用いて，定数 M を $\limsup_{n\to\infty}P^*(\|X_n\|_T>M)<\zeta$ を満たす実数と定める．仮定から，$\epsilon=\epsilon_m,\eta=2^{-m}\zeta$ に対する有限分割 $(T_i)_{i\in[k]}$ を取る．

各 T_i 上定値で，値域が $\left\{0,\pm\epsilon,\cdots,\pm\left\lfloor\frac{M}{\epsilon_m}\right\rfloor\epsilon_m\right\}$ に含まれる関数の全体を $\{z_1,\cdots,z_{p(m)}\}\subset\widetilde{B}(T)$ とし，これを中心とする半径 ϵ_m の閉球の和を K_m とする．このとき，$M<\left\lfloor\frac{M}{\epsilon_m}\right\rfloor\epsilon_m+\epsilon_m$ という関係に注意すれば，

$$\begin{cases}\|X_n\|_T\le M\\ \sup_{i\in[k]}\sup_{s,t\in T_i}|X_n(s)-X_n(t)|\le\epsilon_m\end{cases}\Rightarrow X_n\in K_m$$

が成り立つ．「各ブロック T_i で，ある定数関数 z_i を選べば，ϵ_m 以上そこから離れることはない」ということを閉球を用いて翻訳した．[†10]

K の構成　$K:=\cap_{m\in\mathbb{N}}K_m$ とおくと，これは閉である．

(a)　また，全有界でもある．

(b)　$\widetilde{B}(T)$ は完備なので，従って K はコンパクトである．

(c)　$\forall_{\delta>0}\ \exists_{n\in\mathbb{N}}\ K^\delta\supset\cap_{i=1}^m K_i$ である．

結論　(c) より $X_n\notin K^\delta\Rightarrow\exists_{m\in\mathbb{N}}\ X_n\notin\cap_{i=1}^m K_i$ より，次のように評価できる．

$$\begin{aligned}\limsup_{n\to\infty}P^*(X_n\notin K^\delta)&\le\limsup_{n\to\infty}P^*\left(X_n\notin\cap_{i=1}^m K_i\right)\\&\le\limsup_{n\to\infty}P^*(\|X_n\|_T>M)+\limsup_{n\to\infty}\sum_{i=1}^m P^*\left(\sup_{i\in[k]}\sup_{s,t\in T_i}|X_n(s)-X_n(t)|>\epsilon_m\right)\\&\qquad\because K_m\text{の構成と劣加法性}\\&\le\zeta+\sum_{i=1}^m 2^{-i}\zeta<2\zeta.\end{aligned}$$

■

要諦 6.3.12. タイト性（ある種の σ-有限性）を（擬）距離の言葉に変換することは，複素領域 $D\subset\mathbb{C}$ から距離空間への連続写像の空間 $C(D,S)$ を調べる際に用いたコンパクト開位相を引き起こす距離と同様である．これらを $1/2^m$ を係数として足し合わせる構成も似ている．「最初の有限個以外が効いてこない距離」という定め方である．(2) は連続度の概念に通じるから (1) に戻りやすい．

[†10] これがほとんどの確率で成り立つので，評価への準備ができている．

系 6.3.13 (タイトな確率過程に弱収束する条件の特徴付け 2)**.** 任意の写像列 $(X_n:\Omega_n\to\widetilde{B}(T))_{n\in\mathbb{N}}$ が $\widetilde{B}(T)$ においてタイトな Borel 可測写像 X に弱収束するための必要十分条件は，(1) かつ (2) である：

(1) 任意の有限部分集合 $\{t_1,\cdots,t_k\}\subset T$ に対して，$(X_n(t_1),\cdots,X_n(t_k))\Rightarrow(X(t_1),\cdots,X(t_k))$.

(2) T が全有界となるような準距離 ρ が存在して，全ての $\epsilon>0$ に対して次が成り立つ：

$$\lim_{\delta\searrow 0}\limsup_{n\to\infty}P^*\left(\sup_{\rho(s,t)<\delta}|X_n(s)-X_n(t)|>\epsilon\right)=0.$$

［証明］. 弱収束することの特徴付け 6.3.8 と，漸近的タイト性の特徴付け 6.3.11 より. ■

6.3.4　収束先の確率過程の持つ一様連続性

収束先であるタイトな Borel 可測写像 $X:\Omega\to\widetilde{B}(T)$ の見本過程はある種の連続性を持つ.

記法 6.3.14 (uniform continuity)**.** 一様に ρ-連続な $z\in\widetilde{B}(T)$ の全体からなる集合を

$$UC(T,\rho):=\left\{z\in\widetilde{B}(T)\mid\lim_{\delta\searrow 0}\sup_{\rho(s,t)<\delta}|z(s)-z(t)|=0\right\}$$

で表す.

命題 6.3.15. (T,ρ) を全有界な擬距離空間とする.

(1) $UC(T,\rho)$ は一様距離に関して可分で完備な部分距離空間 (Poland 空間) となる.

(2) $UC(T,\rho)$ は σ-コンパクトである.

定理 6.3.16. 写像列 $(X_n:\Omega_n\to\widetilde{B}(T))_{n\in\mathbb{N}}$ が $\widetilde{B}(T)$ においてタイトな Borel 可測写像 X に弱収束し，ある準距離 ρ について (T,ρ) は全有界になるとする. このとき，次の 2 条件は同値である.

(1) (X_n) は漸近的 ρ-同程度一様確率連続である.

(2) $P(X\in UC(T,\rho))=1$.

［証明］.

(1)⇒(2) (a)　$g:\widetilde{B}(T)\to\widetilde{B}((0,1))$ を

$$\mathrm{ev}_\delta(g(z))=g(z)(\delta):=\sup_{\rho(s,t)<\delta}|z(s)-z(t)|\quad(\delta\in(0,1))$$

で定めると，この写像は連続である：$\|g(y)-g(z)\|_\infty=\sup_{\delta\in(0,1)}|\mathrm{ev}_\delta(g(y))-\mathrm{ev}_\delta(g(z))|\le 2\|y-z\|_T$. すると，合成 $\mathrm{ev}_\delta(g(-)):\widetilde{B}((0,1))\to\mathbb{R}$ も連続だから，連続写像定理 6.2.42 から，$\mathrm{ev}_\delta(g(X_n))\Rightarrow\mathrm{ev}_\delta(g(X))$ である.

(b)　よって，portmanteau 定理 6.2.35(2) から，$\forall_{\delta\in(0,1)}\ \liminf_{n\to\infty}P_*\left(\mathrm{ev}_\delta(g(X_n))>\epsilon\right)\ge P\left(\mathrm{ev}_\delta(g(X))>\epsilon\right)$.

(c)　(X_n) は漸近的 ρ-同程度一様連続という仮定と併せると，任意の $\epsilon,\eta>0$ に対して $\delta\in(0,1)$ が存在して，

$$P\left(\mathrm{ev}_\delta(g(X))>\epsilon\right)\le\liminf_{n\to\infty}P_*\left(\mathrm{ev}_\delta(g(X_n))>\epsilon\right)<\limsup_{n\to\infty}P^*\left(\mathrm{ev}_\delta(g(X_n))>\epsilon\right)<\eta.$$

特に $\epsilon=\eta=2^{-m}$ の場合について考えることで，単調減少な正実数列 $(\delta_m)_{m\in\mathbb{N}}$ を得る.

(d)　すると，Borel-Cantelli の補題より，

$$P\left(\mathrm{ev}_{\delta_m}(g(X))>2^{-m}\ \text{ i.o.}\right)=P\left(\sup_{\rho(s,t)<\delta_m}|X(s)-X(t)|>2^{-m}\ \text{ i.o.}\right)=0.$$

これは $P(X\in UC(T,\rho))=1$ を含意する.

(2)⇒(1) (a)　$UC(T,\rho)$ は完備で可分 6.3.15 だから，ここに値を取る可測写像 X は緊密である（補題 6.2.26）. 従って，(X_n) は漸近的タイト (Prokhorov の定理の逆 6.2.47)：

$$\forall_{\eta,\epsilon>0}\ \exists_{K\overset{\text{cpt}}{\subset}UC(T,\rho)}\ P(X\in K)\ge 1-\eta,\quad\liminf_{n\to\infty}P_*\left(X_n\in K^\epsilon\right)\ge 1-\eta.$$

(b) コンパクト集合は全有界だから，K は $\{z_1,\cdots,z_p\}\subset UC(T,\rho)$ が存在してこれを中心とした半径 $\epsilon/3$ の開球で被覆できる．$z_i\in UC(T,\rho)$ より，$\forall_{i\in[p]}\;\exists_{\delta>0}\;\rho(s,t)<\delta\Rightarrow|z_i(s)-z_i(t)|<\epsilon/3$．よって，任意の $z\in K$ に対して，$\|z-z_i\|_T<\epsilon/3$ を満たす $i\in[p]$ が存在して，

$$\rho(s,t)<\delta\Rightarrow|z(s)-z(t)|\le\frac{2\epsilon}{3}+|z_i(s)-z_i(t)|<\epsilon.$$

よって，K は同程度一様連続である．

(c) さらに，いま任意の $\widetilde{z}\in K^\epsilon$ に対して，$z\in K$ が存在して $\|\widetilde{z}-z\|_T<\epsilon$ を満たすから，$\rho(s,t)<\delta\Rightarrow|\widetilde{z}(s)-\widetilde{z}(t)|\le 2\epsilon+|z(s)-z(t)|<3\epsilon$．すなわち，$K^\epsilon\subset\left\{z\in\widetilde{B}(T)\;\middle|\;\sup_{\rho(s,t)<\delta}|z(s)-z(t)|\le 3\epsilon\right\}$．(a) と併せると，

$$\liminf_{n\to\infty}P_*\left(\sup_{\rho(s,t)<\delta}|X_n(s)-X_n(t)|\le 3\epsilon\right)\ge 1-\eta.$$

これは漸近的 ρ-同程度一様確率連続性に同値．

■

要諦 6.3.17. 全く同様の議論を複素解析学で繰り返したが，これはコンパクト集合の全有界性に訴えていたのか．

6.3.5　準距離と L^p 距離

極限 X の候補は周辺写像の収束から同定出来る．そこで，準距離の見つけ方の指針が必要になる．標準的なものが，L^p が定める準距離 ρ_r である．

定義 6.3.18.

(1) L^p 擬距離を $\rho_r(s,t):=\{E|X(s)-X(t)|^r\}^{1/(r\vee 1)}$，$(r\in(0,\infty))$ と表す．

(2) $\widetilde{B}(T)$ の過程 X が T 上の準距離 ρ に関して r **次平均 ρ-一様連続**であるとは，$\rho(s_n,t_n)\to 0\Rightarrow E|X(s_n)-X(t_n)|^r\to 0$ が成り立つことをいう．

補題 6.3.19. $X:\Omega\to\widetilde{B}(T)$ をタイトな Borel 可測写像とする．

(1) このとき，T 上に準距離 ρ が存在して (T,ρ) は全有界となり，$P(X\in UC(T,\rho))=1$ が成り立つ．

(2) さらに，ある $0<\rho<\infty$ に対して X が r 次平均 ρ-連続であるならば，準距離 ρ_r に対して (T,ρ_r) も全有界となり，$P(X\in UC(T,\rho_r))=1$ が成り立つ．

ただし，全ての見本過程 $X(t,\omega):T\to\mathbb{R}$ は ρ-一様連続であることを仮定する．

［証明］．

(1) 定数列 (X) は漸近的タイト (Prokhorov の定理の逆 6.2.47) だから，漸近的タイト性の特徴付け 6.3.11 と漸近的 ρ-同程度一様確率連続性の特徴付け 6.3.16 より従う．

(2) 方針　$X:T\to\mathrm{Map}(\Omega,\mathbb{R})$ は r 次平均 ρ-一様連続とする：$\forall_{\epsilon>0}\;\exists_{\delta>0}\;\rho(s,t)<\delta\Rightarrow\rho_r(s,t)<\epsilon$．するとまず，$(T,\rho)$ が全有界ならば，(T,ρ_r) も全有界である．ここで，全ての見本過程 $X(t,\omega):T\to\mathbb{R}$ は ρ-一様連続であることを仮定する．準距離空間上の一様連続関数は，閉包上への連続延長を持つから，完備化 $(\overline{T},\overline{\rho})$ 上への連続延長 $\overline{X}(t,\omega)$ を持つ．$\overline{T}$ はやはり全有界なので，コンパクトである．ここで，このとき殆ど全ての見本過程が ρ_r-一様連続であることを示せば，$P(X\in UC(T,\rho_r))=1$ を得る 6.3.16.

殆ど全ての見本過程が ρ_r-一様連続であることの証明

(a) $X(t,\omega)$ が ρ_r-一様連続でないならば，ある $s,t\in\overline{T}$ について $\rho_r(s,t)=0\wedge\overline{X}(s,\omega)\ne\overline{X}(t,\omega)$ であることを示す．$X(t,\omega)$ が ρ_r-一様連続でないとする：$\exists_{\epsilon>0}\;\exists_{\{s_n\},\{t_n\}\subset T}\;[\rho_r(s_n,t_n)\to 0\wedge\forall_{n\in\mathbb{N}}\;|X(s_n,\omega)-X(t_n,\omega)\ge\epsilon|]$．$\overline{T}$ のコンパクト性から，$(s_n),(t_n)$ は $s\ne t\in\overline{T}$ に収束する部分列 $(s_{n'}),(t_{n'})$ を持つ．なお，$|X(s_{n'},\omega)-X(t_{n'},\omega)|\to$

$|X(s,\omega)-X(t,\omega)| \geq \epsilon > 0$ である．次に，$(s_{n'}),(t_{n'})$ は $\overline{X}$ 上の Cauchy 列であり，X は r 次平均 ρ-一様連続だから，$X(s_{n'})$ も $L_r(P)$ 空間上で Cauchy 列になる．よって，$s_{n'} \to s$ のとき，$X(s_{n'})$ はある $X' \in L_r(P)$ に r 次平均収束する．今，$\forall_{\omega\in\Omega}\ X(s_{n'},\omega) \to \overline{X}(s,\omega)$ であるから，殆ど全ての ω に対して $X'(\omega) = \overline{X}(s,\omega)$?．よって $\rho_r(s_{n'},s) \to 0, \rho_r(t_{n'},t) \to 0$．従って，

$$\rho_r(s,t) \leq \rho_r(s_{n'},s) + \rho_r(t_{n'},t) + \rho_r(s_{n'},t_{n'})$$

より，$\rho_r(s,t)=0$．

(b) $N := \left\{\omega\in\Omega \,\middle|\, \exists_{s,t\in\overline{T}}\ \rho_r(s,t)=0 \wedge \overline{X}(s,\omega)\neq\overline{X}(t,\omega)\right\}$ と定めて，$P(N)=0$ を示す．

■

6.3.6 Gauss 過程への弱収束の特徴付け

極限として最も頻繁に出現する確率過程が Gauss 過程と呼ばれるクラスで，この場合 2 次モーメント準距離 ρ_2 が最も取り扱いやすい．そこで，極限過程が Gauss であると仮定して，簡単化を試みる．

定義 6.3.20 (Gaussian process)**.** 確率過程 $X: T \to \mathrm{Meas}(\mathfrak{X},\mathbb{R})$ の全ての有限次元周辺確率ベクトル $(X(t_1),\cdots,X(t_k))$ $(t_1,\cdots,t_k \in T)$ が $\mathbb{R}^k$ 上の多変量正規分布に従うとき，X を **Gauss 過程**と呼ぶ．[†11]

補題 6.3.21. Gauss 過程 $X:\Omega\to\widetilde{B}(T)$ は r 次平均 ρ-一様連続である．

定理 6.3.22. Gauss 過程 $X:\Omega\to\widetilde{B}(T)$ について，次の 2 条件は同値．

(1) X はタイトである．
(2) ある $r>0$ に対して（結局は任意の $r>0$ に対して），(T,ρ_r) は全有界となり，$P(X\in UC(T,\rho_r))=1$ となる．

［証明］．

(2)⇒(1)　全有界な擬距離空間 (T,ρ_r) に対して，$UC(T,\rho_r)\subset\widetilde{B}(T)$ は Polish space である 6.3.15．すると定数列 (X) は漸近的 ρ-同程度一様確率連続である 6.3.16 から，(X) は漸近的タイト 6.3.11，すなわち X はタイト 6.2.47 である．

(1)⇒(2)　X はタイトであるとすると，ある擬距離 ρ が存在して，(T,ρ) は全有界となり，$P(X\in UC(T,\rho))=1$ である 6.3.19．X は r 次平均 ρ-一様連続である 6.3.21 ことから従う．

■

系 6.3.23 (タイトな Gauss 過程に弱収束する条件の特徴付け 3)**.** X を Gauss 過程とし，これが定める自然な準距離を ρ_r とする．任意の写像列 $(X_n:\Omega_n\to\widetilde{B}(T))$ に対して，X のバージョンであるようなタイトな Borel 可測写像 $\widetilde{X}:\Omega'\to\widetilde{B}(T)$ が存在してこれに弱収束するための必要十分条件は，ある $r>0$ が存在して（結局任意の $r>0$ に対して）次の 3 条件が成り立つことである：

(1) 任意の有限部分集合 $\{t_1,\cdots,t_k\}\subset T$ に対して，$(X_n(t_1),\cdots,X_n(t_k)) \Rightarrow (X(t_1),\cdots,X(t_k))$．
(2) (X_n) は漸近的 ρ_r-同程度一様確率連続である．
(3) (T,ρ_r) は全有界な準距離空間である．

［証明］．

十分性　漸近的タイト性の特徴付け 6.3.11 より (1),(2),(3) から (X_n) は漸近的タイトであり，(1) と併せると弱収束の条件 6.3.8(2) より従う．

必要性　弱収束の条件 6.3.8(1) より，(1) は従う．次に $\widetilde{B}$ はタイトな Gauss 過程であるから，定理 6.3.22 よりこれは (T,ρ_r) が全有界で $P(X\in UC(T,\rho_r))=1$ に同値．6.3.16 から (2) も従う．

■

[†11] これは，任意の $X_{t_1},\cdots,X_{t_k}$ の有限線型結合が一変数正規分布に従うことに同値．

6.3.7　Banach 空間値 Gauss 確率要素

何か米田の補題的なものを感じる．

定義 6.3.24. B を Banach 空間とする．$X:\Omega\to B$ を Borel 可測関数とする．このとき，任意の線型形式 $\phi:B\to\mathbb{R}\in B^*$ に対して $\rho(X)$ が正規分布に従うとき，X は B 上の **Gauss 過程**という．

注 6.3.25. 定義 6.3.20 よりも明らかに強い条件である．

命題 6.3.26. タイトな Borel 可測関数 $X:\Omega\to\widetilde{B}(T)$ について，次の 3 条件が同値．

(1) 任意の有限集合 $\{t_1,\cdots,t_k\}\subset T$ に対して，$(X_{t_1},\cdots,X_{t_k})$ は多変量正規分布に従う．
(2) 任意の連続線型写像 $\rho:\widetilde{B}(T)\to\mathbb{R}$ に対して，$\rho(X)$ は正規分布に従う．
(3) 任意の Banach 空間への任意の連続線型写像 $\rho:\widetilde{B}(T)\to B$ に対して，$\rho(X)$ は B 上の Gauss 過程である．

6.4　Glivenko-Cantelli クラスと Donsker クラス

経験測度 $\mathbb{P}_n$ が P に各点 $f\in\mathcal{F}$ で概収束することは，観測値 $X_1,\cdots,X_n$ が母平均に概収束することと同値で，これを大数の強法則という．これが任意の $f\in\mathcal{F}$ について一様に収束する $\sup_{f\in\mathcal{F}}|\mathbb{P}_nf-Pf|\xrightarrow{\text{a.s.}}0$ ためには，$\mathcal{F}$ が大きすぎないことが必要である．
収束先を見るために，$\sqrt{n}$ 倍して得る経験過程 $\mathbb{G}_n=\sqrt{n}(\mathbb{P}_n-P)$ に注目する．これが線型汎関数として $l^\infty(\mathcal{F})$ 上でタイトなものに弱収束するための条件は，まず $\sup_{f\in\mathcal{F}}|\mathbb{P}_nf-Pf|$ の一様収束性が必要なので，Glivenko-Cantelli の場合以上に $\mathcal{F}$ が大きすぎないことが必要である．なお，タイトな過程に弱収束するならば，その極限過程は F-固定端 Brown 運動となる必要がある．

6.4.1　定義と例

一般に $\mathbb{P}_n-P$ は各点収束$=w^*$-収束するが，一様収束$=\|\mathbb{P}_n-P\|_{\mathcal{F}}$ が収束するかは，$\mathcal{F}$ のエントロピーに依る．最も簡単な指標はブラケットエントロピーである．しかしこれが通用しないことがあり，その場合はより一般的な一様被覆数を用いる．

定義 6.4.1 (empirical measure, empirical process)**.** 可測空間 $(\mathcal{X},\mathcal{B})$ に値を取る確率変数 $X_1,\cdots,X_n$ について，

(1) これらが標本空間に定める**経験測度**とは，$\mathbb{P}_n(C):=\dfrac{1}{n}\displaystyle\sum_{i=1}^n 1_{\{X_i\in C\}},\quad(C\in\mathcal{B})$ と定める．これは Dirac 測度の線型結合になっている．
(2) 経験測度が定める写像 $\mathbb{P}_n:\mathrm{Meas}(\mathcal{X},\mathbb{R})\to\mathbb{R}$ を $f\mapsto\mathbb{P}_nf:=\dfrac{1}{n}\displaystyle\sum_{i=1}^n f(X_i)$ と定める．
(3) 標準化したバージョンを $\mathbb{G}_n:=\sqrt{n}(\mathbb{P}_n-P)$ と定めると，この写像 $\mathrm{Meas}(\mathcal{X},\mathbb{R})\supset\mathcal{F}\to\mathbb{R}$ を $\mathcal{F}$ で**添字づけされた経験過程**という．
(4) 符号付測度 $\mathbb{G}_n:=\dfrac{1}{\sqrt{n}}\displaystyle\sum_{i=1}^n(\delta_{X_i}-P)$ も経験過程と同一視する．

要諦 6.4.2. もやは経験過程と呼びたくないほどには一般的な対象となっている．

命題 6.4.3 (各点収束の結果は古典的)**.** $P|f|<\infty$ または $Pf^2<\infty$ とする．

(1) 大数の法則より $\mathbb{P}_nf\xrightarrow{\text{a.s.}}Pf$．

(2) 中心極限定理より $\mathbb{G}_n f \Rightarrow N(0, P(f-Pf)^2)$.

記法 6.4.4 (符号付き測度の一様ノルム). 符号付測度のノルムを $\|Q\|_{\mathcal{F}} := \sup\{|Qf| \in \mathbb{R} \mid f \in \mathcal{F}\}$ と定める.

定義 6.4.5. 空間 $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ を考える.

(1) 一様大数の法則 $\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{\text{a.s.}^*} 0$ を満たす $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ を P-**Glivenko-Cantelli** クラスという.
(2) 一様な外確率収束 $\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{P^*} 0$ を満たす $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ を弱 **Gilvenko-Cantelli** クラスという.
(3) $\forall_{x\in\mathcal{X}} \sup_{f\in\mathcal{F}} |f(x) - Pf| < \infty$ を仮定すると，経験過程 $\mathbb{G}_n$ は確率変数 $\mathbb{N} \to \widetilde{B}(\mathcal{F})$ とみなせる．この確率変数が一様中心極限定理を満たすとき，すなわち，あるタイトな Borel 可測写像 $\mathbb{G}$ に $l^\infty(\mathcal{F})$ 上で弱収束するとき，$\mathcal{F}$ を P-**Donsker** クラスという.
(4) 可測空間 $(\mathcal{X}, \mathcal{A})$ 上の任意の確率測度 P について P-Donsker であるとき，普遍的に Donsker であるという.

命題 6.4.6 (多次元中心極限定理).

(1) $Pf^2 < \infty$ とする．周辺写像 $\mathbb{G}_n f$ は収束し 6.4.3(2), 多次元中心極限定理から, $\forall_{k\in\mathbb{N}} \forall_{\{f_1,\cdots,f_k\}\subset\mathcal{F}} (\mathbb{G}_n f_1, \cdots, \mathbb{G}_n f_k) \Rightarrow N(0, \Sigma)$ が成り立つ．ただし，$N(0,\Sigma)$ は k 変量正規分布とし，Σ は $k\times k$ 行列で $\Sigma_{ij} = P(f_i - Pf_i)(f_j - Pf_j)$ とした.
(2) 極限過程 $\mathbb{G} : \mathcal{F} \to \mathbb{R}$ は平均 0，共分散関数が $E[\mathbb{G}f\mathbb{G}g] = P[(f-Pf)(g-Pg)] = P[fg] - PfPg$ の Gauss 過程である.

［証明］.

(1) 多次元中心極限定理から.
(2) $\widetilde{B}(\mathcal{F})$ 上で収束するならば，全ての有限周辺分布も収束する 6.3.8 から.

■

定義 6.4.7 (Brownian bridge). 補題 6.3.6 より，$\mathbb{G}$ がタイトであるから，分布が well-defined である．これを P-**Brown 橋**という.

要諦 6.4.8. どこを切っても正規分布な Gauss 過程である.

例 6.4.9 (Donsker の結果). $X_1, \cdots, X_n$ は独立同分布に従う $\mathbb{R}^d$ 値確率変数とする.

$$\mathcal{F} := \left\{1_{(-\infty,t]}(x) \in \mathrm{Meas}(\mathbb{R}^d, \mathbb{R}) \mid t \in \overline{\mathbb{R}}^d\right\} \quad ((-\infty, t] := (-\infty, t_1] \times \cdots \times (-\infty, t_d])$$

と定めると，これは Donsker クラスである．すなわち，$\mathbb{P}_n : \mathcal{F} \to \mathbb{R}; 1_{(-\infty,t]} \mapsto \mathcal{P}_n 1_{(-\infty,t]} = \dfrac{1}{n}\sum_{i=1}^n 1_{\{X_i \le t\}}$ はあるタイトな $\widetilde{B}(\mathcal{F})$ 値可測写像に収束する.

6.4.2　Donsker クラスの特徴付け

$\mathcal{F} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ が Donsker であることは，位相の言葉で特徴付けられる．標本空間 $(\mathcal{X}, \mathcal{A}, P)$ 上の実可測関数全体の集合 $\mathrm{Meas}(\mathcal{X}, \mathbb{R})$ に，P に関する積分について L_r-ノルムを入れたものを $L_r(P)$ で表すこととする．ここには，L_r-ノルムの定める距離 $\rho_r(s,t) := \{E|X(s) - X(t)|^r\}^{1/(r\vee 1)}$, $(r \in (0,\infty))$6.3.18 だけでなく，標準偏差がセミノルム ρ_P を定め，これが準距離 d_P を引き起こす.

命題 6.4.10 (標準偏差の定めるセミノルム). 測度 P が定める積分について 2 ノルムが入った関数空間 $L_2(P)$ を考える.

(1) 標準偏差

$$\begin{array}{rcl} \rho_P : L_2(P) & \longrightarrow & \mathbb{R} \\ \rotatebox{90}{$\in$} & & \rotatebox{90}{$\in$} \\ f & \longmapsto & \sqrt{P(f-Pf)^2} \end{array}$$

を定めると，これは $\mathrm{Meas}(\mathcal{X}, \mathbb{R})$ 上の半ノルムとなり，擬距離 d_P を定める.

(2) クラス $\mathcal{F} \subset L_2(P)$ で添字づけられた平均 0，共分散関数 $E[\mathbb{G}f\mathbb{G}g]$ の Gauss 過程 $\mathbb{G} : \mathcal{F} \to \mathbb{R}$ について[†12]，$\rho_2(f,g) = \left(E|\mathbb{G}f - \mathbb{G}g|^2\right)^{1/2} = \rho_P(f-g)$ が成り立つ．

［証明］．

(1) $\|af\| = |a|\|f\|$ は明らか．$\|f+g\| \le \|f\| + \|g\|$ は

$$\sqrt{P((f-Pf)+P(g-Pg))^2} \le \sqrt{P(f-Pf)^2} + \sqrt{P(g-Pg)^2}$$
$$\Leftrightarrow \quad P(f-Pf)^2 + P(g-Pg)^2 + 2P(f-Pf)(g-Pg) \le P(f-Pf)^2 + P(g-Pg)^2 + 2\sqrt{P(f-Pf)^2P(g-Pg)^2}$$

より．$P\{(f-Pf)(g-Pg)\}$ より，$P|f-Pf|\cdot P|g-Pg|$ の方が大きい．

(2) Gauss 過程 $\mathbb{G} : \mathcal{F} \to \mathbb{R}$ の共分散関数について，$E[\mathbb{G}f\mathbb{G}g] = P[(f-Pf)(g-Pg)]$ が成り立つから，

$$\begin{aligned}\rho_P(f-g)^2 &= P((f-Pf)-(g-Pg))^2 \\ &= P(f-Pf)^2 + P(g-Pg)^2 - 2P[(f-Pf)(g-Pg)] \\ &= E[(Gf)^2] + E[(Gg)^2] - 2E[GfGg] \\ &= E[(G(f-g))^2] = \rho_2(f-g).\end{aligned}$$

■

系 6.4.11 (Donsker クラスであることの特徴付け 1)．$\mathcal{F} \subset L_2(P)$ が Donsker クラスであることは，次の 2 条件に同値：

(1) (全有界) $(\mathcal{F}, d_P)$ が全有界である．

(2) (漸近的同程度連続性) $\forall_{\epsilon>0} \lim_{\delta\searrow 0} \limsup_{n\to\infty} P^*\left(\sup_{\rho_P(f-g)<\delta} |\mathbb{G}_n(f-g)| > \epsilon\right) = 0.$

［証明］．$\forall_{f\in\mathcal{F}}\ Pf^2 < \infty$ ならば，$\mathcal{F}$ を添字とした経験過程 $\mathbb{G}_n : \mathbb{N} \to \widetilde{B}(\mathcal{F})$ の任意の有界周辺写像が収束すること 6.4.6(1) より，Gauss 過程が弱収束する条件 6.3.23 のうち (1) は必然的に満たされる． ■

注 6.4.12 (漸近的同程度連続性の特徴付け)．$\mathcal{F}_\delta := \{f-g \in \mathcal{F} \mid f,g \in \mathcal{F}, \rho_P(f-g) < \delta\}$ について，(2) は $\forall_{\epsilon>0} \lim_{\delta\searrow 0} \limsup_{n\to\infty} P^*(\|\mathbb{G}_n\|_{\mathcal{F}_\delta} > \epsilon) = 0$ に同値．よって結局，

$$\forall_{\delta_n\searrow 0}\ \|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$$

と同値．

また，(2) は確率収束でなく，弱収束としても十分 [8](Lemma 8.17).

条件 $\|P\|_{\mathcal{F}} < \infty$ がある場合は，ρ_P 半ノルムの代わりに $L_2(P)$-ノルムを用いることができる．

最終型：

系 6.4.13 (Donsker クラスであることの特徴付け 2 ：2 ノルムで)．$\|P\|_{\mathcal{F}} < \infty$ ならば，$\mathcal{F} \subset L_2(\mathcal{X}, \mathcal{B}, P)$ が Donsker であることは，次の 2 条件に同値：

(1) $\mathcal{F}$ は $\|\cdot\|_{P,2}$ が定める距離に関して全有界である．

(2) $\forall_{\delta_n\searrow 0} \|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$．ただし，$\mathcal{F}_\delta := \{f-g \in \mathbb{G} \mid f,g \in \mathcal{F}, \|f-g\|_{P,2} < \delta\}$ とした．

［証明］．

(1) ρ_P と ρ_2 が同値だから？

(2) 注意参照．

■

[†12] Kolmogorov extension theorem により存在が保証される

6.4.3　漸近的同程度連続性続論

任意の可積分なクラス $\mathcal{F}\subset\mathcal{L}_P(\mathcal{X})$ について，標準的な Gauss 過程 $(G_Pf)_{f\in\mathcal{F}}$ が存在し，平均 0 で共分散 $\mathrm{Cov}(G_P(f),G_P(g))$ を持つ．この共分散は，$\mathcal{L}^2_P(\mathcal{X})$ 上に半内積と擬距離 ρ_P を定める．$\mathcal{X}=[0,1],P=U(\mathcal{X}),\mathcal{F}=\left\{1_{\{[0,x]\}}\mid x\in[0,1]\right\}$ のとき，G_P は標準 Brown 橋といい，$B(x)$ とも表す．

定義 6.4.14 (pre-Gaussian).

(1) クラス $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ が P-プレガウスであるとは，$\mathcal{F}$ を添字に持つタイトな Gauss 過程 $\mathcal{X}^\infty\to l^\infty(\mathcal{F})$ が存在することをいう．あきらかに，$\mathcal{F}$ が P-Donsker であることの必要条件であるが，タイトな Gauss 過程が必ずしも経験過程の極限として得られるとは限らないため，十分ではない．[†13][†14]

補題 6.4.15. クラス $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ について，次の 2 条件は同値．

(1) $\mathcal{F}$ は P-pre-Gaussian である．
(2) $(\mathcal{F},\rho_P)$ は全有界で，標準 Gauss 過程 $(G_Pf)_{f\in\mathcal{F}}:\mathcal{F}\to\mathrm{Map}(\mathcal{X},\mathbb{R})$ のバージョンであって，一様に ρ_P-連続な見本過程を持つものが存在する．

［証明］．収束先の一様連続性とタイト性との関係 6.3.16 より． ■

補題 6.4.16. pre-Gaussian なクラス $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ について，次の 2 条件は同値．

(1) $\mathcal{F}$ は Donsker である．
(2) $\mathbb{G}_n$ は漸近的タイトである．

要諦 6.4.17. そして，漸近的タイト性は，漸近的同程度一様確率連続性によって測る 6.3.11．こうして Donsker クラスにアプローチしていく．

6.5　連続度を抑えるための準備

6.5.1　Orlicz ノルムと最大不等式

道具 1：

大数の法則と中心極限定理を一般化するにあたって，いずれの場合も上限を評価する必要がある．この共通する手法を，Orlicz ノルムを用いて整備する．

経験過程の漸近的同程度連続性を示すのに有用な道具は，最大不等式と呼ばれるクラスの不等式である．そのために，L^p 空間を一般化する Banach 空間である Birnbaum-Orlicz 空間を定める．ほとんどの Sobolev 空間も含む．Orlicz ノルムは，収束についても L_r-ノルムと似た同じ性質を持っている．

6.5.1.1　Orlicz ノルム

L^p ノルムの一般化であるが，このノルムの扱いが難しい．

定義 6.5.1. $\psi:[0,\infty)\to[0,\infty)$ は $\psi(0)=0,\psi\neq0$ を満たす単調増加な凸関数とする．

[†13] P が離散測度であり，さらにいくつかの正則条件を満たすとき，P-preGaussian ならば，P-Donsker である．これは密度の言葉で一般化できる．
[†14] 種々の必要条件から，結局，この上の標準 Gauss 過程 $(G_Pf)_{f\in\mathcal{F}}:\mathcal{F}\to\mathrm{Map}(\mathcal{X},\mathbb{R})$ が，各点 $x\in\mathcal{X}$ について有界で，ρ_P-一様連続で，prelinear であることに同値？ https://en.wikipedia.org/wiki/Pregaussian_class．少なくとも，標準 Gauss 過程のタイトなバージョンが存在することとは同値 [10]．

(1) 確率変数 X の **Orlicz** ノルムまたは ψ-ノルムとは，

$$\|X\|_\psi := \inf\left\{c \in \mathbb{R}_{>0} \;\middle|\; E\left[\psi\left(\frac{|X|}{c}\right) \le 1\right]\right\}$$

と定める．ただし，$\inf\emptyset = \infty$ と定める．

補題 6.5.2 (well-definedness)**.** ψ-ノルムは，確かに $\|X\|_\psi < \infty$ を満たす確率変数がなす集合上でノルムを定める．

［証明］**.**

(1) $\|X\|_\psi = 0 \Rightarrow X = 0$.
(2) $\forall_{a\in\mathbb{R}}\ \|aX\|_\psi = |a|\|X\|_\psi$ は，$\dfrac{|X|}{c}$ という定め方による．
(3) 三角不等式は，Jensen の不等式 $\psi(E[X]) \le E[\psi(X)]$ により，ψ が凸のときから従う．？？

■

例 6.5.3.

(1) $\psi(x) = x^r$ $(r \ge 1)$ とすると，この ψ-ノルムとは L_r-ノルム $\|X\|_r = (E|X|^r)^{1/r}$ をいう．すなわち，r 次のモーメント．
(2) $\psi_r = e^{x^r} - 1$ $(r \ge 1)$ が定める Orlicz ノルムが，X の袖の振る舞いを強調するため，極限での振る舞いを考えたり，最大不等式の議論で重要な役割を果たす．$x^r \le \psi_r(x)$ なので，$\forall_{r\ge1}\ \|X\|_r \le \|X\|_{\psi_r}$．$\psi_2$ ノルムが有限なクラスを劣ガウスといい，ψ_1 ノルムが有限なクラスを劣指数的という．

命題 6.5.4 (Orlicz ノルムと L ノルム)**.** 任意の $1 \le r \le s$ について，

(1) $\|X\|_r \le \lceil r\rceil!\|X\|_{\psi_1}$.
(2) $\|X\|_{\psi_r} \le \|X\|_{\psi_s}(\log 2)^{1/s-1/r}$.

6.5.1.2　Orlicz ノルムと収束

補題 6.5.5 (収束についての性質)**.** $\psi : [0,\infty) \to [0,\infty)$ は $\psi(0) = 0, \psi \ne 0$ を満たす単調増加な凸関数とする．任意の確率変数列 (X_n) に対して，

(1) $0 \le X_n \nearrow X$ a.s. $\Rightarrow \|X_n\|_\psi \nearrow \|X\|_\psi$.
(2) $\|X_n\|_\psi \to 0 \Rightarrow X_n \xrightarrow{p} 0$.

［証明］**.**

(1) 明らか.
(2) $\|X_n\|_\psi \to 0$ ならば，$|X_n|$ が 0 に収束しない集合の測度は 0 である.

■

補題 6.5.6.

$$P(|X| > x) \le P\left(\psi\left(\frac{|X|}{\|X\|_\psi}\right) \ge \psi\left(\frac{x}{\|X\|_\psi}\right)\right) \le 1 \wedge \psi\left(\frac{x}{\|X\|_\psi}\right)^{-1}$$

［証明］**.**

$$\begin{aligned}
P(|X| > x) &= P\left(\frac{|X|}{\|X\|_\psi} > \frac{x}{\|X\|_\psi}\right) \\
&\le P\left(\psi\left(\frac{|X|}{\|X\|_\psi} \ge \psi\left(\frac{x}{\|X\|_\psi}\right)\right)\right) && \psi\text{の定義} \\
&\le \frac{E\left[\psi\left(\frac{|X|}{\|X\|_\psi}\right)\right]}{\psi\left(\frac{x}{\|X\|_\psi}\right)} \wedge 1 && \text{Markov の不等式} \\
&\le \frac{1}{\psi\left(\frac{x}{\|X\|_\psi}\right)} \wedge 1 && \text{ノルムの定義}
\end{aligned}$$

■

命題 6.5.7 (尾部確率の評価)**.** 確率変数 X と任意の $r \ge 1$ に関して，次の 2 条件は同値：

(1) $\|X\|_{\psi_r} < \infty$.
(2) $\exists_{0<C,K<\infty}\ \forall_{x>0}\ P(|X| > x) \le Ke^{-Cx^r}$.

さらに，もし (1),(2) のいずれかが成り立つならば，(2) について $K = 2, C = \|X\|_{\psi_r}^{-r}$ がこれを満たす．また (2) を満たす任意の $0 < C, K < \infty$ について，$\|X\|_{\psi_r} \le \left(\frac{1+K}{C}\right)^{1/r}$ が成り立つ．

［証明］.

(1)⇒(2)　$\|X\|_{\psi_r} < \infty$ と仮定する．簡単にわかる事実 $\forall_{u>0}\ 1 \wedge (e^u - 1)^{-1} \le 2e^{-u}$ と補題より，

$$P(|X| > x) \le 2\exp\left(-\frac{x^r}{\|X\|_{\psi_r}^r}\right)$$

という評価を得る．

(2)⇒(1)　任意の $c \in (0, C)$ について，確率変数 $\psi_r(|X|) = e^{c|X|^r} - 1$ の期待値は，Fubini の定理を用いて

$$\begin{aligned} E\left[e^{c|X|^r} - 1\right] &= E\left[\int_0^{|X|^r} ce^{cs}ds\right] \\ &= \int_0^\infty P(|X| > s^{1/r})ce^{cs}ds && \text{Fubini の定理} \\ &\le \int_0^\infty Ke^{-Cs}ce^{cs}ds && \text{(2) の仮定} \end{aligned}$$

と評価できる．これについて，$c \le \frac{C}{1+K} \Leftrightarrow c^{-1/r} \ge \left(\frac{1+K}{C}\right)^{1/r}$ ならば常に $\frac{Kc}{C-c} \le 1$ と抑えられるから，Orlicz ノルムについて $\|X\|_{\psi_r} \le \left(\frac{1+K}{C}\right)^{1/r} < \infty$ である．

■

要諦 6.5.8. Chebyshev 不等式のような使い方．tail を強調するためである．

6.5.1.3　最大量の動きの評価

$\max_{i\in N}|X_i|^r \le \sum_{i\in N}|X_i|^r$ という不等式から，L_r ノルムに関して

$$\|\max_{i\in\mathbb{N}} X_i\|_r \le \left(E\left[\max_{i\in\mathbb{N}}|X_i|^r\right]\right)^{1/r} \le m^{1/r}\max_{i\in N}\|X_i\|_r$$

を得る．同様の不等式が，Orlicz ノルムについても一般化できる．

［証明］.

$$\begin{aligned} \left\|\max_{1\le i\le m} X_i\right\|_r &= \left(E\left[\left|\max_{1\le i\le m} X_i\right|^r\right]\right)^{1/r} \\ &\le \left(E\left[\max_{1\le i\le m}|X_i|^r\right]\right)^{1/r} \\ &\le \left(E\left[\sum_{1\le i\le m}|X_i|^r\right]\right)^{1/r} \\ &\le \left(m\cdot\max_{1\le i\le m} E[|X_i|^r]\right)^{1/r} = m^{1/r}\max_{1\le i\le m}\|X_i\|_r. \end{aligned}$$

■

命題 6.5.9 (最大不等式). $\psi : [0,\infty) \to [0,\infty)$ は $\psi(0) = 0, \psi \neq 0$ を満たす単調増加な凸関数であって，

$$\exists_{c>0} \limsup_{x,y\to 0} \frac{\psi(x)\psi(y)}{\psi(cxy)} < \infty$$

を満たすとする．この時，任意の確率変数 $X_1, \cdots, X_m$ について，

$$\left\| \max_{1\le i\le m} X_i \right\|_\psi \le \left\| \max_{1\le i\le m} |X_i| \right\|_\psi \le K\psi^{-1}(m) \max_{1\le i\le m} \|X_i\|_\psi .$$

ただし，K は ψ のみに依存する定数であり，$\psi^{-1}(m)$ は逆像の一点を表す．

［証明］．

特別の ψ について　ψ は $\exists_{c>0}\ \forall_{x,y\ge 1}\ \dfrac{\psi(x)\psi(y)}{\psi(cxy)} \le 1, \psi(1) \le \dfrac{1}{2}$ を満たすとする． 1 つ目の仮定より，この時 $\forall_{x\ge y\ge 1} \psi\left(\dfrac{x}{y}\right) \le \dfrac{\psi(cy)}{\psi(x)} \le \dfrac{\psi(cy)}{\psi(y)} \le \dfrac{\psi(cx)}{\psi(y)}$ であるから，任意の $y \ge 1, a > 0$ について，$\dfrac{|X_i|}{a} \ge y \ge 1$ を満たすときと満たさないときに分けて考えると，

$$\begin{aligned} \max_{1\le i\le m} \psi\left(\frac{|X|_i}{ay}\right) &\le \max_{1\le i\le m} \left[\frac{\psi\left(\frac{c|X_i|}{a}\right)}{\psi(y)} + \psi\left(\frac{|X_i|}{ay}\right) 1_{\frac{|X_i|}{ay}<1} \right] \\ &\le \sum_{i=1}^m \frac{\psi\left(\frac{c|X_i|}{a}\right)}{\psi(y)} + \psi(1). \end{aligned}$$

特に $a = c \max_{1\le i\le m} \|X_i\|_\psi$ とすると両辺の期待値を取って，

$$\begin{aligned} E\left[\psi\left(\frac{\max_{1\le i\le m}|X_i|}{ay}\right)\right] &\le \sum_{i=1}^m \frac{E\left[\psi\left(\frac{|X_i|}{\max_{1\le i\le m}\|X_i\|_\psi}\right)\right]}{\psi(y)} + \psi(1) \\ &\le \frac{m}{\psi(y)} + \psi(1). \end{aligned}$$

ただし，最後の評価は $\psi\left(\dfrac{|X_i|}{\max_{1\le i\le m}\|X_i\|_\psi}\right) \le 1$ による．さらに $y \in \psi^{-1}(2m)$ とすると $y \ge 1$ で，右辺は $\dfrac{1}{2} + \dfrac{1}{2} = 1$ となるから，Orlicz ノルムは

$$\| \max_{1\le i\le m} |X_i| \|_\psi \le ay = c\psi^{-1}(2m) \max_{1\le i\le m} \|X_i\|_\psi$$

となる．ψ の凸性から $\psi^{-1}(2m) \le 2\psi^{-1}(m)$ より，$K := 2c$ とすれば成立．

一般の ψ について　補題の仮定を満たす関数 ψ から，$\exists_{c>0}\ \forall_{x,y\ge 1}\ \dfrac{\phi(x)\phi(y)}{\phi(cxy)} \le 1, \phi(1) \le \dfrac{1}{2}$ を満たす単調増加関数 $\phi, \phi(0) = 0, \phi \neq 0$ を構成できることを示す．

■

要諦 6.5.10 (確率変数列の最大値という特性量の振る舞いの記述). 特に ψ_r $(r \ge 1)$ の場合を考える．$c = 1$ とすると命題の条件を満たすから，$\psi_r^{-1}(m) = (\log(m+1))^{1/r}$ なので，ψ-ノルムの最大量の増大レートは対数関数的である．

補題 6.5.11 (ψ_r-ノルムで尾部を評価する). m 個の確率変数 $X_1, \cdots, X_m$ を考える．尾部確率 $P(|X_i| > x)$ は，

$$\exists_{a,b\ge 0}\ \forall_{i\in[m]}\ \forall_{x>0}\ P(|X_i| > x) \le 2e^{-\frac{1}{2}\frac{x^2}{b+ax}}$$

を満たすとする．この時，

$$\left\| \max_{1\le i\le m} |X_i| \right\|_{\psi_1} \le K \left\{ a\log(1+m) + \sqrt{b}\sqrt{\log(1+m)} \right\}$$

が成り立つ．ここで，K は a, b や確率変数に無関係な普遍定数である．

［証明］．

$a, b > 0$ のとき　$x \le \frac{b}{a} \Rightarrow b + ax \le 2b$ なので，$P(|X_i| > x) \le 2\exp\left(-\frac{x^2}{4b}\right)$．$x > \frac{b}{a} \Rightarrow \frac{b}{x} + a \le 2a$ なので，$P(|X_i| > x) \le 2\exp\left(-\frac{x}{4a}\right)$ が仮定からわかる．よって，任意の $x > 0$ について，

$$P\left(|X_i|1_{\{|X_i|\le\frac{b}{a}\}} > x\right) \le 2\exp\left(-\frac{x^2}{4b}\right), \qquad P\left(|X_i|1_{\{|X_i|<\frac{b}{a}\}} > x\right) \le 2\exp\left(-\frac{x}{4a}\right).$$

命題 6.5.7 より，

$$\||X_i|1_{\{|X_i|\le\frac{b}{a}\}}\|_{\psi_2} \le \sqrt{12b}, \qquad \||X_i|1_{\{|X_i|<\frac{b}{a}\}}\|_{\psi_1} \le 12a.$$

よって，三角不等式と命題 6.5.4(2)，そして最大不等式より，

$$\begin{aligned}\|\max_{1\le i\le m}|X_i|\|_{\psi_1} &\le (\log 2)^{-1/2}\|\max_{1\le i\le m}|X_i|1_{\{|X_i|\le\frac{b}{a}\}}\|_{\psi_2} + \|\max_{1\le i\le m}|X_i|1_{\{|X_i|<\frac{b}{a}\}}\|_{\psi_1} \\ &\le \left(\frac{\sqrt{12b}}{\sqrt{\log 2}}\psi_2^{-1}(m) + 12a\psi_1^{-1}(m)\right)K.\end{aligned}$$

その他の場合　$a > 0, b = 0$ の場合，$a = 0, b > 0$ の場合は $a, b > 0$ の場合の極限を取ることで得る．$a = b = 0$ の場合は確率 1 で $X_i = 0$ より，自明．

■

要諦 6.5.12. 尾部確率は ψ_r-ノルムによって評価できる．大きな x については $\exp(-x/(4a))$ で，0 に近い x については $\exp(-x^2(4b))$ となる．

補題 6.5.13 (Bernstein). 有界な区間に値を取る確率変数 $Y_1, \cdots, Y_n \to [-M, M]$ は独立で平均 0 がであるとする．この時，任意の $x > 0, v \ge \mathrm{Var}(Y_1 + \cdots + Y_n)$ に対して，

$$P(|Y_1 + \cdots + Y_n| > x) \le 2e^{-\frac{1}{2}\frac{x^2}{v+Mx/3}}.$$

［証明］.

尾部確率の評価　$S := Y_1 + \cdots + Y_n$ とおく．Chebyshev の不等式 $\forall_{x,t>0}\ P(S \ge x) \le \frac{E[e^{tS}]}{e^{tx}} = \frac{\prod_{i=1}^n E[e^{tY_i}]}{e^{tx}}$ に注意．各 i について，$E[Y_i] = 0$ したがって $\mathrm{Var}[Y_i] = E[Y_i^2]$ に注意すると，広義一様収束より項別積分して，

$$\begin{aligned}E[e^{tY_i}] &= 1 + E[Y_i] + \sum_{k=2}^\infty \frac{t^k}{k!}E[Y_i^2Y_i^{k-2}] \\ &\le 1 + \mathrm{Var}[Y_i]\sum_{k=2}^\infty \frac{t^k}{k!}M^{k-2}\end{aligned}$$

であるから，$g(t) := \frac{e^{tM} - 1 - tM}{M^2}$ とおくと，$E[e^{tY_i}] \le \exp(\mathrm{Var}[Y_i]g(t))$ を得る．以上を併せて，$P(S \ge x) \le \exp(vg(t) - tx)$．同様にして $P(-S \ge x) \le \exp(vg(t) - tx)$．

変形　以上より，$B(\lambda) := 2\frac{(1+\lambda)\log(1+\lambda) - \lambda}{\lambda^2}$ とおくと，

$$P(|S| > x) \le 2\inf\exp(vg(t) - tx) = 2\exp\left(-\frac{1}{2}\frac{x^2 B\left(\frac{Mx}{v}\right)}{v}\right)$$

と評価できる．これと $\forall_{\lambda>0}\ \frac{1}{1+\lambda/3}$ を併せれば良い．

■

要諦 6.5.14. 実際の補題 6.5.11 が使える具体例として，有界な範囲に値を取る，独立で平均 0 な確率変数列の部分和が，部分分散の正規分布に従うことなどから現れる．

6.5.1.4　被覆数と梱包数

一般最大不等式の導出

確率過程はいわば無限の確率変数を持つから，最大不等式はそのままでは役に立たない．Kolmogorov による連鎖 (chaining) と呼ばれる手法で有限への帰着を行う．添字集合 T に備わる擬距離が定める計量エントロピーに引き戻す．エントロピーの対数は，指数関数の引き戻しとして自然に現れるのかな．

定義 6.5.15 (covering number, separated, packing number, entropy number)**.** (T,d) を擬距離空間とする．

(1) T を被覆するのに必要な半径 ϵ の閉球の最小個数を**被覆数**とよび，$N(\epsilon;T,d)$ で表す．
(2) $S\subset T$ が **ϵ-分離的**であるとは，$\forall_{x,y\in S}\;x\ne y\Rightarrow d(x,y)>\epsilon$ が成り立つことをいう．
(3) $D(\epsilon;T,d):=\max\left\{|T|\in\mathbb{N}\mid T\subset S\text{ は}\epsilon\text{-分離的}\right\}$ を**パッキング数**という．ただし，S のどの 2 点も ϵ-分離的に取れないならば，$D=1$ となることに注意．
(4) 被覆数，及び，パッキング数の対数を，**エントロピー数**という．

注 6.5.16. この用語を用いると，(T,d) が全有界であることは，$\forall_{\epsilon>0}\;N(\epsilon;T,d)<\infty$ に同値かつ $\forall_{\epsilon>0}\;D(\epsilon;T,d)<\infty$ に同値．

補題 6.5.17.

$$N(\epsilon;T,d)\le D(\epsilon;T,d)\le N(\epsilon/2,T,d)$$

［証明］．

$N(\epsilon;T,d)\le D(\epsilon;T,d)$　T は $D:=D(\epsilon;T,d)$ とすると，ある $x_1,\cdots,x_D$ を中心とする半径 ϵ の閉球で覆われる．もし被覆しない（漏れ出る点がある）とすると，D がパッキング数であること（その最大性）に矛盾．

$D(\epsilon;T,d)\le N(\epsilon/2;T,d)$　T は $N:=N(\epsilon/2;T,d)$ 個の半径 $\epsilon/2$ の閉球 $B_1,\cdots,B_N$ で被覆できる．このとき，任意の ϵ-分離的な集合 $S\subset T$ は，$|B_i\cap S|\le 1$ を満たすことを示せば良いが，これは明らか．

■

注 6.5.18. N,D が距離関数として同値，つまり同じ位相を定める，すなわち $\epsilon\searrow 0$ における振る舞いが本質的に同等であることがわかる．

6.5.2　一般化最大不等式

sup よりは積分の方がマジかなあ．

定理 6.5.19 (generalized maximal inequality)**.** $\psi:[0,\infty)\to[0,\infty)$ は $\psi(0)=0,\psi\ne 0$ を満たす単調増加な凸関数であって，

$$\exists_{c>0}\;\limsup_{x,y\to 0}\frac{\psi(x)\psi(y)}{\psi(cxy)}<\infty$$

を満たすとする．確率過程 $X:T\to\widetilde{B}(\Omega)$ は T 上のある擬距離 d に関して可分で，$\forall_{s,t\in T}\;\exists_{r\in\mathbb{R}}\;\|X(s)-\|(\|t)\|_\psi\le rd(s,t)$ を満たすとする．このとき，

(1) 任意の $\eta,\delta>0$ に対して，

$$\left\|\sup_{d(s,t)\le\delta}|X(s)-X(t)|\right\|\le K\left[\int_0^\eta\psi^{-1}(D(\epsilon;T,d))d\epsilon+\delta\psi^{-1}(D(\eta;T,d)^2)\right].$$

ここで，$K<\infty$ は ψ,r のみに依存する定数．

(2)

$$\left\|\sup_{s,t\in T}|X(s)-X(t)|\right\|_\psi\le 2K\int_0^{\operatorname{diam}T}\psi^{-1}(D(\epsilon;T,d))d\epsilon.$$

［証明］．

(1) 右辺が可積分な場合について示せば良い．各 T_j が $\eta 2^{-j}$-分離的になるように有限集合の増大列 $(T_j)_{j\in\mathbb{N}}$ を取ると，T_∞ は T 上で稠密になる．

連鎖の評価　各点 $t_{j+1}\in T_{j+1}$ に対して，$d(t_j,t_{j+1})\le\eta 2^{-j}$ を満たす点 $t_j\in T_j$ を選び，写像の列 $t_{j+1}\to t_j\to\cdots\to t_0$ を構成する．これを連鎖とよび，各連鎖全体の集合を $T(j+1,j)\subset\mathrm{Map}(T_{j+1},T_j)$ と表すと，任意の $s_{k+1},t_{k+1}\in T_{k+1}$ について，

$$\begin{aligned}|\{X(s_{k+1}-X(t_{k+1}))\}-\{X(s_0)-X(t_0)\}| &= \left|\sum_{j=0}^k\{X(s_{j+1})-X(s_j)\}-\sum_{j=0}^k\{X(t_{j+1})-X(t_j)\}\right| \\ &\le 2\sum_{j=0}^k\max_{u,v\in T(j+1,j)}|X(u)-X(v)|\end{aligned}$$

が成り立つ．最大不等式 6.5.9 より，$K_0\in\mathbb{R}$ が存在して，この max の Orlicz ノルムは，仮定より $\|X(s)-X(t)\|_\psi\le rd(u,v)\le r\eta 2^{-j}$ だから，

$$\begin{aligned}\left\|\max_{s,t\in T_{k+1}}|\{X(s)-X(s_0)\}-\{X(t)-X(t_0)\}|\right\|_\psi &\le K_0\sum_{j=0}^k\psi^{-1}(D(\eta 2^{-j-1};T,d))\eta 2^{-j} \\ &= 4K_0\sum_{j=0}^k\psi^{-1}(D(\eta 2^{-k+j-1};T,d))\eta 2^{-k+j-2} \\ &\le 4\eta K_0\int_0^1\psi^{-1}(D(\eta u;T,d))du = 4K_0\int_0^\eta\psi^{-1}(D(\epsilon;T,d))d\epsilon.\end{aligned}$$

T_0 に引き戻して評価　三角不等式より，

$$X(s_0)-X(t_0)\le|\{X(s_0)-X(s_{k+1})-\{X(t_0)-X(t_{k+1})\}\}|+|X(s_{k+1})-X(t_{k+1})|$$

これらを併せて，最大不等式より，$C\in\mathbb{R}$ が存在して，

$$\begin{aligned}\left\|\max_{s,t\in T_{k+1},d(s,t)<\delta}|X(s)-X(t)|\right\|_\psi &\le \left\|\max_{s,t\in T_{k+1}}|\{X(s)-X(s_0)\}-\{X(t)-X(t_0)\}|\right\|_\psi+|X(s_0)-X(t_0)| \\ &\le 8K_0\int_0^\eta\psi^{-1}(D(\epsilon;T,d))d\epsilon+\left\|\max_{s_{k+1},t_{k+1}\in T_{k+1}}|X(s_{k+1})-X(t_{k+1})|\right\|_\psi \\ &\le 8K_0\int_0^\eta\psi^{-1}(D(\epsilon;T,d))d\epsilon+C\delta\psi^{-1}(D(\eta;T,d)^2)\end{aligned}$$

よって，$K:=8K_0\vee C$ とおくと，

$$\left\|\sup_{s,t\in T_{k+1},d(s,t)<\delta}|X(s)-X(t)|\right\|_\psi\le K\left[\int_0^\eta\psi^{-1}(D(\epsilon;T,d))d\epsilon+\delta\psi^{-1}(D(\eta;T,d)^2)\right]$$

X が可分のとき，$\sup_{s,t\in T_\infty}=\sup_{s,t\in T}$ が成立．

(2) $\delta=\eta=\mathrm{diam}T$ とおくと，$D(\eta;T,d)^2=D(\eta;T,d)=1$ であり，

$$\delta\psi^{-1}(D(\eta;T,d)^2)=\int_0^\eta\psi^{-1}(D(\eta;T,d))d\epsilon\le\int_0^\eta\psi^{-1}(D(\epsilon;T,d))d\epsilon.$$

■

要諦 6.5.20 (可分確率過程)**.** 可分な確率過程とは，値域が有限個の点で近似可能であることをいう．

可分性の仮定が現れるが，cadlag 確率過程 $X:\Omega\to D([a,b])$ は可分である．可分な確率過程は確率 1 で可分な部分空間に乗るから，全ての $\delta>0$ に対して，確率 1 で

$$\sup_{s,t\in T;d(s,t)<\delta}|X(s)-X(t)|=\sup_{s,t\in T^*;d(s,t)<\delta}|X(s)-X(t)|$$

が成り立つということになる．

ある $\eta>0$ について $\int_0^\eta\psi^{-1}(D(\epsilon;T,d))d\epsilon<\infty$ という可積分性条件を満たすならば，この定理の (1) は X がほとんど確実に d-一様連続な見本過程を持つことを示している．これは ψ-ノルムでの収束は確率収束を表す 6.5.5 ことによる．(2) は，X が確率 1 で $UC(T,d)$ に値を取ることを言っている，$\forall_{t_0\in T}\;\|\sup_{t\in T}|X(t)|\|_\psi-\|X(t_0)\|_\psi\le\|\sup_{s,t\in T}|X(s)-X(t)|\|_\psi$ と併せればわかる．(T,d) が全有界のとき，$UC(T,d)$ は σ-コンパクトであるから，X が Borel 可測であることとタイトであることは同値 6.2.26.

6.5.3　劣 Gauss 過程に関する最大不等式

興味ある対象のクラスとして劣 Gauss 過程を定義し，これについて一般化最大不等式を考えると，自然に逆関数として のエントロピーが現れる．Hoeffding の不等式により，Rademacher 過程が劣 Gauss であることを導く．このように，Hoeffding の不等式は特に有用な集中不等式である．

定義 6.5.21 (sub-Gaussian process, Rademacher 過程)**.**

(1) 確率過程 $X:T\to\widetilde{B}(\Omega)$ が

$$\forall_{s,t\in T}\;\forall_{x>0}\;P(|X(s)-X(t)|>x)\le 2\exp\left(-\frac{1}{2}\frac{x^2}{d(s,t)^2}\right)$$

満たすとき，X は T の擬距離 d に関して**劣 Gauss 過程**であるという．

(2) Rademacher 確率分布に従う独立同分布な確率変数列 $\epsilon_1,\cdots,\epsilon_n$ に対して，$X:=\sum_{i=1}^n\epsilon_i$ と定めた $X:\mathbb{R}^n\to\mathbb{R}$ を **Rademacher 過程**という．

要諦 6.5.22. 標準正規過程 X は $P(X>x)=O\left(\frac{e^{-x^2/2}}{x}\right)$ $(x\to\infty)$ の尾部確率をもつことから，係数の 2 と $-\frac{1}{2}$ を得る．

例 6.5.23.

(1) 平均 0 の Gauss 過程は明らかに，標準偏差が定める擬距離 $d(s,t)=\sigma(X(s)-X(t))=\mathrm{Var}(X(s)-X(t))^{1/2}$ に関して劣 Gauss である．

(2) $[0,1]$ 上の Brown 運動は $d(s,t)=|s-t|^{1/2}$ に関して可分な劣 Gauss 過程である．

命題 6.5.24 (Hoeffding の不等式)**.** $\epsilon_1,\cdots,\epsilon_n$ を独立な Rademacher 確率変数とする．

$$\forall_{a=(a_1,\cdots,a_n)^T\in\mathbb{R}^n}\;\forall_{x>0}\;P\left(\left|\sum_{i=1}^n\epsilon_ia_i\right|>x\right)\le 2\exp\left(-\frac{1}{2}\frac{x^2}{\|a\|^2}\right).$$

特に，$\left\|\sum_{i=1}^n\epsilon_ia_i\right\|_{\psi_2}\le\sqrt{6}\|a\|$.

［証明］．任意の Rademacher 確率変数 ϵ について，$\forall_{\lambda\in\mathbb{R}}\;E[e^{\lambda\epsilon}]=\frac{e^\lambda+e^{-\lambda}}{2}=\sum_{i=0}^\infty\frac{\lambda^{2i}}{(2i)!}\le e^{\frac{\lambda^2}{2}}$．ただし，$(2i)!\ge 2^ii!$ より従う．故に，Markov の不等式と併せると，任意の $\lambda>0$ に関して，

$$P\left(\sum_{i=1}^n\epsilon_ia_i>x\right)\le\frac{E\left[e^{\lambda\sum_{i=1}^n\epsilon_ia_i}\right]}{e^{\lambda x}}\le e^{\frac{\lambda^2}{2}\|a\|^2-\lambda x}$$

最良の上界は $\lambda=\frac{x}{\|a\|^2}>0$ とすれば良い．$\epsilon_1,\cdots,\epsilon_n$ に -1 をかけても結合分布は不変（対称）なので，結論を得る．

ψ_2 ノルムについては 6.5.7 による． ■

注 6.5.25. 評価 $\|X(s)-X(t)\|_{\psi_2}\le\sqrt{6}d(s,t)$ は一般の劣 Gauss 過程について成り立つ．$\psi_2^{-1}(x)=\sqrt{\log(1+x)}$ なので，劣 Gauss 過程の一般最大不等式はエントロピー積分が出現する．

系 6.5.26 (劣 Gauss 過程に関する一般化最大不等式)**.** 擬距離 d に関して可分な劣 Gauss 過程 $X:T\to\widetilde{B}(\Omega)$ について，

(1) $\forall_{\delta>0}\;E\left(\sup_{d(s,t)\le\delta}|X(s)-X(t)|\right)\le K\int_0^\delta\sqrt{\log D(\epsilon;T,d)}d\epsilon$．ここで K は普遍定数である．

(2) $\forall_{t_0\in T}\;E\left(\sup_{t\in T}|X(t)|\right)\le E|X(t_0)|+K\int_0^{\mathrm{diam}T}\sqrt{\log D(\epsilon;T,d)}d\epsilon$.

［証明］．

(1) 一般化最大不等式 6.5.19 について $\psi := \psi_2, \eta = \delta$ とすると，

$$\delta\psi_2^{-1}(D(\delta,T,d)^2) \le \delta\sqrt{2}\delta\psi_2^{-1}(D(\delta;T,d)) \le \sqrt{2}\int_0^\delta \psi_2^{-1}(D(\epsilon;T,d))d\epsilon$$

という特性より，Orlicz ノルムが

$$\begin{aligned}\left\|\sup_{d(s,t)\le\delta}|X(s)-X(t)|\right\| &\le \left[\int_0^\delta \psi_2^{-1}(D(\epsilon;T,d))d\epsilon + \delta\psi_2^{-1}(D(\delta;T,d)^2)\right] \\ &\le K'\int_0^\delta \sqrt{\log(1+D(\epsilon;T,d))}d\epsilon\end{aligned}$$

と評価できる．いま，$\forall_{0\le\epsilon\le\mathrm{diam}T}\ D(\epsilon;T,d)\ge 2$，$\forall_{m\ge 2}\ \log(1+m)\le 2\log m$ に注意すると，$\delta<\mathrm{diam}T$ ならば，K' を十分大きく取ることにより対数の引数の中の 1 は消去できる．$\forall_{s,t\in T}\ d(s,t)\le\mathrm{diam}T$ であることと，L_1 ノルムと ψ_2 ノルムの関係 6.5.5 より，

$$E\left(\sup_{d(s,t)\le\delta}|X(s)-X(t)|\right) \le (\log 2)^{-1/2}\left\|\sup_{d(s,t)\le\delta}|X(s)-X(t)|\right\|.$$

よって，結論を得る．

(2) $\delta := \mathrm{diam}T$ と取ることによる．

■

6.5.4　対称化不等式と可測性

道具 2：対称化という技法

一般の経験過程は対称化された経験過程で抑え，対称化された経験過程は Rademacher 過程であるから，系 6.5.26 の形の最大不等式で抑えられる．一般の経験過程の上限は可測ではないから，対称化された経験過程で抑える際に，外積分が必要となる．

6.5.4.1　対称化不等式

経験過程を対称化すると劣 Gauss 過程になるから，前節までで議論し，なんとか単純な組み合わせ論的な形に還元した最大化不等式を使える．

定義 6.5.27. 経験測度 $\mathbb{P}_n$ に対して，

(1) $\mathbb{P}_n - P : \mathcal{F}\to\mathbb{R}; f\mapsto(\mathbb{P}_n-P)f = \dfrac{1}{n}\displaystyle\sum_{i=1}^n\left(f(x_i)-\int_\Omega fdP\right)$ を経験過程という．

(2) $\mathbb{P}_n^\circ : \mathcal{F}\to\mathbb{R}; f\mapsto\mathbb{P}_n^\circ f := \dfrac{1}{n}\displaystyle\sum_{i=1}^n\epsilon_i f(x_i)$ を対称化された経験過程という．

ただし，$\epsilon_1,\cdots,\epsilon_n$ は互いに独立で，$X_1,\cdots,X_n$ とも独立な Rademacher 確率変数とした．

補題 6.5.28 (経験過程を対称化すると劣 Gauss である).

(1) どちらの過程も，平均関数は 0 となる．

(2) 固定された $X_1,\cdots,X_n$ について $\mathbb{P}_n^\circ$ は Rademacher 過程であり，したがって特に劣 Gauss である．

定理 6.5.29 (対称化不等式). 直積空間を $\Omega := (\mathcal{X}^n,\mathcal{B}^n,P^n)\times(\mathcal{Z},\mathcal{C},Q)\times(\mathcal{X}^n,\mathcal{B}^n,P^n)$ とする．この上での期待値を E で表す．任意の単調増加な凸関数 $\Phi:\mathbb{R}\to\mathbb{R}$ と，可測関数の族 $\mathcal{F}\subset\mathrm{Meas}(\Omega,\mathcal{X})$ に対して，

$$E^*[\Phi(\|\mathbb{P}_n-P\|_{\mathcal{F}})] \le E^*[\Phi(2\|\mathbb{P}_n^\circ\|_{\mathcal{F}})].$$ [†15]

[†15] 特に左辺の外積分は $(\mathcal{X}^n,B^n,P^n)$ に関して，右辺の外積分は $(\mathcal{X}^n,B^n,P^n)\times(\mathcal{Z},\mathcal{C},Q)$ に関するもの．

［証明］.

方針　確率変数 X_i は，の前半への射影 $\mathrm{pr}_i:\Omega\to\mathcal{X}$ とし，$Y_1,\cdots,Y_n$ を後半への射影で，$X_1,\cdots,X_n$ と同じもの（コピー）とする．$\epsilon_1,\cdots,\epsilon_n:\Omega\to\mathbb{Z}$ を互いに独立な Rademacher 確率変数とする．すると，射影は完全である（補題 6.2.22）から，この定理の主張は可測な拡大 $\Omega\to\Omega'$ に対して影響を受けないことに注意.

最初の評価　任意の $X_1,\cdots,X_n$ について，

$$\begin{aligned}\|P_n-P\|_{\mathcal{F}}&=\sup_{f\in\mathcal{F}}\frac{1}{n}\left|\sum_{i=1}^n(f(X_i)-E[f(Y_i)])\right| && \text{定義}\\ &\le E_Y^*\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\left|\sum_{i=1}^n(f(X_i)-f(Y_i))\right|\right] && f(Y_i)\text{ は可測でも，sup で包むと可測とは限らない}\end{aligned}$$

と評価できる．さらに

$$\begin{aligned}\Phi(\|P_n-P\|_{\mathcal{F}})&\le\Phi\left(E_Y^*\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\left|\sum_{i=1}^n(f(X_i)-f(Y_i))\right|\right]\right) && \Phi\text{の単調性}\\ &\le E_Y\left[\Phi\left(\left\|\frac{1}{n}\sum_{i=1}^n(f(X_i)-f(Y_i))\right\|_{\mathcal{F}}^{*Y}\right)\right] && \text{Jensen の不等式}\\ &=E_Y^*\left[\Phi\left(\left\|\frac{1}{n}\sum_{i=1}^n(f(X_i)-f(Y_i))\right\|_{\mathcal{F}}\right)\right] && \Phi\text{が連続の場合は等しい.}\end{aligned}$$

最後に $X_1,\cdots,X_n$ についても期待値を取ると，

$$E^*[\Phi(\|\mathbb{P}_n-P\|_{\mathcal{F}})]\le E_X^*\left[E_Y^*\left[\Phi\left(\left\|\frac{1}{n}\sum_{i=1}^n(f(X_i)-f(Y_i))\right\|_{\mathcal{F}}\right)\right]\right]$$

を得る．Fubini の定理 6.2.24 から，右辺はさらに重外積分 E_{XY}^* を超えない．

対称化による評価

■

6.5.4.2　可測性の問題

対称化不等式の右辺が可測になってくれると，応用上さらにうまくいく．そのためには，各点可測クラスを取れば良いことが分かる．

定義 6.5.30 (measurable class). $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ が，次を満たすとき，**P-可測クラス**であるという：$\forall_{n\in\mathbb{N}}\ \forall_{(e_1,\cdots,e_n)\in\mathbb{R}^n}$ 関数 $(X_1,\cdots,X_n)\mapsto\left\|\sum_{i=1}^n e_if(X_i)\right\|_{\mathcal{F}}$ が $(\mathcal{X}^n,\mathcal{B}^n,P^n)$ の完備化について可測.

要諦 6.5.31. 先に ϵ に関して，次に X に関して積分したいので，$\|\mathbb{P}_n^\circ\|_{\mathcal{F}}$ が $X_1,\cdots,X_n,\epsilon_1,\cdots,\epsilon_n$ について可測である時が一番理想的である．すなわち，任意の $(e_1,\cdots,e_n)\in\{\pm1\}^n$ に対して $(X_1,\cdots,X_n)\mapsto\left\|\sum_{i=1}^n e_if(X_i)\right\|_{\mathcal{F}}$ が可測であることを仮定する．なお，$\mathcal{X}$ の完備化に関して可測であれば（6.2.24 よりも強い意味での）Fubini の定理には十分.

定義 6.5.32 (pointwise measurable class). $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ が，次を満たすとき，**各点可測クラス**であるという：可算部分集合 $\mathbb{G}\subset\mathcal{F}$ が存在して，$\forall_{f\in\mathcal{F}}\ \exists_{\{g_m\}\subset\mathbb{G}}\ \forall_{x\in\mathcal{X}}\ g_m(x)\to f(x)$.

補題 6.5.33. $\mathcal{F}$ が各点可測ならば，$\mathcal{F}$ は任意の確率測度 P に関して P-可測クラスである.

［証明］. $\mathcal{F}$ が各点可測である時，$\left\|\sum_{i=1}^n e_if(X_i)\right\|_{\mathcal{F}}=\left\|\sum_{i=1}^n e_if(X_i)\right\|_{\mathbb{G}}$ が成り立つため.　■

例 6.5.34. 定義関数のクラス $\mathcal{F}=\left\{1_{(-\infty,t]}(x)\mid t\in\mathbb{R}\right\}$ は各点可測である.

実際，可算部分集合を

$$\mathcal{G} := \left\{1_{(-\infty,t]} \in \mathrm{Meas}(\mathcal{X},\mathbb{R}) \mid t \in \mathbb{Q}\right\}$$

と定めると，任意の $1_{(-\infty,t]} \in \mathcal{F}$ に対して，ある $\{s_n\}_{n\in\mathbb{N}} \subset \mathbb{Q}$ が存在して $\lim_{n\to\infty} s_n = t$ であるから，$\{1_{(-\infty,s_n]}\}_{n\in\mathbb{N}} \subset \mathcal{G}$ は f に各点収束する．

補題 6.5.35 (各点可測性の保存)**.** $\mathcal{F}_1,\cdots,\mathcal{F}_k \subset \mathrm{Meas}(\mathcal{X},\mathbb{R})$ を各点可測とする．この時，任意の連続関数 $\phi:\mathbb{R}^k \to \mathbb{R}$ に対して，クラス

$$\phi\circ(\mathcal{F}_1,\cdots,\mathcal{F}_k) := \{\phi\circ f(x) \mid f = (f_1,\cdots,f_k) \in \mathcal{F}_1\times\cdots\times\mathcal{F}_k\}$$

は各点可測である．

要諦 6.5.36. ほとんどの構成（特に可測関数の構成と同じような max や基本的な演算など）は連続である．

6.5.4.3　Donsker の定理で必要な可測性

実際に可測性が必要になるのは，$\mathcal{F}_\delta, \mathcal{F}^2_\infty$ なる集合であるので，これについての各論を展開する．

記法 6.5.37. クラス $\mathcal{F}$ に対して，

(1) $\mathcal{F}_\delta := \{f-g \in \mathrm{Meas}(\mathcal{X},\mathbb{R}) \mid f,g \in \mathcal{F}, \|f-g\|_{P,2} < \delta\}$ $(\delta \in (0,\infty])$.
(2) $\mathcal{F}^2_\infty := \left\{h^2 \in \mathrm{Meas}(\mathcal{X},\mathbb{R}) \mid h \in \mathcal{F}_\infty\right\}$.

と定める．

定義 6.5.38 (envelope function)**.** クラス $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X},\mathbb{R})$ の包絡関数とは，次を満たす関数 $F:\mathcal{X}\to\mathbb{R}$ である：$\forall_{x\in\mathcal{X}}\ \forall_{f\in\mathcal{F}}\ |f(x)| \le F(x)$.

例 6.5.39. $x \mapsto \sup_{f\in F}(f(x))$ は最小の包絡関数である．

命題 6.5.40. クラス $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X},\mathbb{R})$ について，$\mathcal{F}$ が各点可測で，そのある包絡関数 F が $P^*F^2 < \infty$ を満たすならば，全ての $0 < \delta \le \infty$ に対して $\mathcal{F}_\delta$ と F^2_∞ は各点可測である．

6.5.4.4　セミパラメトリック回帰モデルで必要な可測性

定義 6.5.41. 任意の $X = (Y,Z) \in \mathcal{X} := \mathbb{R}\times\mathbb{R}^k$ の分布 $P \in \mathcal{M}(\mathcal{X})$ に関して，クラス

$$\mathcal{F} := \left\{1_{\{Y-\theta^T Z\le t\}} \mid \theta \in \mathbb{R}^k, t \in \mathbb{R}\right\}$$

はセミパラメトリック回帰モデルの分析で大事になる．

命題 6.5.42. $\mathcal{X}$ 上の任意の確率測度 P について，クラス $\mathcal{F},\mathcal{F}_\delta,\mathcal{F}^2_\infty$ はいずれも各点可測である．

6.5.5　ブラケットエントロピー

道具３：

クラス $\mathcal{F}$ が Glivenko-Cantelli クラスであるか，または Donsker クラスであるかは，基本的にはその大きさに依存する．そこで，関数空間の部分集合の「大きさ」を定量化する方法を考える．有限集合は濃度で測れるが，無限集合はどうするか．これに対して Kolmogorov らロシアの学派が始めた取り扱いである．ノルムによるもの（球）とブラケットによるものの 2 つの定義があるが，この 2 つをつなぐのが Riesz の性質 $|f| \le |g| \Rightarrow \|f\| \le \|g\|$ である．

6.5.5.1　定義と性質

記法 6.5.43. $(\mathcal{F}, \|-\|)$ を，$\mathrm{Map}(X, \mathbb{R})$ 内のノルム空間のある部分集合とする．例えば確率測度 $Q \in \mathcal{M}(\mathcal{X})$ に関する $L_r(Q)$-ノルム空間など．

定義 6.5.44 (covering number, entropy without bracketing, ϵ-net)**.**

(1) $N(\epsilon, \mathcal{F}, \|-\|) := \min\left\{N \in \mathbb{N} \mid (U_n)_{n\in N} \text{は}\epsilon\text{-開球の族で } \mathcal{F} \text{ を被覆する}\right\}$．ただし，$\epsilon$-開球の中心が $\mathcal{F}$ に属する必要はない．
(2) $\log N(\epsilon, \mathcal{F}, \|-\|)$ を（ブラケットなし）エントロピーという．
(3) 部分集合 $\mathcal{G} \subset \mathcal{F}$ が ϵ-網であるとは，$\forall_{f\in\mathcal{F}}\ \inf_{g\in\mathcal{G}} \|f-g\| < \epsilon$ が成り立つことをいう．

要諦 6.5.45. ϵ-網というのは，$\mathcal{F}$ 内部に存在する「被覆する開球の中心として取れば良い点」となる．逆に，全ての開被覆に対して同じ濃度の ϵ-網が対応する．したがって，任意の ϵ-網に対して，同数で被覆するブラケットを構成できる算譜を示すことが基本的な手法となる．

定義 6.5.46 (bracket, bracketing number, bracketing entropy)**.**

(1) 関数の組 $l \le u \in \mathrm{Map}(X, \mathbb{R})$ に対して，集合 $[l, u] := \{f \in \mathrm{Map}(X, \mathbb{R}) \mid l \le f \le u\}$ をブラケットという．
(2) $\|u - l\| \le \epsilon$ を満たすブラケット $[l, u]$ を ϵ-ブラケットという．
(3) $\mathcal{F}$ を覆うために必要な ϵ-ブラケットの最小個数を**ブラケット数**といい，$N_{[]}(\epsilon, \mathcal{F}, \|-\|)$ で表す．l, u は $\mathcal{F}$ の元である必要はない．
(4) $\log N_{[]}(\epsilon, \mathcal{F}, \|-\|)$ をブラケットエントロピーという．

補題 6.5.47 (Riesz の性質)**.**

(1) $L_r(Q)$-ノルムは Riesz の性質を満たす：$|f| \le |g| \Rightarrow \|f\| \le \|g\|$.
(2) Riesz の性質を満たすノルム $\|-\|$ について，任意の $\epsilon > 0$ と任意のクラス $\mathcal{F} \subset \mathrm{Map}(\mathcal{X}, \mathbb{R})$ について $N(\epsilon, \mathcal{F}, \|-\|) \le N_{[]}(2\epsilon, \mathcal{F}, \|-\|)$.
(3) ノルム $\|-\|$ は一様ノルム $\|-\|_\infty$ に優越されるとする：$\forall_{f\in\mathcal{F}}\ \|f\| \le \|f\|_\infty$．この時，(2) の逆の不等式も成り立つ：$\forall_{\epsilon>0}\ N_{[]}(2\epsilon, \mathcal{F}, \|-\|) \le N(\epsilon, \mathcal{F}, \|-\|\infty)$.

［証明］.

(1) 1
(2) 2ϵ-ブラケット $[l, u]$ について，$f \in [l, u] \Rightarrow U_\epsilon\left(\dfrac{l+u}{2}\right)$.
(3) 任意の ϵ-網に対して，同じ個数で被覆するブラケットが構成できる算譜を示せば良い．$\mathcal{F}$ を被覆する一様 ϵ-網 $\{f_1, \cdots, f_m\}$ を考えると，$\|-\|$ に関する 2ϵ-ブラケット $[f_1-\epsilon, f_1+\epsilon], \cdots, [f_m-\epsilon, f_m+\epsilon]$ は $\mathcal{F}$ を覆う．

■

6.5.5.2　Sobolev クラス

補題 6.5.48.

6.5.5.3　Lipschitz 連続なクラスのブラケット

> 添字集合 T に関してある連続性条件を仮定すると，6.5.47(3) の性質を拡張できる．

定理 6.5.49. 距離空間 (T, d) 上のクラス $\mathcal{F} = \{f_t\}_{t\in T} \subset \mathrm{Map}(\mathcal{X}, \mathbb{R})$ を考える．ある関数 $F : \mathcal{X} \to \mathbb{R}$ が存在して，$\forall_{x\in\mathcal{X}}\ |f_s(x) - f_t(x)| \le F(x) \cdot d(s, t)$ が成り立つとする (添字に関する Lipschitz 連続性)．この時，任意の $\epsilon > 0$ と $\mathcal{F}$ 上の任意のノルム $\|-\|$ で $\|F\| < \infty$ を満たすものに関して $N_{[]}(2\epsilon\|F\|, \mathcal{F}, \|-\|) \le N(\epsilon, T, d)$ が成り立つ．

［証明］. (T, d) を被覆する ϵ-網 $t_1, \cdots, t_k$ に対して，$2\epsilon\|F\|$ 個で被覆するブラケットが必ず構成できることを示す．$[f_{t_1} - \epsilon F, f_{t_1} +$

$\epsilon F], \cdots, [f_{t_k} - \epsilon F, f_{t_k} + \epsilon F]$ とすれば良い． ■

要諦 6.5.50. そりゃそう．この本の基本手法が見えてきた，写像 F とか言って種々の手法を抽象化して抽象的な定理を示してから，具体的に使う．しかもその具体例は 1,2 個しかない．

注 6.5.51. $\|-\|$ が $\|-\|_\infty$ で優越されるとは，$T = \mathcal{F}, d = \|-\|_\infty, F = \mathrm{id}_{\mathcal{F}}$ の場合に当たる．

6.5.5.4　和と積に対する保存性

補題 6.5.52. クラス $\mathcal{F}, \mathcal{G} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ と確率測度 $Q \in \mathcal{M}(\mathcal{X})$ に対して，

(1) $\forall_{1\le r\le\infty}\ \forall_{\epsilon>0}\ N_{[]}(2\epsilon, \mathcal{F}+\mathcal{G}, L_r(Q)) \le N_{[]}(\epsilon, \mathcal{F}, L_r(Q))N_{[]}(\epsilon, \mathcal{G}, L_r(Q))$.

(2) $\sup_{f\in\mathcal{F}} \|f\|_\infty \le 1, \sup_{g\in\mathcal{G}} \|g\|_\infty \le 1$ が成り立つ時，$\forall_{1\le r\le\infty}\ \forall_{\epsilon>0}\ N_{[]}(2\epsilon, \mathcal{F}\cdot\mathcal{G}, L_r(Q)) \le N_{[]}(\epsilon, \mathcal{F}, L_r(Q))N_{[]}(\epsilon, \mathcal{G}, L_r(Q))$.

6.6　Glivenko-Cantelli 型定理

被覆数が有限とは，全有界と同じ雰囲気を感じる．$\mathcal{F}$ が本質的に有限だから成り立つ性質であるという消息を暴いている．

6.6.1　ブラケットナンバーの十分条件

定理 6.6.1. クラス $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ が $\forall_{\epsilon>0}\ N_{[\,]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$ を満たすとする．この時，$\mathcal{F}$ は P-Glivenko-Cantelli クラスである：$\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{\text{a.s.}^*} 0$.

例 6.6.2 (Glivenko-Cantelli の定理)**.** 経験分布関数 $\mathbb{F}_n : \overline{\mathbb{R}} \to [0,1]$ に関する一様収束性 $\|\mathbb{F}_n - F\|_\infty \xrightarrow{\text{a.s.}} 0$ は Glivenko-Cantelli の定理と呼ばれる．経験分布関数は，経験測度の $\mathcal{F} := \left\{1_{(-\infty,t]}(x) \in \mathrm{Meas}(X, \mathbb{R}) \mid t \in \overline{\mathbb{R}}\right\}$ への制限とみなせるから，$\|\mathbb{P}_n - P\|_{\mathcal{F}} \xrightarrow{\text{a.s.}} 0$ に同値．これについて定理を考えると，真の分布関数 F の高々有限個の ϵ 以上のジャンプを分点としてブラケットを取れば，$\mathcal{F}$ を被覆することができるから $N_{[\,]}(\epsilon, \mathcal{F}, L_1(P)) < \infty$．よって，確かに $\mathcal{F}$ は Gilvenko-Cantelli クラスである．

6.6.2　被覆エントロピーの十分条件

被覆数については，任意の $\epsilon > 0$ について有限であるだけでなく，n に関して緩増加であることも必要．

定理 6.6.3. クラス $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ が

(a) P-可測クラス 6.5.30 で，

(b) その包絡関数 F は $E^*[F] < \infty$ を満たすとする 6.5.38.

(c) $M > 0$ について，$\mathcal{F}_M := \left\{f 1_{\{F\le M\}} \mid f \in \mathcal{F}\right\}$ と定める．

(d) $\forall_{\epsilon>0}\ \forall_{M<\infty}\ \log N(\epsilon, \mathcal{F}_M, L_1(\mathbb{P}_n)) = o_P^*(n)$

ならば，次の 2 条件が成り立つ．

(1) $E\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \to 0$.

(2) $\|\mathbb{P}_n - P\|_{\mathcal{F}}^* \xrightarrow{\text{a.s.}} 0$.

特に，$\mathcal{F}$ は P-Glivenko-Cantelli クラスである．

［証明］．

方針　まず，$\Phi := \mathrm{id}_X$ に関する対称不等式と Fubini の定理より，

$$E^*\left[\|\mathbb{P}_n - P\|_{\mathcal{F}}\right] \le 2E_X\left[E_\epsilon\left[\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right\|_{\mathcal{F}}\right]\right] \quad \because P\text{-可測より外積分が取れる}$$

$$
\begin{aligned}
&\le 2E_X^*\left[E_\epsilon\left[\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i f 1_{\{F\le M\}}(X_i)\right\|_{\mathcal{F}}\right]\right] \\
&\qquad + 2E_X^*\left[E_\epsilon\left[\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i f 1_{\{F> M\}}(X_i)\right\|_{\mathcal{F}}\right]\right] && \text{三角不等式} \\
&\le 2E_X^*\left[E_\epsilon\left[\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right\|_{\mathcal{F}_M}\right]\right] \\
&\qquad + 2E^*\left[F1_{\{F>M\}}\right].
\end{aligned}
$$

仮定 (b) より，第二項は $M>0$ を十分大きく取ることで任意に小さくできる．第一項が任意の $M>0$ に対して 0 に収束することを示せば (1) が示せる．(2) は次の補題による．

第一項の評価　$E_\epsilon\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right\|_{\mathcal{F}_M}$ を評価する．仮定 (d) より，δ-net $\mathcal{G}\subset\mathcal{F}$ を取れば，

$$
E_\epsilon\left[\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i)\right\|_{\mathcal{F}_M}\right] \le E_\epsilon\left[\left\|\frac{1}{n}\sum_{i=1}^n \epsilon_i g(X_i)\right\|_{\mathcal{F}}\right] + \delta.
$$

■

補題 6.6.4. 確率空間 $(\mathcal{X}^\infty,\mathcal{B}^\infty,P^\infty)$ について，

(a) クラス $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ は包絡関数 F をもち，$E^*[F]<\infty$ を満たす．
(b) フィルトレーション Σ_n は，最初の n 個の変数の任意の置換について対称な全ての可測関数が生成する σ-加法族とする．$\Pi_n := \{\sigma\in\mathrm{Sym}(\mathbb{N}) \mid \forall_{i\ge n+1}\ \sigma(i)=i\}$ について，$H_n := \left\{h\in\mathrm{Meas}(\mathcal{X}^\infty,\mathbb{R}) \mid \forall_{\sigma\in\Pi_n}\ h((x_{\sigma(i)})_{i=1}^\infty)=h((x_i)_{i=1}^\infty),(x_i)\in\mathcal{X}^\infty\right\}$ としたとき，$\Sigma_n:=\sigma[h;h\in H_n]\subset\mathcal{B}^\infty$．

この時，

(1) $E[\|\mathbb{P}_n-P\|_{\mathcal{F}}^*|\Sigma_{n+1}]\ge\|\mathbb{P}_{n+1}-P\|_{\mathcal{F}}^*$ a.s.
(2) $\|\mathbb{P}_n-P\|_{\mathcal{F}}^*$ のバージョンで，この filtration に適合するもの$=\Sigma_n$-可測なものが存在する．
(3) このようなバージョンは全て逆進 sub-martingale で，ある P^∞ 可積分確率変数に確率 1 で収束する．

6.6.3　一様エントロピーの十分条件

経験分布だけでなく，（考え得る）あらゆる測度について被覆数を考える．すると，十分条件は少しは簡単になる．そしてこの十分条件は意外と使える．

定義 6.6.5 (uniform covering number).

(1) クラス $\mathcal{F}$ の包絡関数 F に対して，$\|F\|_{Q,r}>0$ を満たす有限離散確率測度 Q の全体を $\mathcal{Q}_{F,r}$ とする．
(2) $\sup_{Q\in\mathcal{Q}_{\mathcal{F},r}} N(\epsilon\|F\|_{Q,r},\mathcal{F},L_r(Q))$ を一様被覆数という．
(3) その対数を一様エントロピーという．

定理 6.6.6. 可測関数のクラス $\mathcal{F}$ は P-可測クラスで包絡関数 $F\in L^1(P)$ を持つとし，全ての $\epsilon>0$ に関して一様被覆数は有限とする：$\sup_{Q\in\mathcal{Q}_{F,1}} N(\epsilon\|F\|_{Q,1},\mathcal{F},\|-\|_{L^1(Q)})<\infty$．このとき，$\mathcal{F}$ は P-Gilvenko-Cantelli クラスである．

6.7 Donsker 型定理

11/8/2021

6.7.1 第一定理：一様エントロピー

定義 6.7.1 (uniform entropy integral). 包絡関数 F を持つ可測関数のクラス $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X},\mathbb{R})$ について，一様エントロピーの積分を**一様エントロピー積分**という：

$$J(\delta,\mathcal{F},L_r) := \int_0^\delta \sqrt{\log \sup_{Q\in\mathcal{Q}_{F,r}} N(\epsilon\|F\|_{Q,r},\mathcal{F},L_r(Q))}d\epsilon, \quad \delta\in(0,\infty].$$

補題 6.7.2. $\mathcal{X}$ 上の有限離散確率測度の全体を $\mathcal{Q}$ とする．

$$\forall_{0<\delta\le\infty} \int_0^\delta \sqrt{\log\sup_{Q\in\mathcal{Q}} N(\epsilon\|F_c\|_{Q,r},\mathcal{F},L_r(Q))}d\epsilon \le J(\delta,\mathcal{F},L_r)$$

記法 6.7.3. $\mathcal{F}_\delta := \{f-g\in\mathrm{Meas}(\mathcal{X},\mathbb{R}) \mid f,g\in\mathcal{F}, \|f-g\|_{P,2}<\delta\}$ と表す．

定理 6.7.4. 可測関数のクラス $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X},\mathbb{R})$ は包絡関数 $F\in L^2(P)$ をもち，$J(1,\mathcal{F},L_2)<\infty$ を満たすとする．任意の $0<\delta\le\infty$ について，クラス $\mathcal{F}_\delta, \mathcal{F}_\infty^2 := \left\{h^2\in\mathrm{Meas}(\mathcal{X},\mathbb{R}) \mid h\in\mathcal{F}_\infty\right\}$ は P-可測とする．このとき，$\mathcal{F}$ は P-Donsker クラスである．

6.7.2 第二定理：ブラケットエントロピー

定義 6.7.5 (bracketing integral). 可測関数のクラス $\mathcal{F}$ に対して，次を**ブラケット積分**という：

$$J_{[\,]}(\delta,\mathcal{F},L_r(P)) := \int_0^\delta \sqrt{\log N_{[\,]}(\epsilon,\mathcal{F},L_r(P))}d\epsilon, \quad \delta\in(0,\infty].$$

一様エントロピー積分の定義の際に必要な $Q\in\mathcal{Q}_{\mathcal{F},r}$ は考えず，真の測度 P のみを含む．

定理 6.7.6. 可測関数のクラス $\mathcal{F}$ は $J_{[\,]}(\infty,\mathcal{F},L_2(P))<\infty$ を満たすとする．このとき，$\mathcal{F}$ は P-Donsker クラスである．

補題 6.7.7. 2 乗可積分な有界関数 f からなる有限集合 $\mathcal{F}$ に対して，

$$E[\|\mathbb{G}_n\|_\mathcal{F}] \lesssim \max_{f\in\mathcal{F}} \frac{\|f\|_\infty}{\sqrt{n}}\log(1+|\mathcal{F}|) + \max_{f\in\mathcal{F}}\|f\|_{P,2}\sqrt{\log(1+|\mathcal{F}|)}.$$

［証明］. 任意の $x>0$ について，Bernstein の不等式 6.5.13 を

$$P(|\mathbb{G}_nf|>x) \le 2e^{-\frac{1}{2}\frac{x^2}{Pf^2+\|f\|_\infty x/\sqrt{n}}}$$

の形で利用する． ■

補題 6.7.8. 可測関数のクラス $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ は包絡関数 F をもち，全ての $f\in\mathcal{F}$ に対して $Pf^2<\delta^2$ を満たすとする．このとき，

$$a(\delta) := \frac{\delta}{\sqrt{1\vee\log N_{[\,]}(\delta,\mathcal{F},L_2(P))}}$$

に対して，

$$E^*[\|\mathbb{G}_n\|_\mathcal{F}] \lesssim J_{[\,]}(\delta,\mathcal{F},L_2(P)) + \sqrt{n}E^*[F1_{F>\sqrt{n}a(\delta)}].$$

注 6.7.9 (全ての Donsker クラスは Gilvenko-Cantelli クラスである). 系 6.4.13 より，Donsker クラスであることは $(\mathcal{F},d_P)$ が全有界で，$\forall_{\delta_n\searrow0}\ \|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}} \xrightarrow{P^*} 0$ に同値．後者の条件は $\forall_{\delta_n\searrow0}\ E^*[\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}}]\to0$ と同値．よって，Donsker 性の $\lim_{n\to\infty}x^2\sup P^*(\|\mathbb{G}_n\|_\mathcal{F}>x)=0$ より，$\forall_{0<r<2}\ E^*[\|f-Pf\|_\mathcal{F}^r]<\infty\wedge E^*[\|\mathbb{G}_n\|^r_\mathcal{F}]\to E[\|\mathbb{G}\|^r_\mathcal{F}]<\infty$ が成り立つ．よって，$\mathcal{F}$ は弱 P-Gilvenko-Cantelli クラスである．あとは補題 6.6.4 による．

6.7.3　例

例 6.7.10 (区間の定義関数クラス)**.** 経験分布関数 $\mathbb{F}_n(t):\mathbb{R}\to[0,1]$ は，経験測度 $\mathbb{P}_n:\mathcal{B}(\mathbb{R})\to[0,1]$ の定義関数のクラス $\mathcal{F}:=\left\{1_{(-\infty,t]}\in\mathrm{Meas}(\mathbb{R},[0,1])\mid t\in\overline{\mathbb{R}}\right\}$ への制限と同一視できるのであった．経験分布関数は有界な càdlàg 関数であるから，任意の $\epsilon\in(0,1]$ に対して，$\mathbb{R}$ の分割 $(t_i)_{i=0,\cdots,k}$ が存在して，$F(t_i-)-F(t_{i-1})\le\epsilon$ を満たす．この分割を用いて，ϵ-ブラケットの族を $([1_{(-\infty,t_{i-1}]},1_{(-\infty,t_i)}])_{i=0,\cdots,k}$ と定めると，それぞれの $i=0,\cdots,k$ について $\|1_{(-\infty,t_i)}-1_{(-\infty,t_{i-1})}\|_{L_1(P)}=F(t_i-)-F(t_{i-1})\le\epsilon$ を満たし，そのサイズ k は $k\le 1+\lfloor 1/\epsilon\rfloor\le 2/\epsilon$ などと評価できる（よって特に，$\mathcal{F}$ は P-Glivenko-Cantelli クラスである 6.6.1）．

第二定理からの確認　$0\le f\le 1$ のとき $0\le f^2\le f\le 1$ より，一般に $P[f^2]\le P[f]$ であるから，$L_1(P)$ ノルムについての ϵ-ブラケットは $\|1_{(-\infty,t_i)}-1_{(-\infty,t_{i-1})}\|_{L_2(P)}=(P[1^2_{(t_{i-1},t_i)}])^{1/2}\le\sqrt{\epsilon}$ より，そのまま $L_2(P)$ ノルムについての $\sqrt{\epsilon}$ ブラケットの例となっている．したがって，$N_{[\,]}(\epsilon,\mathcal{F},L_2(P))\le N_{[\,]}(\epsilon^2,\mathcal{F},L_1(P))\le\dfrac{2}{\epsilon^2}$．よって，ブラケット積分は，$\epsilon>1$ のときは $N_{[\,]}(\epsilon,\mathcal{F},L_2(P))=1$ であることに注意すると，

$$\begin{aligned}
J_{[\,]}(\infty,\mathcal{F},L_2(P))&=\int_0^\infty\sqrt{\log N_{[\,]}(\epsilon,\mathcal{F},L_2(P))}d\epsilon\\
&\le\int_0^1\sqrt{\log 2-2\log\epsilon}\,d\epsilon\\
&\le\int_0^{1/e}(\log 2-2\log\epsilon)d\epsilon+\int_{1/e}^1\sqrt{\log 2-2\log\epsilon}\,d\epsilon && \because \log 2-2\log\epsilon\ge\log 2+2>1\\
& && (0\le\epsilon\le 1/e)\\
&\le\frac{\log 2}{e}+\frac{4}{e}+2\left(1-\frac{1}{e}\right)=2+\frac{2+\log 2}{e} && \because 0\le\sqrt{\log 2-2\log\epsilon}\le\sqrt{\log 2+2}<2\\
& && (1/e\le\epsilon\le 1)
\end{aligned}$$

というように評価できる．よって，$\mathcal{F}$ は Donsker クラスでもある（定理 6.7.6）．

第一定理からの確認　また，明らかに $\forall_{f\in\mathcal{F}}\ \forall_{x\in\mathbb{R}}\ |f(x)|\le 1$ であるから，定数関数 $F:=1$ は $\mathcal{F}$ の可積分な包絡関数である．一様エントロピー積分については，一般の確率測度 Q に対して，$L_r(Q)$ $(1\le r\le\infty)$ ノルムは Riesz の性質 $|f|\le|g|\Rightarrow\|f\|\le\|g\|$ を満たすから，$N(\underbrace{\epsilon\|F\|_{Q,2}}_{=\epsilon},\mathcal{F},\|-\|_{L_r(Q)})\le N_{[\,]}(2\epsilon,\mathcal{F},\|-\|_{L_2(Q)})$ である 6.5.47 から，

$$\begin{aligned}
J(1,F,L_2)&=\int_0^1\sqrt{\log\sup_{Q\in\mathcal{Q}_{F,2}}N(\epsilon\|F\|_{Q,2},\mathcal{F},L_2(Q))}d\epsilon\\
&\le\int_0^1\sqrt{\log N_{[\,]}(2\epsilon,\mathcal{F},\|-\|_{L_2(Q)})}d\epsilon=J_{[\,]}(\infty,\mathcal{F},L_2(P))<\infty
\end{aligned}$$

と評価できる．$\mathcal{F}$ は各点可測 6.5.34 で，包絡関数が $P[F^2]<\infty$ を満たすから，$\forall_{\delta\in(0,\infty]}\ \mathcal{F}_\delta,\mathcal{F}^2_\infty$ は各点可測 6.5.40．よって，ほとんど等価な議論だが，定理 6.7.4 からも P-Donsker クラスであることがわかる．

例 6.7.11 (パラメータに関する Lipschitz クラス)**.** パラメータの空間 (Θ,d) によって添字づけられたクラス $\mathcal{F}:=\{f_\theta\}_{\theta\in\Theta}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ であって，ある $F:\mathcal{X}\to\mathbb{R}$ が存在して，Lipschitz 連続性

$$\forall_{\theta,\theta'\in\Theta}\ \forall_{x\in\mathcal{X}}\ |f_\theta(x)-f_{\theta'}(x)|\le d(\theta,\theta')F(x)$$

を満たすものを考える．

(1) ある $r\in[1,\infty]$ に対して，もし $\|F\|_{P,r}<\infty$ を満たすように取れるならば，定理 6.5.49 より

$$\forall_{\epsilon>0}\quad N_{[\,]}(2\epsilon\|F\|_{P,r},\mathcal{F},L_r(P))\le N(\epsilon,\Theta,d)$$

と評価できる．

(2) 例えば Θ を $\mathbb{R}^d$ 内の有界集合とすると，$N(\epsilon,\Theta,d)$ が有限であるから，(1) を使って $\mathcal{F}$ が Donsker であることを示せる．まず，Θ は 1 辺 ϵ の立方体を $\lceil\mathrm{diam}\Theta/\epsilon\rceil^d$ 個によって被覆できるので，$\forall_{\epsilon>0}\ \exists_{K\in\mathbb{R}}\ N(\epsilon,\Theta,d)\le K\left(\dfrac{\mathrm{diam}\Theta}{\epsilon}\right)^d$．よって，

$\exists_{K'\in\mathbb{R}}\ \forall_{0<\epsilon<\mathrm{diam}\Theta}\ N_{[\,]}(\epsilon,\mathcal{F},L_2(P))\le\dfrac{K'}{\epsilon^d}$ と評価でき，$\epsilon>\mathrm{diam}\Theta$ については $N_{[\,]}(\epsilon,\mathcal{F},L_2(P))=1$ だから，

$$\begin{aligned}
J_{[\,]}(\infty,\mathcal{F},L_r(P)) &= \int_0^\infty\sqrt{\log N_{[\,]}(\epsilon,\mathcal{F},L_r(P))}d\epsilon\\
&\le \int_0^{\min\{\mathrm{diam}\Theta,1\}}\sqrt{\log K-d\log\epsilon}d\epsilon\\
&\le \int_0^\alpha(\log K-d\log\epsilon)d\epsilon+\int_\alpha^1\sqrt{\log K-d\log\epsilon}d\epsilon \quad 0<\alpha<1\text{ は }\log K-d\log x=1\text{ の解とした}\\
&\le \alpha\log K-\underbrace{d\alpha\log\alpha}_{=\alpha(\log K-1)}+d\alpha+(1-\alpha)=1+\alpha d
\end{aligned}$$

よって，定理 6.7.6 より，Donsker であることが従う．

例 6.7.12 (Sobolev クラス)**.**

$$\mathcal{F}_k:=\left\{\eta:[0,1]\to\mathbb{R}\ \middle|\ \begin{array}{l}\|\eta\|_\infty\le1\text{ かつ }\eta^{(k-1)}\text{ は絶対連続で}\\ \int_0^1(\eta^{(k)}(x))^2dx\le1\text{ を満たす}\end{array}\right\}\quad k\in\mathbb{N}$$

は P-Donsker クラスである．

(1)

$$\exists_{K\in\mathbb{R}}\quad\forall_{\epsilon>0}\ \log N(\epsilon,\mathcal{F}_k,\|-\|_\infty)\le\frac{K}{\epsilon^{1/k}}$$

(2) 補題 6.5.48 より，一様ノルムについてはブラケット数は

$$\forall_{\epsilon>0}\quad\log N_{[\,]}(2\epsilon,\mathcal{F},\|-\|_\infty)\le\log N(\epsilon,\mathcal{F},\|-\|_\infty)\le K\epsilon^{-1/k}$$

と評価できるから，ブラケット積分は，

$$\begin{aligned}
J_{[\,]}(\infty,\mathcal{F},L_2(P)) &:= \int_0^\infty\sqrt{\log N_{[\,]}(\epsilon,\mathcal{F},L_2(P))}d\epsilon\\
&\le\sqrt{C}\int_0^1x^{-\frac{1}{2k}}dx+\int_1^\infty\sqrt{\log N_{[\,]}(\epsilon,\mathcal{F},L_2(P))}d\epsilon\\
&=\sqrt{C}\int_0^1x^{-\frac{1}{2k}}dx+\int_1^\infty\sqrt{\log N_{[\,]}(\epsilon,\mathcal{F},L_2(P))}d\epsilon
\end{aligned}$$

第一項は $\dfrac{2k}{2k-1}$ に収束し，第二項もブラケットエントロピーは十分大きな ϵ に関しては 0 なので有界である．定理 6.7.6 より，任意の $k\in\mathbb{N}$ に関して，$\mathcal{F}_k$ は Donsker であることが従う．

3 つの例に共通するパターン

最後の 2 つの例では被覆数を抑えてから，ブラケット数に関する条件に還元してから，第二定理 6.7.6 の方を用いている．ブラケット積分 $J_{[\,]}(\infty,\mathcal{F},L_2(P))$ の収束は，ブラケットエントロピーは十分大きい ϵ に関しては 0 であるから上に有界であることと，

$$\int_0^1x^\alpha dx=\begin{cases}\dfrac{1}{\alpha+1}, & -1<\alpha,\\ \infty, & \alpha\le-1.\end{cases}$$

より下にも有界であることを組み合わせて示す．

6.8 Glivenko-Cantelli クラスと Donsker クラスの構成

統計モデルとは，$\mathbf{x}$ を y に変換する関数の集合で，統計的な推論を行うために「とりあえず」考えておくものです．次の問題は，学習誤差から，予測誤差を見積もる事になる．これは大数の法則の漸近展開である．漸近展開ができないならば，大数の法則の精緻化である Hoeffding の不等式などで評価するしかない．最大不等式は最悪評価に使う．
次に推定量の収束スピードなど諸々がわかったのちには，モデル選択が問題になる．基本的にモデルが大きいと小さくなる学習誤差の項と，エントロピーの形をした項がある．
「表現力＝ VC 次元が高すぎると学習がうまく進まない」

記法 6.8.1. $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ に対して，$\mathcal{F}$ の列 $(f_m)_{m\in\mathbb{N}}$ の各点収束極限でもあり，$L_r(P)$ 収束極限でもあるような元に関する閉包を $\overline{\mathcal{F}}^{P,r}$ と表す．

6.8.1 Glivenko-Cantelli クラスの簡単な構成

命題 6.8.2. $\mathcal{F},\mathcal{G}$ を P-Glivenko-Cantelli クラスとする．

(1) $\mathcal{H}\subset\mathcal{F}$ は P-Glivenko-Cantelli である．
(2) $\mathcal{F}\cup\mathcal{G}$ は P-Glivenko-Cantelli である．
(3) $\mathcal{F}+\mathcal{G}$ は P-Glivenko-Cantelli である．
(4) $\overline{\mathcal{F}}^{P,1}$ は P-Glivenko-Cantelli である．

［証明］．

(1) 測度のノルムは $\|Q\|_{\mathcal{F}}=\sup_{f\in\mathcal{F}}|Qf|$ と定めたから，$\|\mathbb{P}_n-P\|_{\mathcal{H}}\le\|\mathbb{P}_n-P\|_{\mathcal{F}}\xrightarrow{\text{a.s.}^*}0$ より．
(2) $\|\mathbb{P}_n-P\|_{\mathcal{F}\cup\mathcal{G}}\le\max\{\|\mathbb{P}_n-P\|_{\mathcal{F}},\|\mathbb{P}_n-P\|_{\mathcal{G}}\}\xrightarrow{\text{a.s.}^*}0$.
(3) 積分の線形性より，$\sup_{f+g\in\mathcal{F}+\mathcal{G}}|Q(f+g)|=\sup_{f+g\in\mathcal{F}+\mathcal{G}}|Qf+Qg|\le\sup_{f\in\mathcal{F}}|Qf|+\sup_{g\in\mathcal{G}}|Qg|$ だから，$\|\mathbb{P}_n-P\|_{\mathcal{F}+\mathcal{G}}\le\|\mathbb{P}_n-P\|_{\mathcal{F}}+\|\mathbb{P}_n-P\|_{\mathcal{G}}\xrightarrow{\text{a.s.}}0$.
(4) $\sup_{f\in\mathcal{F}}|(\mathbb{P}_n-P)f|=\sup_{\widetilde{f}\in\overline{\mathcal{F}}^{P,1}}|(\mathbb{P}_n-P)\widetilde{f}|$ より．$\le$ は $\mathcal{F}\subset\overline{\mathcal{F}}^{P,1}$ より明らかで，逆は任意の $\widetilde{f}\in\overline{\mathcal{F}}^{P,1}$ に対して，これに各点収束しかつ $L_1(P)$ 収束する $\mathcal{F}$ の列 (f_n) が存在するため．

■

6.8.2 更なる構成

定理 6.8.3. クラス $\mathcal{F}_1,\cdots,\mathcal{F}_k\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ は P-Glivenko-Cantelli で，$\max_{1\le j\le k}\|P\|_{\mathcal{F}_j}<\infty$ を満たすとし，関数 $\phi:\mathbb{R}^k\to\mathbb{R}$ を連続とする．このとき，クラス

$$\phi\circ(\mathcal{F}_1,\cdots,\mathcal{F}_k)=\{\phi\circ f\in\mathrm{Meas}(\mathcal{X},\mathbb{R})\mid f:=(f_1,\cdots,f_k)\in\mathcal{F}_1\times\cdots\times\mathcal{F}_k\}$$

に可積分な包絡関数が存在するならば，P-Glivenko-Cantelli である．

系 6.8.4. $\mathcal{F},\mathcal{G}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ は可積分な包絡関数 F,G をそれぞれ持つ P-Glivenko-Cantelli クラスとする．

(1) $P[FG]<\infty$ ならば，$\mathcal{F}\cdot\mathcal{G}$ も P-Glivenko-Cantelli である．
(2) 値域の合併を $R:=\cup_{f\in\mathcal{F}}\mathrm{Im}\,f\subset\mathbb{R}$ とし，$\psi:\overline{R}\to\mathbb{R}$ をその閉包上に定まった連続関数とする．クラス $\psi\circ\mathcal{F}$ が可積分な包絡関数を持つならば，これは P-Glivenko-Cantelli である．

［証明］．$\mathcal{F},\mathcal{G}$ に可積分な包絡関数 F,G が存在するとき，$\|\mathcal{F}\|_{\mathcal{F}}=\sup_{f\in\mathcal{F}}|Pf|\le PF<\infty$ を満たすことに注意．

(1) FG を可積分な包絡関数として，積 $\cdot:\mathbb{R}^2\to\mathbb{R}$ を連続関数 ϕ と見れば良い．

(2) 関数 $\psi:\overline{R}\to\mathbb{R}$ の連続な延長 $\overline{\psi}:\mathbb{R}\to\mathbb{R}$ であって，$\overline{\psi}\circ F$ が可積分になるようなものを任意に 1 つ取れば良い (Tietze の拡張定理)．実際，$\forall_{f\in\mathcal{F}}\ \psi(f)=\overline{\psi}(f)$ より，$\overline{\psi}\circ\mathcal{F}=\psi\circ\mathcal{F}$ は可積分な包絡関数 $\overline{\psi}\circ F$ を持つ P-Glivenko-Cantelli クラスである．

■

6.8.3　Donsker クラスの簡単な構成

系 6.4.11 の特徴付けを用いて，初等的な結果を得る．

定義 6.8.5 (convex hull, symmetric convex hull). $\mathcal{F}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ に対して，

(1) $\mathrm{Conv}\mathcal{F}:=\left\{f=\sum_{i=1}^m\alpha_if_i\in\mathrm{Meas}(\mathcal{X},\mathbb{R})\;\middle|\;f_i\in\mathcal{F},\alpha_i\in\mathbb{R}_+,\sum_{i=i}^m\alpha_i\le 1\right\}$ を凸包という．

(2) $\mathrm{Sconv}\mathcal{F}:=\left\{f=\sum_{i=1}^m\alpha_if_i\in\mathrm{Meas}(\mathcal{X},\mathbb{R})\;\middle|\;f_i\in\mathcal{F},\alpha_i\in\mathbb{R},\sum_{i=i}^m|\alpha_i|\le 1\right\}$ を対称凸包という．

注 6.8.6. 通常の凸包の定義と違い，「原点も含んだ最小の凸集合」を定めている．A の対称凸包とは，原点について対称な集合 A' と，原点とを併せた $A\cup A'\cup\{0\}$ を含む最小の凸集合に等しい．

命題 6.8.7. $\mathcal{F}$ を P-Donsker クラスとする．

(1) $\mathcal{G}\subset\mathcal{F}$ は P-Donsker である．

(2) $\overline{\mathcal{F}}^{P,2}$ は P-Donsker である．

(3) $\overline{\mathrm{Sconv}\mathcal{F}}^{P,2}$ は P-Donsker である．

［証明］．

(1) (1) : クラス $\mathcal{F}$ が d_P に関して全有界ならば，$\mathcal{G}$ も全有界．(2) の漸近的同程度連続性も，$\forall_{\delta_n\searrow 0}\ \|\mathbb{G}_n\|_{\mathcal{G}_{\delta_n}}\le\|\mathbb{G}_n\|_{\mathcal{F}_{\delta_n}}\xrightarrow{P^*}0$ からわかる．

(2) **全有界**　$\mathcal{F}$ が全有界ならば $\overline{\mathcal{F}}^{P,2}$ も全有界．

漸近的同程度連続　クラス $\mathcal{F}$ は平均 0 のクラスであり，したがって $\rho_2=d_P$ であると仮定しても一般性は失われない．実際，$E[\mathbb{G}]=:c\in\mathbb{R}$ としたとき，$\mathcal{F}$ が Donsker であることと，$\mathcal{F}-c$ が Donsker であることは同値である．よって，ρ_P-セミノルムの代わりに，$L_2(P)$-ノルムで考える．

$\mathcal{G}\subset\mathrm{Meas}(\mathcal{X},\mathbb{R})$ の連続度を
$$w_{\mathcal{G}}(\delta):=\sup_{f,g\in\mathcal{G},\|f-g\|_{P,2}<\delta}|\mathbb{G}_n(f-g)|$$
と定める．$\forall_{\delta>0}\ w_{\overline{F}^{P,2}}(\delta)\le w_F(2\delta)$ を示せば，$\mathcal{F}$ の漸近的同程度連続性から $\overline{F}^{P,2}$ の漸近的同程度連続性が従う．任意に $\delta>0,\epsilon>0$ を取る．するとまず，連続度の定義より，$\|f-g\|_{P,2}<\delta$ を満たす $f,g\in\overline{\mathcal{F}}^{P,2}$ が存在して，
$$w_{\overline{F}^{P,2}}(\delta)-|\mathbb{G}_n(f-g)|<\frac{\epsilon}{3}$$
をみたす．次に，$\overline{F}^{P,2}$ は $L_2(P)$-閉包に含まれるから，$\|f-f_*\|_{P,2}<\dfrac{\delta}{2},\|g-g_*\|_{P,2}<\dfrac{\delta}{2}$ を満たす $f_*,g_*\in\mathcal{F}$ が存在して，
$$|\mathbb{G}_n(f-f_*)|<\frac{\epsilon}{3},\quad|\mathbb{G}_n(g-g_*)|<\frac{\epsilon}{3},$$
を満たす．このとき，$\|f_*-g_*\|_{P,2}\le 2\delta$ である．よって，
$$\begin{aligned}w_{\overline{F}^{P,2}}(\delta)-\mathbb{G}_n(f_*-g_*)&=\left|w_{\overline{F}^{P,2}}(\delta)-\mathbb{G}_n(f-g)+\mathbb{G}_n(f-f_*)+\mathbb{G}_n(g-g_*)\right|\\&\le\left|w_{\overline{F}^{P,2}}(\delta)-\mathbb{G}_n(f-g)\right|+|\mathbb{G}_n(f-f_*)|+|\mathbb{G}_n(g-g_*)|<\epsilon.\end{aligned}$$
すなわち，$w_{\overline{F}^{P,2}}(\delta)\le\mathbb{G}_n(f_*-g_*)+\epsilon\le w_F(2\delta)$．$\epsilon>0$ は任意としたから，$w_{\overline{F}^{P,2}}(\delta)\le w_F(2\delta)$ を得る．

(3) $\{\psi_i\}$ を $\mathcal{F}$ を含む $L_2(P)$ 部分空間 V の直交基底とし，$Z_1, Z_2, \cdots$ を標準正規分布に従う i.i.d. な確率変数とする．

(a) $\mathcal{F}$ は特に pre-Gaussian だから，$f \mapsto \sum_{i=1}^{\infty}(P(f-Pf)\psi_i)Z_i \in l^\infty(\mathcal{F})$ は一様収束極限で，あるタイトな Brown 橋を表現する．

(b) したがって，$\mathbb{G}_k := \sum_{i=1}^{k}(P(f-Pf)\psi_i)Z_i$ と定めると，$\|\mathbb{G}_k - \mathbb{G}_l\|_{\overline{\mathrm{Sconv}\mathcal{F}}} = \|\mathbb{G}_k - \mathbb{G}_l\|_{\mathcal{F}} \xrightarrow[k,l\to\infty]{\text{a.s.}}$ であるから，列 $(\mathbb{G}_k)_{k\in\mathbb{N}}$ は $l^\infty(\overline{\mathrm{Sconv}\mathcal{F}})$ 上でも至る所 Cauchy 列であるから，共通の極限 $\mathbb{G}$ に概収束する．各 $\mathbb{G}_k$ は線型汎函数で $\overline{\mathrm{Sconv}\mathcal{F}}$ 上 ρ_P-一様連続であるから，$\mathbb{G}$ も同じ性質を確率 1 で満たす．よって，$\overline{\mathrm{Sconv}\mathcal{F}}$ は pre-Gaussian である．

(c) 概表現定理 1.10.4 より，確率空間 $(\Omega, \mathcal{U}, P)$ と完全写像 $\phi_n : \Omega \to \mathcal{X}^n$ と $\phi : \Omega \to$ が存在して，

$$\left\|\frac{1}{\sqrt{n}}\sum_{i=1}^{n}(f(\phi_n(\omega)_i) - Pf) - \mathbb{G}(f, \phi(\omega))\right\|_{\mathcal{F}} \xrightarrow{\text{a.s.}^*} 0.$$

これは $\mathcal{F}$ を $\overline{\mathrm{Sconv}\mathcal{F}}$ としても成り立つ．すなわち，$\mathbb{G}$ という一様連続な見本過程に，$l^\infty(\overline{\mathrm{Sconv}\mathcal{F}})$ 上外概収束する経験過程のバージョンが存在する．これは $\overline{\mathrm{Sconv}\mathcal{F}}$ が Donsker であることに他ならない．

■

要諦 6.8.8. (1),(2) は $l^\infty(\mathcal{F})$ 上の $*$-弱収束の特徴付け 6.3.13（任意の周辺分布が収束し，漸近的同程度連続）からすぐに従う．議論は，新たなクラスでも連続度が変わらないことから従う．

$\mathcal{F}$ が pre-Gaussian とは，一様中心極限定理 $\mathbb{G}_n := \sqrt{n}(\mathbb{P}_n - P) \Rightarrow \mathbb{G} \in l^\infty(\mathcal{F})$ を満たすような確率変数 $\mathbb{G} : \Omega \to l^\infty(\mathcal{F})$ が存在する，すなわち，平均 0 共分散 $E[\mathbb{G}f\mathbb{G}g]$ の Gauss 過程は Kolmogorov の拡張定理により常に存在するが，そのタイトで Borle 可測なバージョン $\Omega \to l^\infty(\mathcal{F})$ がある確率空間 Ω 上に存在することをいう．$\mathcal{F}$ が pre-Gaussian であることは，$(\mathcal{F}, \rho_P)$ が全有界であり，かつ，$\mathbb{G}$ のバージョンであって見本過程 $f \mapsto \mathbb{G}(f)$ が一様 ρ_P 連続であるものが存在することに同値．$\mathcal{F}$ が Donsker であるためには（すなわち，その極限過程 $\mathbb{G}$ に実際に弱収束するには），$\mathcal{F}$ が pre-Gaussian であることに加えて，漸近的タイトであることが必要．2 つ併せると悪名高い漸近的同程度連続性となる．

定理 6.8.9 (almost sure representation theorem). $X_\alpha : \Omega_\alpha \to D$ をネットとし，極限 X_∞ を Borel 可測で可分な確率変数とする．$X_\alpha \Rightarrow X_\infty$ と弱収束するならば，ある確率空間 $(\overline{\Omega}, \overline{\mathcal{A}}, \overline{P})$ とネット $\overline{X}_\alpha : \overline{\Omega} \to D$ が存在して，

(1) $\overline{X}_\alpha \xrightarrow{\text{au}} X_\infty$，

(2) $\forall_{f\in l^\infty(D)}\ \forall\alpha \in \Lambda\ E^*[f(\overline{X}_\alpha)] = E^*[f(X_\alpha)]$.

6.8.4　更なる構成

定理 6.8.10. クラス $\mathcal{F}_1, \cdots, \mathcal{F}_k \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ は P-Donsker で，$\max_{1\le j\le k}\|P\|_{\mathcal{F}_j} < \infty$ を満たすとし，関数 $\phi : \mathbb{R}^k \to \mathbb{R}$ は

$$\forall_{f,g\in\mathcal{F}_1\times\cdots\times\mathcal{F}_k}\ \forall_{x\in\mathcal{X}}\ \exists_{c\in\mathbb{R}}\quad |\phi\circ f(x) - \phi\circ g(x)| \le c^2\sum_{i=1}^{k}(f_i(x) - g_i(x))^2$$

を満たすとする．このとき，$\exists_{f\in\mathcal{F}_1\times\cdots\times\mathcal{F}_k}\ \|\phi\circ f\|_{P,2} < \infty$ を満たすならば，クラス

$$\phi\circ(\mathcal{F}_1, \cdots, \mathcal{F}_k) = \{\phi\circ f \in \mathrm{Meas}(\mathcal{X}, \mathbb{R}) \mid f := (f_1, \cdots, f_k) \in \mathcal{F}_1\times\cdots\times\mathcal{F}_k\}$$

は P-Donsker である．

注 6.8.11. 条件 $\max_{1\le j\le k}\|P\|_{\mathcal{F}_j} < \infty$ が満たされるとき，系 6.4.13 より，$\mathcal{F}_i$ はそれぞれ 2-ノルム $\|-\|_{P,2}$ に関して全有界，特に有界であることに注意．

系 6.8.12. $\mathcal{F}, \mathcal{G}$ を P-Donsker クラスとする．

(1) $\mathcal{F}\cup\mathcal{G}, \mathcal{F}+\mathcal{G}$ は P-Donsker である．

(2) $\|P\|_{\mathcal{F}\cup\mathcal{G}} < \infty$ ならば，$\mathcal{F}\wedge\mathcal{G}, \mathcal{F}\vee\mathcal{G}$ は P-Donsker である．

(3) $\mathcal{F}, \mathcal{G}$ が共に一様有界ならば，$\mathcal{F}\cdot\mathcal{G}$ は P-Donsker である．

(4) $R := \cup_{f\in\mathcal{F}}\mathrm{Im}\,f$ の閉包上の関数 $\psi : \overline{R} \to \mathbb{R}$ は Lipschitz 連続であるとする．$\exists_{f\in\mathcal{F}}\ \|\psi\circ f\|_{P,2} < \infty$ ならば，$\psi\circ\mathcal{F}$ は P-Donsker である．

(5) $\|P\|_{\mathcal{F}} < \infty$ ならば，任意の有界な可測関数 g に対して，$\mathcal{F}\cdot g$ は P-Donsker クラスである．

［証明］．

(1) $f\in\mathrm{Meas}(\mathcal{X},\mathbb{R})$ に対して，$\dot{f} := f - Pf$ とおく．$\dot{\mathcal{F}} := \left\{\dot{f}\in\mathrm{Meas}(\mathcal{X},\mathbb{R}) \mid f\in\mathcal{F}\right\}$ と表すと．$\mathbb{G}_n\dot{f} = \mathbb{G}_nf - \mathbb{G}_n(Pf) = \mathbb{G}_nf$ より，$\mathcal{F}\cup\mathcal{G},\mathcal{F}+\mathcal{G}$ が Donsker であることと $\dot{\mathcal{F}}\cup\dot{\mathcal{G}},\dot{\mathcal{F}}+\dot{\mathcal{G}}$ が Donsker であることは同値．
   (a) いま，$\forall_{f\in\dot{\mathcal{F}}\cup\dot{\mathcal{G}}}\ P\dot{f} = Pf - P(Pf) = 0$ より，$\|P\|_{\dot{\mathcal{F}}} = \|P\|_{\dot{\mathcal{G}}} = \|P\|_{\dot{\mathcal{F}}\cup\dot{\mathcal{G}}} = 0$.
   (b) Lipschitz 関数 $+ : \mathbb{R}^2\to\mathbb{R}$ について，$+\circ(\dot{\mathcal{F}},\dot{\mathcal{G}})$ を考えれば良い．
   (c) $\|P\|_{\dot{\mathcal{F}}\cup\dot{\mathcal{G}}} = 0 < \infty$ であるから，$\dot{F}$ は $L_2(P)$-ノルムについて全有界，特に有界である（Donsker クラスの特徴付け 6.4.13）．よって，三角不等式より，$\|f+g\| \le \|f\| + \|g\| < \infty$ より，３つ目の条件も確かに満たされている．

   $\dot{\mathcal{F}}\cup\{0\}\subset\overline{\mathrm{Sconv}\dot{\mathcal{F}}}^{P,2}$ も Donsker である 6.8.7(3) ことより，$\dot{\mathcal{F}}\cup\dot{\mathcal{G}}\subset(\dot{\mathcal{F}}\cup\{0\}) + (\{0\}\cup\dot{\mathcal{G}})$ より，$\dot{\mathcal{F}}\cup\dot{\mathcal{G}}$ も $\mathcal{F}\cup\mathcal{G}$ も Donsker.

(2) (a) $\{\underset{\|P\|_{\mathcal{F}\wedge\mathcal{G}},\|P\|_{\mathcal{F}\vee\mathcal{G}}\le\|P\|_{\mathcal{F}\cup\mathcal{G}}<\infty}{\max}\}$.
   (b) $f = (f_1,g_1), g = (f_2,g_2)\in\mathcal{F}\times\mathcal{G}$ とし，$\phi(f(x)) = \max\{f_1(x),g_1(x)\}$ とおく．すると，$|\phi(f(x)) - \phi(g(x))|^2 \le \|f-g\|_{l^2}^2$ は明らか．
   (c) $\|\phi(f)\|_{P,2} = \max\|f_1\|_{P,2},\|g_1\|_{P,2} < \infty$.

(3) (a) $\exists_{c_1,c_2\in\mathbb{R}}\ c_1 < \mathcal{F}\cup\mathcal{G} < c_2$ より，$\max\{\|P\|_{\mathcal{F}},\|P\|_{\mathcal{G}}\} \le \max\{|c_1|,|c_2|\}$.
   (b) 乗算 $\cdot : \mathbb{R}^2\to\mathbb{R}$ は有界集合 $[c_1,c_2]^2$ に制限すると Lipschitz 連続である．
   (c) $\|\phi(f)\|_{P,2} \le \|f_1\|_{P,2}\|g_1\|_{P,2} \le \max\{|c_1|,|c_2|\}^2$

(4) a

(5) $f_1,f_2\in\mathcal{F}$ について，$|f_1(x)g(x) - f_2(x)g(x)| \le \|g\|_\infty|f_1(x)-f_2(x)|$ より，Lipschitz 連続性を満たし，$\|f\cdot g\|_{P,2} < \infty$ もみたす．

■

6.9　BUEI クラスと PE クラスの構成

有界な一様エントロピー積分を持つクラス (Bounded Uniform Entropy Integral) と各点可測 (Pointwise Measurable) なクラスの構成は並行に行える．前者は VC クラスを含む．VC クラスが Donsker であるためには，各点可測性さえ満たせば良い．

6.10　クラスの例とエントロピー評価

VC 理論は 60-90 に Vladimir Vapnik と Alexey Chervonenkis によって創られた統計的学習理論で，Dudley と Vapnik が後に経験過程論にも応用した．Vapnik はサポートベクトルマシンの発明者でもある．エントロピー評価によって，その経験過程がどのクラスに属するかが判る．

6.10.1　Vapnik-Červonenkis class

VC class という組み合わせ論的に定義されるクラスについては，被覆数の増大が多項式で抑えられるので，一様エントロピー積分の条件が満たされる．
集合系 $\mathcal{C}\subset P(X)$ が VC クラスであるとは，分解性能が高すぎないことをいう．

定義 6.10.1 (VC class: pick out, shatter, VC-index). 集合族 $\mathcal{C}\subset P(\mathcal{X})$ について，VC 次元 $V : P(\mathcal{X})\to\mathbb{N}\cup\{\infty\}$ を定義する．

$n \geq 1$ とする．

(1) 有限部分集合 $A \subset \{x_1, \cdots, x_n\} \subset \mathcal{X}$ について，$\exists_{C \in \mathcal{C}}\ A = C \cap \{x_1, \cdots, x_n\}$ を満たすとき，$\mathcal{C}$ は A を**抽出する**という．

(2) 各有限部分集合 $\{x_1, \cdots, x_n\}$ について，$\mathcal{C}$ によって抽出される部分集合の個数を $\Delta_n(\mathcal{C}; x_1, \cdots, x_n)$ で表す．$\Delta_n(\mathcal{C}; x_1, \cdots, x_n) = 2^n$ であるとき，$\mathcal{C}$ は $\{x_1, \cdots, x_n\}$ を**完全分解**するという．

(3)

$$\begin{aligned} V(\mathcal{C}) &:= \min\left\{n \in \mathbb{N} \mid \forall_{\{x_1,\cdots,x_n\} \subset \mathcal{X}}\ \Delta_n(\mathcal{C}; x_1, \cdots, x_n) < 2^n\right\} \\ &= \min\left\{n \in \mathbb{N} \mid \text{どの大きさ } n \text{ の有限部分集合 } \{x_1, \cdots, x_n\} \subset \mathcal{X} \text{ も } \mathcal{C} \text{ によって完全分解されない}\right\} \\ &= \sup\left\{n \in \mathbb{N} \mid \exists_{\{x_1,\cdots,x_n\} \subset \mathcal{X}}\ \Delta_n(\mathcal{C}; x_1, \cdots, x_n) = 2^n\right\} \end{aligned}$$

を **VC 次元**という．なお，右辺が空集合である時は $V(\mathcal{C}) = \infty$ とおく．

(4) $V(\mathcal{C}) < \infty$ を満たす $\mathcal{C}$ を **VC クラス**という．

例 6.10.2. $\mathcal{X} = \mathbb{R}^d$ の部分集合族 $\mathcal{C} = ((-\infty, c])_{c \in \mathbb{R}^d}$ の VC-指数は $d+1$，$\mathcal{C} = ((a, b])_{a,b \in \mathbb{R}^d}$ の VC-指数は $2d+1$ となる．ただし，$a = {}^t(a_1, \cdots, a_d), b = {}^t(b_1, \cdots, b_d)$ について $((a, b])$ とは，$(a, b] = (a_1, b_1] \times \cdots \times (a_d, b_d]$ を表す．

(1) $\mathcal{X} = \mathbb{R}$ の部分集合族 $\mathcal{C} := ((-\infty, c])_{c \in \mathbb{R}}$ を考えると，$n = 0, 1$ の場合は完全分解するが，$n = 2$ のとき，$\{x_1, x_2\} \subset \mathbb{R}$ $(x_1 < x_2)$ について，$\{x_2\} \subset \{x_1, x_2\}$ を抽出できない．よって，$V(\mathcal{C}) = 2$.

(2) $\mathcal{X} = \mathbb{R}$ の部分集合族 $\mathcal{C} := ((a, b])_{a,b \in \mathbb{R}}$ を考えると，$n = 0, 1, 2$ の場合は完全分解する．$n = 3$ の場合は $\{x_1, x_2, x_3\} \subset \mathbb{R}$ $(x_1 < x_2 < x_3)$ について，$\{x_2\} \subset \{x_1, x_3\}$ を抽出できない．よって，$V(\mathcal{C}) = 3$.

［証明］．

(1) n 元の部分集合 $\left\{\begin{pmatrix} x_1^1 \\ \vdots \\ x_d^1 \end{pmatrix}, \cdots, \begin{pmatrix} x_1^n \\ \vdots \\ x_d^n \end{pmatrix}\right\}$ であって，辞書式順序について $\begin{pmatrix} x_1^1 \\ \vdots \\ x_d^1 \end{pmatrix} < \cdots < \begin{pmatrix} x_1^n \\ \vdots \\ x_d^n \end{pmatrix}$ を満たすとする．$n \leq d$ のとき，

■

例 6.10.3.

(1) $\mathbb{R}^d$ の半空間の全体 $\mathcal{C} := \left\{x \in \mathbb{R}^d \mid \langle x, u \rangle \leq c\right\}_{u \in \mathbb{R}^d, c \in \mathbb{R}}$ は，$V(\mathcal{C}) = d+2$ を満たす VC-クラスである．

(2) $\mathbb{R}^d$ の閉球の全体 $\mathcal{C} := \left\{x \in \mathbb{R}^d \mid \|x - u\| \leq c\right\}_{u \in \mathbb{R}^d, c \in \mathbb{R}_{\geq 0}}$ は，$V(\mathcal{C}) \leq d+3$ を持つ VC-クラスである．

記法 6.10.4. $\mathcal{X}$ の可測集合の族 $\mathcal{C}$ に対応する定義関数のクラスを $1_{\{\mathcal{C}\}} := \{1_C(x) \in \mathrm{Meas}(\mathcal{X}, \mathbb{R}) \mid C \in \mathcal{C}\}$ と表す．また定理の証明中では，$\mathcal{C}$ 自体と適宜同一視する．

定理 6.10.5. ある実数 $K \in \mathbb{R}$ が存在して，任意の $\mathcal{C}$ が VC-クラスと，任意の確率測度 Q と実数 $r \geq 1, 0 < \epsilon < 1$ について，

$$N(\epsilon, 1_{\{\mathcal{C}\}}, L_r(Q)) \leq K \cdot V(\mathcal{C})(4e)^{V(\mathcal{C})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}.$$

例 6.10.6. $F := 1$ は $1_{\{\mathcal{C}\}}$ の可積分な包絡関数だから，

(a) 一様被覆数は $\forall_{\epsilon>0} \sup_{Q \in \mathcal{Q}_{F,1}} N(\epsilon\|F\|_{Q,1}, 1_{\{\mathcal{C}\}}, L_1(Q)) < \infty$ より，$1_{\{\mathcal{C}\}}$ は P-Glivenko-Cantelli である 6.6.6.

(b) 一様エントロピー積分について

$$\begin{aligned} J(1, 1_{\{\mathcal{C}\}}, L_2) &= \int_0^1 \sqrt{\log \sup_{Q \in \mathcal{Q}_{F,2}} N(\epsilon\|F\|_{Q,2}, 1_{\{\mathcal{C}\}}, L_2(Q))} d\epsilon \\ &\leq \int_0^1 \sqrt{\log K \cdot V(\mathcal{C})(4e)^{V(\mathcal{C})}(1/\epsilon)^{2(V(\mathcal{C})-1)}} d\epsilon \\ &= \int_0^1 \sqrt{\log K \cdot V(\mathcal{C})(4e)^{V(\mathcal{C})} + 2(V(\mathcal{C})-1)\log(1/\epsilon)} d\epsilon \\ &\leq \int_0^1 \log K \cdot V(\mathcal{C})(4e)^{V(\mathcal{C})} d\epsilon + \int_0^1 \sqrt{2(V(\mathcal{C})-1)\log(1/\epsilon)} d\epsilon \qquad \because \text{関数 } y = \sqrt{x} \text{は上に凸} \end{aligned}$$

$$= A + B\int_0^1 \sqrt{\log(1/\epsilon)}d\epsilon \qquad (A, B \in \mathbb{R})$$

と評価でき，$x := \log(1/\epsilon)$ と置換することにより，

$$\int_0^1 \sqrt{\log(1/\epsilon)}d\epsilon = \int_0^\infty x^{1/2}e^{-x}dx = \Gamma\left(\frac{3}{2}\right) = \frac{1}{2}\Gamma\left(\frac{1}{2}\right) = \frac{\sqrt{\pi}}{2}.$$

とわかるから，$J(1, 1_{\{\mathcal{C}\}}, L_2) < \infty$．よって，$(1_{\{\mathcal{C}\}})_\delta, (1_{\{\mathcal{C}\}})^2_\infty$ $(\delta \in (0, \infty])$ が P-可測ならば，$1_{\{\mathcal{C}\}}$ は Donsker でもある 6.7.4.

VC クラスは漸近的可測性に関する条件を満たせば，十分余裕を持って，普遍 Donsker である．

6.10.2　定理の証明

経験測度が任意の確率測度に収束することが大きい．議論を経験測度に限れば，これは所詮離散分布であるため，事象をうまく設定しなおせば一様測度に均すことが出来る．

補題 6.10.7 (Sauer's lemma).

(1) n 元集合 $\{x_1, \cdots, x_n\} \subset \mathcal{X}$ について，

$$\Delta_n(\mathcal{C}; x_1, \cdots, x_n) \le |\{A \in P(\{x_1, \cdots, x_n\}) \mid \Delta_m(\mathcal{C}; A) = 2^m, m \le n\}|$$

(2) (Sauer's lemma)

$$\max_{x_1, \cdots, x_n \in \mathcal{X}} \Delta_n(\mathcal{C}; x_1, \cdots, x_n) \le \sum_{j=0}^{V(\mathcal{C})-1} \binom{n}{j}.$$

(3)

$$\forall_{n \ge V(\mathcal{C})-1} \quad \max_{x_1, \cdots, x_n \in \mathcal{X}} \Delta_n(\mathcal{C}; x_1, \cdots, x_n) \le \left(\frac{ne}{V(\mathcal{C}) - 1}\right)^{V(\mathcal{C})-1}.$$

要諦 6.10.8. 今後度々使う手法として，「遺伝的集合」と呼ばれるような集合系 $\mathcal{C}$ に議論を還元する手法を整備する．遺伝的集合 $\mathcal{C}$ を符号化すると，$\{0,1\}^n$ に一対一対応させることができる．

［証明］．

(1) (a) $\mathcal{X}$ を $\mathcal{X}' := \{x_1, \cdots, x_n\}$ と取り直し，$\mathcal{C}$ を $\mathcal{C}' := \{C' \in P(\{x_1, \cdots, x_n\} \mid C' = C \cap \mathcal{X}', C \in \mathcal{C}\}$ と取り直すことより，任意の $A \subset \{x_1, \cdots, x_n\}$ について $A \cap C = A \cap C'$ なので，$\mathcal{C}$ が $\Delta_n(\mathcal{C}; x_1, \cdots, x_n) = |\mathcal{C}|$ を満たすと仮定しても，一般性を失わない．

(b) 集合系 $\mathcal{C}$ が遺伝的であるとき，すなわち，$\forall_{C \in \mathcal{C}}\ B \subset C \Rightarrow B \in \mathcal{C}$ をみたすとき，$\mathcal{C}$ の元はすべて完全分解されるから，不等式は必ず成立する．問題は一般の集合系 $\mathcal{C}$ についてであるが，濃度 $|\mathcal{C}|$ と，完全分解される $A \subset \{x_1, \cdots, x_n\}$ の数とを変えない集合系 $\mathcal{D} \subset P(\{x_1, \cdots, x_n\})$ であって遺伝的であるものに変換できることを示せば良い．

(c) 写像 $T_i : \mathcal{C} \to P(\{x_1, \cdots, x_n\})$ $(1 \le i \le n)$ を，

$$T_i(C) := \begin{cases} C \setminus \{x_i\}, & C \setminus \{x_i\} \notin \mathcal{C}, \\ C, & C \setminus \{x_i\} \in \mathcal{C}. \end{cases}$$

で定めると，これは単射である．また，任意の $A \subset \{x_1, \cdots, x_n\}$ について，A が $\mathcal{C}$ によって完全分解されることと，$T_i(\mathcal{C})$ によって完全分解されることは同値である．実際，$x_i \notin A$ ならば，T_i は $\mathcal{C}$ の各要素について x_i が属するかのみを変化させるので，$\forall_{C \in \mathcal{C}}\ C \cap A = T_i(C) \cap A$ より，A を完全分解する性質は変えない．$x_i \in A$ のとき，$\mathcal{C}$ が A を完全分解するならば，$\forall_{B \subset A}\ \exists_{C \in \mathcal{C}}\ C \cap A = B$ である．$C \setminus \{x_i\} \in \mathcal{C}$ のとき，$T_i(C) = C$ より，これが B を抽出する．$C \setminus \{x_i\} \notin \mathcal{C}$ のとき，$T_i(C) = C \setminus \{x_i\}$ で，$T_i(C) \cap A = B \setminus \{x_i\} \ne B$ であるが，A は $\mathcal{C}$ によって完全分解されるから，ある $C' \in \mathcal{C}$ が存在して，$C' \cap A = B \setminus \{x_i\}$．この C' は $x_i \notin C'$ より，$C' \setminus \{x_i\} = C' \in \mathcal{C}$ だから，$T_i(C') = C' \in \mathrm{Im}\ T_i$．よって，$T_i(\mathcal{C})$ も任意の $B \subset A$ を抽出する．逆に $T_i(\mathcal{C})$ が A を完全分解する時，$\forall_{B \subset A}\ \exists_{C \in \mathcal{C}}\ T_i(C) \cap A = B \cup \{x_i\}$ である．このとき $x_i \in T_i(C)$ だから，$T_i(C) = C$ かつ $C \setminus \{x_i\} \in \mathcal{C}$．よって，$C \cap A = B \cup \{x_i\}, (C \setminus \{x_i\}) \cap A = B \setminus \{x_i\}$．よって，$x_i \in B$ が成り立つかの如何に依らず，任意の A の部分集合 $B \subset A$ 抽出できる．

(d) したがって，一般の $\mathcal{C}$ が (1) を満たすことは，$T_1(\mathcal{C})$ が満たすことに同値．これを繰り返すことで，$T := T_n \circ \cdots \circ T_1$ について，$T(\mathcal{C})$ が満たすことに同値．$T : P(\{x_1, \cdots, x_n\}) \to P(\{x_1, \cdots, x_n\})$ は不動点 $\mathcal{D}$ を持ち，これは遺伝的になる．

(2) $V(\mathcal{C})$ 元以上の集合は，いずれも完全分解されない．したがって，$\min\{V(\mathcal{C}) - 1, n\}$ 元以下の集合の総数によって抑えられる．

(3) 次の評価と (2) より．$S := V(\mathcal{C}) - 1$ とおく．

$$\begin{aligned}
\sum_{i=0}^{S} \binom{n}{i} &\le \sum_{i=0}^{S} \binom{n}{i} \left(\frac{n}{S}\right)^{S-i} && \because S \le n \\
&\le \sum_{i=0}^{n} \binom{n}{i} \left(\frac{n}{S}\right)^{S-i} && \because S \le n \\
&\le \left(\frac{n}{S}\right)^{S} \sum_{i=0}^{n} \binom{n}{i} \left(\frac{S}{n}\right)^{i} \\
&\le \left(\frac{n}{S}\right)^{S} \left(1 + \frac{S}{n}\right)^{n} < \left(\frac{n}{S}\right)^{S} e^{S} && \because \text{二項定理}
\end{aligned}$$

■

要諦 6.10.9. $\Delta_n(\mathcal{C}; x_1, \cdots, x_n) < 2^n$ ならば，特に多項式オーダーで抑えられる．そもそも，完全分解可能な集合の最大濃度は多項式オーダーで増加することが組み合わせ論的に解っている．ここに二項係数が現れる理由は (1) にて，S 元集合の大きさ n 以下の冪集合の数を数え上げているためである．

補題 6.10.10. $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X}, \mathbb{R})$ は，ある経験測度 Q と関数 $g : (0,1) \to \mathbb{R}$ について，$\forall_{\epsilon \in (0,1)}\ D(\epsilon, \mathcal{F}, L_r(Q)) \le g(\epsilon)$ を満たすとする．このとき，任意の確率測度 Q について，同様の不等式が成り立つ．

［証明］．

パッキング数は有限である　任意の $\epsilon > 0$ と確率測度 P について，パッキング数は有限であるから，$D(\epsilon, \mathcal{F}, L_r(P)) = m \in \mathbb{N}$ とする．仮に有限でないとすると，任意の $f \in \mathcal{F}$ について，大数の強法則より，$\mathbb{P}_n|f|^r \xrightarrow{\text{a.s.}} P|f|^r$ より，$\forall_{f \in \mathcal{F}}\ \|f\|_{L_r(\mathbb{P}_n)} \to \|f\|_{L_r(P)}$ であるから，任意の $m \in \mathbb{N}$ に対して，十分大きな $n \in \mathbb{N}$ と $X_1, \cdots, X_n \in \mathcal{X}$ が存在して，$L_r(\mathbb{P}_n)$ ノルムについて，$\mathcal{F}$ では ϵ-分離的な m 点が取れる．ϵ-分離的な m 点 $f_1, \cdots, f_m$ について，$d := \min_{i \ne j \in [m]} (\|f_i - f_j\|_{L_r(P)} - \epsilon) > 0$ とし，$\max_{i,j \in [m]} |\|f_i - f_j\|_{L_r(\mathbb{P}_N)} - \|f_i - f_j\|_{L_r(P)}| < d$ $(N \ge n)$ となる n を取れば，この m 点 $f_1, \cdots, f_m$ はノルム $L_r(\mathbb{P}_n)$ が定める距離に関して引き続き ϵ-分離的である．しかしこれは，$D(\epsilon, \mathcal{F}, L_r(\mathbb{P}_n)) \le g(\epsilon) < \infty$ に矛盾．よって，パッキング数は有限．

証明　任意の確率測度 P について，$D(\epsilon, \mathcal{F}, L_r(P)) = m \in \mathbb{N}$ のとき，$m \le g(\epsilon)$ であることを示す．パッキング数の定義より，m 個の点 $f_1, \cdots, f_m \in \mathcal{F}$ が取れて，$P|f_i - f_j|^r > \epsilon^r$ $(i \ne j)$ を満たす．大数の強法則より，$\mathbb{P}_n|f_i - f_j|^r \xrightarrow{\text{a.s.}} P|f_i - f_j|^r$ であるから，ある $n \in \mathbb{N}, X_1, \cdots, X_n \in \mathcal{X}$ が存在して，$\mathbb{P}_{X_1, \cdots, X_n}|f_i - f_j|^r > \epsilon^r$ が成り立つ．したがって，$f_1, \cdots, f_m$ は $L_r(\mathbb{P}_n)$ ノルムについても ϵ-分離的である．よって，$m \le D(\epsilon, \mathcal{F}, L_r(\mathbb{P}_n)) \le g(\epsilon)$.

■

記法 6.10.11 (グラフ).

(1) 集合 A に対して，$[A]^k \subset P(A)$ を，A の部分集合であって k 元よりなるもの全体からなる集合とする．

(2) 有向グラフ $G = (V, E)$ に対して，$E(v)$ とは，点 $v \in V$ から出る辺全体からなる集合とする．

補題 6.10.12. $\mathcal{Z} \subset \{0,1\}^n$ を集合 $\{x_1, \cdots, x_n\}$ 上の VC クラス $\mathcal{C}$ の符号化（特性関数のグラフと同一視したもの）とする．

(1) $\mathcal{Z}$ が定めるグラフ $G = (V, E)$ には，任意の頂点に対して，多くとも $V(\mathcal{C}) - 1$ の辺が出ているような，すなわち $\forall_{z \in V}\ |E(z)| < V(\mathcal{C})$ を満たすように向きを入れることが出来る．

(2) $Z : \Omega \to \mathcal{Z}$ を確率変数とする．このとき，次が成り立つ：

$$\sum_{i=1}^{n} E[\mathrm{Var}[Z_i | Z_j, j \ne i]] \le V(\mathcal{C}) - 1.$$

要諦 6.10.13. (1) の組み合わせ論的な結果を使えば (2) の証明は速いが，(1) と独立にも示せる．

［証明］.

(1) 略．遺伝的な場合に議論を還元し，Hall の結婚定理を適用する．

(2) **分散の評価**　$v,w\in\mathfrak{X}$ は Hamming 距離が $1/n$ であるとする．すなわち，ある $i\in[n]$ が存在して，$v_i=0, w_i=1, v_j=w_j$ $(j\neq i)$ を満たすとき，分布 P の密度関数を $p(z):=P(Z=z)$ と表すとすると，確率変数 $Z_i:\Omega\to\{0,1\}$ は，$p:=\dfrac{p(w)}{p(v)+p(w)}$ についての Bernoulli 変数であるから，その分散は $\mathrm{Var}[Z_i|Z_j,j\neq i]=p(1-p)$.
仮に，Hamming 距離が $1/n$ となる 2 点が $\mathfrak{X}$ 内に存在しないとする．$m:=\min\{d(w,z)\in[0,1]\mid w,z\in\mathfrak{X},w\neq z\}$ とし，$d(v,w)=m$ とする．このとき，v,w で値が異なる nm 個の成分の集合を $J\subset[n]$ とすると（$\forall_{j\in J}\;v_j\neq w_j$），やはり $\mathrm{Var}[Z_i|Z_j,j\notin J]=p(1-p)$ となる．（この場合は $\mathrm{Var}[Z_i|Z_j,j\neq i]=0$ だから考えなくてよい？）

和の評価　$\mathfrak{X}\subset\{0,1\}^n$ は，$\{0,1\}^n$ を n-次元立方体のなすグラフとみると，その部分グラフ $G=(V,E)$ を，

$$E:=\left\{\{w,z\}\in[\mathfrak{X}]^2\mid d(w,z)=m\right\},\qquad V:=\{z\in\mathfrak{X}\mid\exists_{w\in\mathfrak{X}}\,d(w,z)=m\},$$

によって定める．$E_i:=\{(w,z)\in E\mid w_i\neq z_i\}$ $(i\in[n])$ とする．ただし，$[\mathfrak{X}]^2:=\{A\in P(\mathfrak{X})\mid |A|=2\}$ とした．
このとき，

$$\begin{aligned}\sum_{i=1}^n E[\mathrm{Var}[Z_i|Z_j,i\neq j]]\sum_{i=1}^n\sum_{\{v,w\}\in E_i}(p(v)+p(w))\frac{p(v)}{p(v)+p(w)}\frac{p(w)}{p(v)+p(w)}&=\sum_{i=1}^n\sum_{\{v,w\}\in E_i}\frac{p(v)p(w)}{p(v)+p(w)}\\&\leq\sum_{\{v,w\}\in E}p(v)\wedge p(w)\end{aligned}$$

と評価できるが，(1) より，$\forall_{z\in\mathfrak{X}}\,E(z)\leq V(\mathcal{C})-1$ を満たすような向きをグラフ G に入れられるから，これについて

$$\sum_{\{v,w\}\in E}p(v)\wedge p(w)=\sum_{z\in V}\sum_{(z,w)\in E}p(v)\wedge p(w)\leq V(\mathcal{C})-1$$

と評価できる．ただし，グラフ G について，$E(v)$ を点 v から出る辺全体の集合とした．
一方で，(1) の事実を使わなくとも証明できる．

(1) の別導出　(a)　$\mathcal{C}$ が遺伝的である場合を考える．V に定まる順序 $v<w:\Leftrightarrow\forall_{i\in[n]}\,v_i<w_i$ について，$(v,w)\in E:\Leftrightarrow v>w$ によって向きが定まり，$|E(v)|\leq V(\mathcal{C})-1$ を満たす．実際，$\mathcal{C}$ が遺伝的である場合は $m=1$ で $V=\mathfrak{X}$ となり，$<$ は V 上にたしかに順序を定め，また $\mathcal{C}$ は遺伝的であるから，$\{w\in\mathfrak{X}\mid(z,w)\in E\}\cup\{z\}\in\mathcal{C}$ は完全分解されるため，$|\{w\in\mathfrak{X}\mid(z,w)\in E\}\cup\{z\}|\leq V(\mathcal{C})$.

(b)

■

定理 6.10.14. ある実数 $K\in\mathbb{R}$ が存在して，任意の VC-クラス $\mathcal{C}$ と，任意の確率測度 Q と実数 $r\geq1, 0<\epsilon<1$ について，

$$N(\epsilon,1_{\{\mathcal{C}\}},L_r(Q))\leq K\cdot V(\mathcal{C})(4e)^{V(\mathcal{C})}\left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{C})-1)}.$$

要諦 6.10.15. 実は，$r=1$ で Q は経験測度の場合のみについて示せば十分である．経験測度は，これを定める標本 $X_1,\cdots,X_n$ に注目することにより，$[n]$ 上の一様分布と等価になり，この上に等長同型な空間 $(\widetilde{C},L_1(\widetilde{Q}))$ を構成できる．そこで，この自然数上の一様測度（離散空間 2^n 上の Hamming 距離に等価）に限って考えればよく，ここまで還元されるとパッキング数を使った組み合わせ論的な議論がしやすい．なぜならば，「抽出」が射影によって表現できることが非常に議論を簡明にする．評価の途中で分散に注目する手法の要諦はまだ見えない．

［証明］.

証明の方針

(1)　$1_{\{\mathcal{C}\}}$ の元である定義関数について，ノルムは

$$\forall_{C,D\in\mathcal{C}}\quad\|1_C-1_D\|_{Q,r}=\left(\int|1_C-1_D|^r dQ^{1/r}\right)=(Q(C\triangle D))^{1/r}$$

と表わせ，$\mathrm{Im}\,Q\subset[0,1]$ なので，r が大きくなるほど $\|1_C-1_D\|_{Q,r}$ は単調に増加する．したがって，$r=1$ の場合について示せば十分．

(2) 補題より，Q は任意の経験測度として示せば十分である．そこで，Q は $\{y_1,\cdots,y_k\}\subset\mathfrak{X}$ に，値 $Q(\{y_i\})=\dfrac{l_i}{n}$ $(l_i\in\mathbb{N}_{>0})$ を持つとする．このとき，$\mathcal{C}$ を $\mathcal{C}':=\{C'\in P(\{y_1,\cdots,y_k\})\mid C'=C\cap\{y_1,\cdots,y_k\},C\in\mathcal{C}\}$ と取り直すことで，$\mathcal{C}\subset P(\{y_1,\cdots,y_k\})$ であるとして証明すれば十分である．実際，この場合について $N(\epsilon,1_{\{\mathcal{C}'\}},L_1(Q))=N\in\mathbb{N}$ ならば，全く同じ N 個の ϵ-球について $1_{\{\mathcal{C}\}}$ も被覆されるから，$N(\epsilon,1_{\{\mathcal{C}\}},L_1(Q))=N$ である．以降，この意味での $\mathcal{C}'$ を $\mathcal{C}$ と表す．

(3) そこで，標本 $X_1,\cdots,X_n$ の重複を展開して，問題を添字集合である自然数の集合 $[n]:=\{1,2,\cdots,n\}$ 上に還元することを考える．[†16] $x_1,\cdots,x_n$ を，y_i を l_i 個含んだ多重集合とする．このとき，任意の $C\subset\{y_1,\cdots,y_k\}$ に対して，多重集合 $\widetilde{C}$ が，次のように写像 ϕ に関して定まる：写像 $\phi:[n]\to\{y_1,\cdots,y_k\}$ は，$\left|\phi^{-1}(y_i)\right|=l_i$ を満たすように任意に取る（標準的なものは 1 つさだまるが）．これに対して，逆像 $\widetilde{C}:=\phi^{-1}(C)$ とし，各 $x_i\in\mathfrak{X}$ は添字 $i\in[n]$ と同一視する．この対応は，写像 $\widetilde{-}:P(\{y_1,\cdots,y_k\})\to P([n])$ を定めるが，この写像による $\mathcal{C}\subset\{y_1,\cdots,y_k\}$ の像を，$\widetilde{\mathcal{C}}$ で表す．

(4) (3) での構成の仕方から，次が成り立つ：

　ある $J\subset[n]$ が $\widetilde{\mathcal{C}}$ によって完全分解されるならば，制限 $\phi|_J:J\to\{y_1,\cdots,y_k\}$ は単射であり，また像 $\phi(J)$ は $\mathcal{C}$ によって完全分解される．

この対偶命題を考えることより，$V(\widetilde{\mathcal{C}})\le V(\mathcal{C})<\infty$ で，$\widetilde{\mathcal{C}}$ も VC-クラスであることがわかる．また同様に構成より，$\mathfrak{X}$ 上の経験測度 Q に対応する $[n]$ 上の測度 $\widetilde{Q}$ は $[n]$ 上の一様測度で，$Q(C\triangle D)=\widetilde{Q}(\widetilde{C}\triangle\widetilde{D})$ が成り立つ．[†17] これにより，2 つの距離空間 $(\mathcal{C},L_1(Q)),(\widetilde{\mathcal{C}},L_1(\widetilde{Q}))$ は等長同型である．以上より，$N(\epsilon,\widetilde{\mathcal{C}},L_1(\widetilde{Q}))=N(\epsilon,\mathcal{C},L_1(Q))$ である．

(5) 記法と問題設定を整理する．

以降，$\widetilde{Q}$ を Q と同一視し，$\widetilde{\mathcal{C}}$ を $\mathcal{C}$ と同一視する．$\mathcal{C}\subset P([n])$ の元 $C\in\mathcal{C}$ は，特性関数 $1_C:[n]\to\{0,1\}$ を考えることで，$\{0,1\}^n=(\partial[0,1])^n$ の元と同一視出来る．したがって，$\mathcal{C}\subset P(\{y_1,\cdots,y_k\})$ は，$P([n])$ の部分集合とも，2^n の部分集合 $\mathcal{Z}$ とも見れる．以降特に，$\mathcal{C}$ を 2^n の部分集合と見たときの部分空間 $\mathcal{Z}$ を，各要素を列ベクトルに持つ $n\times\#\mathcal{C}$-行列と見る．行ベクトルの組を $J\subset[n]$ と指定した時，$\mathcal{Z}_J$ によって，J に含まれる行のみを残し，列ベクトルについても一次独立でない行は削除して得る行列とすると，$\mathcal{Z}_J$ は，射影 $\mathrm{pr}_J:\mathcal{Z}\hookrightarrow[0,1]^n\twoheadrightarrow[0,1]^J$ の像に他ならない．具体的には，

$$\mathcal{Z}_J=\{C\cap\{x_j\in\mathfrak{X}\mid j\in J\}\mid C\in\mathcal{C}\}$$

と表せる．このとき，次が成り立つ：

　$\mathcal{Z}_J$ の列が $2^{|J|}$ 個あるか，または，$\mathcal{Z}_J=(\partial\{0,1\})^J$ であるとき，$\mathcal{Z}_J$ は $\mathcal{C}\subset P(\{y_1,\cdots,y_k\})$ によって完全分解される．

$\mathcal{Z}_J$ 完全分解されるとき，$|J|<V(\mathcal{C})$ が必要であるから，$S:=V(\mathcal{C})-1$ とおくと $|J|\le S$ が成り立つ．

(6) $\{0,1\}^n$ 上に Hamming 距離（の正規化）$d(w,z):=\dfrac{1}{n}\displaystyle\sum_{i=1}^n|w_i-z_i|$ $(w,z\in 2^n)$ を定めると，$(\mathcal{Z},d)$ は $(\mathcal{C},L_1(Q)),(\widetilde{\mathcal{C}},L_1(\widetilde{Q}))$ のいずれとも等長同型である．実際，$C,D\in\mathcal{C}$ に対応する点を $z,w\in\mathcal{C}$ とすると，

$$\|1_C-1_D\|_{L_1(Q)}=Q(C\triangle D)=d(w,z).$$

証明　$\mathcal{C}$ の ϵ-分離的な最大の部分集合を取る[†18]．簡単のため，これを改めて $\mathcal{C}$ とし，$\mathcal{C}$ 自身を ϵ-分離的と考える．このとき，$D(\epsilon,\mathcal{C},L_1(Q))=|\mathcal{C}|$ を上から評価することで，パッキング数と被覆数の関係 6.5.17 より，同じ値が被覆数の上界となる．

(i) $Z:\Omega\to\mathcal{Z}$ を，$\mathcal{Z}$ 上の一様分布に従う確率変数とする．すると，各成分 $Z=(Z_1,\cdots,Z_n)$ $(Z_i\in\{0,1\})$ は，2 項分布 $B(n,1/2)$ に従う Bernoulli 変数となる．$n\le S$ であるとき，$\displaystyle\sum_{i=1}^n E[\mathrm{Var}[Z_i|Z_j,j\ne i]]\le n\le S$ は成り立つ．?

$S<n$ の場合も同様の不等式が成り立つことを示す．$S\le m<n$ を満たす整数 $m\in\mathcal{Z}$ を任意に取り，それに対して $|I|=m+1\le n$ を満たす部分集合 $I\subset[n]$ を任意に取る．補題より，確率変数 $Z_I:=\mathrm{pr}_I\circ Z:\Omega\to\mathcal{Z}_I$ について，

$$\sum_{i\in I}E[\mathrm{Var}[Z_i|Z_{I\setminus\{i\}}]]\le S$$

[†16] この n は経験測度を定める標本 $X_1,\cdots,X_n$ の数であることに注意．

[†17] 積み上げられたブロックをばらして一列にしたので，その上の測度は一様になる．これを狙って (3) のように構成した．

[†18] 距離を経験測度で測っているため，有限に取れてしまう．

が成り立つ．条件を満たす部分集合 $I \subset [n]$ の取り方は $\binom{n}{m+1}$ 通りあるから，それぞれの不等式を足し合わせて

$$\sum_{J \in P([n]), |J|=m} E\left[\sum_{i \notin J} \mathrm{Var}[Z_i | Z_J]\right] \leq \binom{n}{m+1} S$$

を得る．

(ii) 各 $\sum_{i \notin J} \mathrm{Var}[Z_i | Z_J = s]$ $(s \in \mathcal{Z}^J)$ を下から評価する．このときに $\#\mathcal{Z}$ なる項が最終的に左辺に残るようにすることで，最終的な不等式の証明に向かう．

$Z_J = s \in \mathcal{Z}^J$ のとき，確率変数 $Z : \Omega \to \mathcal{Z}$ は部分集合 $\mathcal{W} := \{z \in \mathcal{Z} \mid z_J = s\}$ 上に一様分布する．$|\mathcal{W}| =: N_s \in \mathbb{N}$ とおく．

いま，$W, \widetilde{W} : \Omega' \to \mathcal{W}$ を，同様に $\mathcal{W}$ 上の一様分布に従う独立な確率変数とすると，$\mathcal{Z}$ は ϵ-分離的としたから，$d(W, \widetilde{W}) > \epsilon$ $(W \neq \widetilde{W})$ が成り立つ．$W \neq \widetilde{W}$ が成り立つ確率は $1 - 1/N_S$ であることと，$\mathrm{Var}[Z_i | Z_J = s] = \mathrm{Var}[W_i] = E\left[\frac{(W_i - \widetilde{W}_i)^2}{2}\right]$ $(i \notin J)$ に注意すると，

$$\begin{aligned}
\sum_{i \notin J} \mathrm{Var}[Z_i | Z_J = s] &= \frac{1}{2} \sum_{i \notin J} E[(W_i - \widetilde{W}_i)^2] \\
&= \frac{1}{2} E\left[n \cdot d(W, \widetilde{W})\right] > \frac{1}{2} n \cdot \epsilon \left(1 - \frac{1}{N_s}\right)
\end{aligned}$$

と評価できる．

(iii) (ii) の式を，さらに $s \in \mathcal{Z}_J$ 上で期待値を取り，各 $J \in P([n]), |J| = m$ について足し合わせたものが，(i) の式の左辺である．一様分布 $P(Z_J = s) = \frac{N_s}{|\mathcal{Z}|}$ を仮定したから，(ii) の右辺は，集合 $\mathcal{Z}_J$ の元の数の平均を $\overline{|\mathcal{Z}_J|} = E[|\mathcal{Z}_J|]$ と表せば，

$$\begin{aligned}
\sum_{J \in P([n]), |J|=m} \sum_{s \in \mathcal{Z}_J} \frac{N_s}{|\mathcal{Z}|} \frac{1}{2} \epsilon n \left(1 - \frac{1}{N_s}\right) &= \sum_{J \in P([n]), |J|=m} \frac{1}{2} \epsilon n \sum_{s \in \mathcal{Z}_J} \left(\frac{N_s}{|\mathcal{Z}|} - \frac{1}{|\mathcal{Z}|}\right) \\
&= \binom{n}{m} \frac{1}{2} \epsilon n \left(1 - \frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}\right)
\end{aligned}$$

となる．

(iv) 以上，(i),(ii),(iii) の結果を併せると，

$$\begin{aligned}
& \binom{n}{m} \frac{1}{2} \epsilon n \left(1 - \frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}\right) < \binom{n}{m+1} S \\
\Leftrightarrow \quad & \frac{n!}{m!(n-m)!} \frac{1}{2} \epsilon n \left(1 - \frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}\right) < \frac{n!}{(m+1)!(n-m-1)!} S \\
\Leftrightarrow \quad & (m+1) \frac{\epsilon n}{2} \left(1 - \frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|}\right) < (n-m) S \\
\Leftrightarrow \quad & 1 - \frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|} < \frac{2(n-m)S}{(m+1)\epsilon n} \\
\Leftrightarrow \quad & \frac{\overline{|\mathcal{Z}_J|}}{|\mathcal{Z}|} > 1 - \frac{2(n-m)S}{(m+1)\epsilon n} = \frac{n(\epsilon m - 2S) + \epsilon n + 2mS}{(m+1)\epsilon n} \\
\Rightarrow \quad & |\mathcal{Z}| < \overline{|\mathcal{Z}_J|} \frac{\epsilon m}{|\epsilon m - 2S|} \cdot ()
\end{aligned}$$

と評価できる．最終行

(i) で狙ったとおり，$|\mathcal{Z}_J|$ とは，$\mathcal{Z}_J = \{C \cap \{x_j \in \mathcal{X} \mid j \in J\} \in P(\mathcal{X}) \mid C \in \mathcal{C}\}$ であったから，$\mathcal{C}$ によって $\{x_j\}_{j \in J}$ から抽出される部分集合の個数 $\Delta_{|J|}(\mathcal{C}; (x_j)_{j \in J})$ を表す．よって，$m \geq S$ に注意すれば，Sauer の補題より，

$$|\mathcal{Z}_J| \leq \sum_{j=0}^{S} \binom{m}{j} \leq \left(\frac{em}{S}\right)^S.$$

これと併せて，$|\mathcal{Z}|$ は，任意の整数 $m \in [S,n) \cap \mathbb{Z}$ に関して，

$$|\mathcal{Z}| \le \left(\frac{e}{S}\right)^S \frac{\epsilon m^S(m+1)}{\epsilon m - 2S}$$

と評価できる．特に $m = \dfrac{2(S+1)}{\epsilon}$（整数とは限らない）の場合を考えると，

$$\left(\frac{e}{S}\right)^S \frac{1}{2}\left(\frac{2(S+1)}{\epsilon}\right)^{S+1} \epsilon = \left(\frac{2e}{\epsilon}\right)^S (S+1)\left(\frac{1}{S}+1\right)^S < \left(\frac{2e}{\epsilon}\right)^{V(\mathcal{C})-1} V(\mathcal{C})e.$$

実際には m は整数であるが，$\dfrac{2(S+1)}{\epsilon}$ に最も小さい整数を取ることは，上式の右辺にある普遍定数倍をすることで変動を抑えられる．

■

6.10.3　VC-subgraph class

VC クラスは集合系に関する大きさの評価であった．関数空間の大きさを評価するために，subgraph を通じて VC 次元の概念を流入させ，$V : P(\mathrm{Meas}(\mathcal{X},\mathbb{R})) \to \mathbb{N} \cup \{\infty\}$ を定義する．

定義 6.10.16 (VC-subgraph class)**.**

(1) 可測関数のクラス $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X},\mathbb{R})$ が **VC-サブグラフクラス**であるとは，$f \in \mathcal{F}$ のサブグラフがなす集合族 $\{(x,t) \in \mathcal{X} \times \mathbb{R} \mid t < f(x)\}_{f\in\mathcal{F}}$ が VC-クラスであることをいう．

(2) 確率過程 $(f_t)_{t\in T} : T \to \mathrm{Meas}(\mathcal{X},\mathbb{R})$ が VC-サブグラフクラスであるとは，$\mathcal{F} := \{f_t\}_{t\in T}$ が VC-サブグラフクラスであることをいう．

例 6.10.17 (線型空間との関係)**.** $\mathcal{F} \subset \mathrm{Meas}(\mathcal{X},\mathbb{R})$ を有限次元線型部分空間とする．$\mathcal{F}$ は $V(\mathcal{F}) \le \dim \mathcal{F} + 2$ を満たす VC-サブグラフクラスである．

［証明］**.** $n := \dim \mathcal{F} + 2$ 個の点 $(x_1,t_1),\cdots,(x_n,t_n)$ を任意にとり，これが $\mathcal{F}$ のサブグラフによって完全分解されないことを示せば良い．

いま，$(f(x_1)-t_1,\cdots,f(x_n)-t_n)^T$ は $f \in \mathcal{F}$ に依らず，$\mathbb{R}^n$ のある $n-1$ 次元部分空間 V 内に入る．これは，$\mathcal{F}$ の基底 $f_1,\cdots,f_{n-2}$ を用いて，$(f_1(x_1),\cdots,f_1(x_n)),\cdots,(f_{n-2}(x_1),\cdots,f_{n-2}(x_n)),(t_1,\cdots,t_n)$ が生成する部分空間を取れば良い．このとき，$a \in V^\perp \setminus \{0\}$ を取ると，

$$\forall_{f\in\mathcal{F}} \quad \{(x_i,t_i) \in \mathcal{X}\times\mathbb{R} \mid a_i > 0\} \ne \{(x_i,t_i) \in \mathcal{X}\times\mathbb{R} \mid t_i < f(x_i)\}$$

が成り立ち，$\{(x_i,t_i) \in \mathcal{X}\times\mathbb{R} \mid a_i > 0\}$ で定まる点は $\mathcal{F}$ のサブグラフによって抽出できない．なぜならば，$a \perp V$ より，

$$\forall_{f\in\mathcal{F}} \quad \sum_{a_i>0} a_i(f(x_i)-t_i) = \sum_{a_i<0}(-a_i)(f(x_i)-t_i)$$

が成り立つが，仮に $\{(x_i,t_i) \in \mathcal{X}\times\mathbb{R} \mid a_i > 0\}$ が抽出できたとすると，左辺が正で，右辺が非正となり，矛盾． ■

例 6.10.18. $X : \mathbb{R} \supset T \to \mathrm{Meas}(\mathcal{X},\mathbb{R})$ を単調増加な確率過程とすると，これは $V(X) = 2$ の VC-サブグラフクラスである．

［証明］**.** $(x_1,t_1),(x_2,t_2) \in \mathcal{X}\times\mathbb{R}$ を任意に取る．この 2 点を X のサブグラフが完全分解するためには，特に，2 つの見本過程 $X_t(x_1), X_t(x_2) : T \to \mathbb{R}$ が $\mathbb{R}^2$ 上に定める関数 $(X_t(x_1),X_t(x_2)) : T \to \mathbb{R}^2; t \mapsto \begin{pmatrix} X_t(x_1) \\ X_t(x_2) \end{pmatrix}$ が，点 $(t_1,t_2) \in \mathbb{R}^2$ に対して，4 つの象限 $\left\{(x,y) \in \mathbb{R}^2 \mid x > t_1, y > t_2\right\}, \left\{(x,y) \in \mathbb{R}^2 \mid x > t_1, y < t_2\right\}, \left\{(x,y) \in \mathbb{R}^2 \mid x < t_1, y > t_2\right\}, \left\{(x,y) \in \mathbb{R}^2 \mid x < t_1, y < t_2\right\}$ のいずれをも通ることが必要．しかしこれは，見本過程 $T \to \mathbb{R}$ がいずれも T について単調であり，したがって $(X_t(x_1),X_t(x_2)) : T \to \mathbb{R}^2; t \mapsto \begin{pmatrix} X_t(x_1) \\ X_t(x_2) \end{pmatrix}$ のグラフも $\mathbb{R}^2$ 上で単調であることに矛盾する．したがって，$V(X) = 2$. ■

定理 6.10.19. ある実数 $K \in \mathbb{R}$ が存在して，任意の二乗可積分な包絡関数 F を持つ VC-サブグラフクラス $\mathcal{F}$ と，実数 $r \ge 1, 0 < \epsilon < 1$ と，$\|F\|_{Q,r} > 0$ を満たす任意の確率測度 Q について，

$$N(\epsilon\|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \le K \cdot V(\mathcal{F})(16e)^{V(\mathcal{F})} \left(\frac{1}{\epsilon}\right)^{r(V(\mathcal{F})-1)}.$$

［証明］. $\mathcal{C} := \{C_f \subset \mathcal{X} \times \mathbb{R} \mid C_f = \{(x,t) \in \mathcal{X} \times \mathbb{R} \mid t < f(x)\}, f \in \mathcal{F}\}$ とおく.

$r = 1$ のとき　λ を $\mathbb{R}$ 上の Lebesgue 測度とすると，Fubini の定理より，任意の $\mathcal{X}$ 上の確率測度 Q に対して，

$$Q|f - g| = (Q \times \lambda)(C_f \triangle C_g).$$

よって，$P := \dfrac{Q \times \lambda}{2 \cdot QF}$ と定めると，$\mathcal{Y} := \{(x,t) \in \mathcal{X} \times \mathbb{R} \mid |t| \le F(x)\}$ 上の確率測度となり（包絡関数は $F \ge 0$ であることに注意すると，$Q \times \lambda(\mathcal{Y}) = 2 \cdot QF$），定める L_1-距離は $\dfrac{1}{2 \cdot QF}$ 倍になる．よって，定理より，

$$N(\epsilon 2QF, \mathcal{F}, L_1(Q)) = N(\epsilon, \mathcal{C}, L_1(P)) \le KV(\mathcal{F}) \left(\frac{4e}{\epsilon}\right)^{V(\mathcal{F})-1}$$

一般の r のとき　R を，密度関数 $\dfrac{F^{r-1}}{QF^{r-1}}$ が定める $\mathcal{X}$ 上の確率測度とすると，三角不等式より $|f - g| \le |f| + |g| \le 2F$ だから，

$$Q|f - g|^r \le Q|f - g|(2F)^{r-1} = 2^{r-1} R|f - g| QF^{r-1}$$

と表現できるから，新たな確率測度 R についての L_1-ノルムによって

$$\|f - g\|_{Q,r} \le 2(QF^{r-1})^{1/r} \|f - g\|_{R,1}^{1/r}$$

と評価できる．$RF = \displaystyle\int_{\mathcal{X}} F dR = \int_{\mathcal{X}} \frac{F^r}{QF^{r-1}} dQ = \frac{QF^r}{QF^{r-1}}$ であることに注意すると，

$$\frac{R|f - g|}{Q|f - g|^r} \ge \frac{RF}{2^r QF^r}$$

より，$L_r(Q)$-ノルムで半径 $\epsilon 2\|F\|_{Q,r}$ である球は，$L_1(R)$-ノルムでいうと半径は $\epsilon^r RF$ より小さい．よって，

$$N(\epsilon 2\|F\|_{Q,r}, \mathcal{F}, L_r(Q)) \le N(\epsilon^r RF, \mathcal{F}, L_1(R)).$$

よって，$1/\epsilon$ の指数を r 倍すれば，定理が成り立つ．

■

定理 6.10.20. 統計モデル $\mathcal{F}$ を用いて学習を行うとき，任意の確率分布に対して予測誤差 $L(\widetilde{f})$ が $\inf_{f \in \mathcal{F}} L(f)$ に一様に確率収束することと，$\mathrm{VC}(\mathcal{F})$ が有限であることとは同値.

要諦 6.10.21. なんでも説明できるような反証不可能な仮説 $\mathrm{VC}(\mathcal{F}) = \infty$ は，科学的仮説とはいえず，予測能力は持たない」ということを表していると考えられる．分解能とトレードオフ？

6.10.4　滑らかな関数族

VC クラスは組み合わせ論的な議論で被覆数に関する結果を得たが，解析的な議論でブラケット数に関する結果を得る手法もある．

6.10.5　単調関数族

6.11　確率過程の Chaining

Chaining は一般の確率過程 $(X_t)_{t\in T}$ に関して，一様な上界を得る方法である．経験過程の枠に囚われずに考えてみたい．統計的学習理論では，被覆数などの幾何学的性質よりも，さらに組み合わせ論的な VC 次元が応用に近い．さらに評価を精密にし logarithmic gap を消すためには，より幾何学的な Talagrand 汎関数 $\gamma_2(T)$ を考える，これを"generic chaining"という．

6.11.1　Dudley の不等式

定義 6.11.1 (sub-gaussian increments)**.** 距離空間 (T,d) 上の確率過程 $(X_t)_{t\in T}$ が**劣ガウス増分**を持つとは，

$$\exists_{K\geq 0}\ \forall_{s,t\in T}\ \|X_t-X_s\|_{\psi_2}\leq Kd(t,s)$$

を満たすことをいう．

第 7 章

推測理論

第 8 章

機械学習

機械学習とは，統計推測 with an attitude である．

8.1 データセットと教師あり学習

定義 8.1.1 (data set).

(1) データセット $\mathscr{D}$ とは，データ生成分布（母集団）[†1]からサンプリングされた確率変数の実現値の集合をいう．$(x_n, y_n) \sim P(x, y)$ である時，$\mathscr{D} := \{(x_b, y_n)\}_{n=1}^{N}$ となる．

(2) $N := |\mathscr{D}|$ をデータセットサイズという．

(3) 変数 x, y は実験者の意味論から，入力変数と出力変数と呼び分けられることがある．

(4) データ生成分布を記述するのに用いる関数の集合を**仮設集合**や**モデル**という．パラメトリックモデルとは，有限の添字により族とみなせるモデルである．

定義 8.1.2 (supervised learning, over learning / over fitting, generalization, inductive bias).

(1) **訓練データ**と呼ばれる有限集合 D_{train} から，$P(y|x)$ を統計的に推定する問題をいう．

(2) この問題に対する訓練の失敗として，訓練データのみにしか使えない場合を**過学習**や**過剰適応**という．実際訓練データは有限だから，組み合わせ論的な解は存在する．

(3) 一方で理想的な状況を**汎化**という．こちらが人間にとって本当に解きたい問題であり，数学の存在意義でもある．[†2]

(4) 未知のデータに対して行う何らかの仮定を**帰納バイアス**という．

(5) 誤差関数に罰金 (penalty) 項を付加することにより係数が大きな値になることを防ぐことを**正則化**という．[†3]

定理 8.1.3 (no free lunch theorem).

注 8.1.4 (Transformer). Transformer というモデルは，万能帰納バイアスとして注目を集めている．万能であるということは人間がモデルの中に仕込んだ帰納バイアスが弱いということであり，そのため Transformer を訓練するためには大量のデータを要する．しかしながら CNN (Convolutional Neural Network) などの他の強い帰納バイアスとは異なり Transformer には，データや計算資源のスケールアップに従い単調に性能が向上するというべき則が成立すると考えられている．

[†1] ただし，独立同分布に従うとする．

[†2] 現実に使われている NN はパラメタ数がデータより非常に多く，一見すると自由度が高すぎて過学習が起きやすいモデルに思えます．それにも関わらず，なぜ NN の学習は上手くいくのでしょうか？実はパラメタ数が多くなっても，学習された NN は過学習を起こしにくいことが実験的に示唆されています．例えば，Neyshabur ら [1] による実験では，NN のパラメタ数が増えても未知のデータに対する予測精度は保たれていて，過学習が防げていることが分かります．Zhang ら [2] による別の実験でも， 2-正則化などの明示的な正則化がなくても学習後の NN の性能は良いことが示されています．これらの結果は，パラメタ数が多く明示的に正則化が行われない状況でも，何らかの形で過学習が抑制される正則化がかかっていることを示唆しています．https://tech.preferred.jp/ja/blog/implicit-bias/

[†3] 暗黙的な正則化（implicit bias）の正体は何なのでしょうか？ Neyshabur ら [1] は最適化アルゴリズムの性質によるものだという仮説を提示しました．例えば，確率的勾配降下法 (Stochastic Gradient Descent; SGD) は連続的にパラメタを更新していくアルゴリズムなため，初期値からあまり離れることができません．そのため，初期値が非常に小さい場合は，学習されたパラメタのノルムが小さくなると期待されます（図 1 ）．この小ノルム性が正則化として機能し，未知のデータに対する汎化性能に効いているのだというのが彼らの仮説です．Zhang ら [2] も同様に SGD に起因する小ノルム性に基づいた議論を行っています．

第 9 章

有効推定

参考文献

[1] 久保木久孝・鈴木武『セミパラメトリック推測と経験過程』(朝倉書店 統計ライブラリー，2015)
[2] 吉田朋広『数理統計学』(朝倉書店，2006)
[3] 竹村彰道『現代数理統計学』（学術図書，2020）
[4] 久保川達也『現代数理統計学の基礎』（共立出版，2017）
[5] 渡辺澄夫『代数幾何と学習理論』（森北出版，2006）
[6] Bickel - Efficient and Adaptive Estimation for Semiparametric Models (1993) 厳密な議論がなされている．実証科学者には難しすぎるとのことで Tsiatis が書かれた．
[7] Pfanzagl - Estimation In Semiparametric Model
[8] Michael R. Kosorok "Introduction to Empirical Processes and Semiparametric Inference" (Springer, 2008)
[9] Anastasios A. Tsiatis "Semiparametric Theory and Missing Data" (2006) 最初の 5 章は欠測データのないセミパラメトリックモデルを扱い，パラメータに対する推定量の構成法を議論する．特に幾何的な手法を見る．
[10] van der Vaart and Wellner "Weak Convergence and Empirical Processes"
[11] Van der Vaart, A. W. (2000). Asymptotic statistics, volume 3. Cambridge university press.
[12] Tsybakov - Introduction to Nonparametric Estimation
[13] Semiparametric Theory