A1　確　率　論

目　次

(A1-2)

§1. 測度論的確論

（〔1〕第1章，第2章，〔3〕，〔10〕，〔11〕，第1章参照）

甲乙2人が表が出るまで交互に銅貨を投げつづけ，先に表を出した方を勝とするゲームを考えてみる。最初に甲が投げることとし，表が出ることをH，裏が出ることをTで表わせば，例えば“1回目に表が出て甲が勝つ”場合はHで，“最初甲が裏を出し次に乙が表を出して乙が勝つ”場合はTHで“甲乙共に裏を出し3回目に甲が表を出して甲が勝つ”場合はTHHというように表わすことができる。このような表わし方をするとゲーム終了までに起り得るどんな場合も

$$\omega_n = \underbrace{TT\cdots\cdots T}_{(n-1)回}H$$

なる形で表わすことができる。またこのような形のω_nが与えられればそれによってゲームの経過がわかり，nの奇，偶により甲，乙，何れの勝ちであるかも判定できる。そこでω_nをゲームの一つの見本と見なしてこれを“見本点”（*Sample point*）とよび見本点全体の作る空間

$$\Omega = \{\omega_n = \underbrace{TT\cdots\cdots T}_{(n-1)回}H\ ;\ n = 1,\ 2,\ \cdots\cdots\}$$

を“見本空間”（*sample space*）という。このとき例えば“n回目まで勝負がつく”ということは“見本点がnヶ以下文字で表わされてる”という見本点に関する条件で表わすことができる。そこで見本点に関する条件を“事象”（*event*）という。次に事象に対応する確率を考えてみる。銅貨を1回投げたとき表又は裏の出る確率と共に$\frac{1}{2}$とすれば“1回投げただけで甲が勝つ”確率は$\frac{1}{2}$とするのが妥当である。そこでこの事象に対応する見本点ω_1に確率$P(\omega_1)$として$P(\omega_1)=\frac{1}{2}$を与える。ω_2，ω_3，……に対しても同様に考えて$P(\omega_n)=\frac{1}{2n}$とする。また“n回目迄に勝負がつく”確率としては$P(\omega_1)+P(\omega_2)+P(\omega_n=1-\frac{1}{2n}$を考えるのが妥当であろう。それはいうまでもなく“$n$回目までに勝負がつく”ということは，“見本点$\omega_1$，$\omega_2$……，$\omega_n$の何れかで表わされる状態のどれかが起っている”ことと同等であり，しかもこれ等の見本点はそれぞれ異なるゲーム経過に対応するものであって，同時に2つ以上の見本点に対応するゲーム経過は起り得ないからである。同様な考察から，“いつか勝負がつく”確率は$P(\omega_1)+P(\omega_2)+\cdots\cdots P(\omega_n)+\cdots\cdots=1$を，“$n$回目迄には勝負がつかない（（$n+1$）回目以後に勝負の

つく)”確率は $1-\{P(\omega_1)+P(\omega_2)+\cdots\cdots+P(\omega_n)\}=\frac{1}{2}\ (=P(\omega_{n+1})+P(\omega_{n+2})+\cdots\cdots)$ とすればよい。一般にある事象(条件) $\mathcal{E}$ に対してはEをみたす見本点の全体 $E\ (\subset\Omega)$ を考え

$$P(\mathcal{E})=\sum_{\omega_n\in E}P(\omega_n)$$

を“$\mathcal{E}$のおこる(なりたつ)”確率とすればよい.これらの事情を一般化して測度論的取扱いの可能な確率の定義を与えたのは A. Kolmogorov である。次にその定義を述べる.

Ωを抽象空間とし、$\mathcal{B}$をΩの部分集合からなるボレル集合体(Borel-field), Pを$\mathcal{B}$で定義された集合函数で次の条件をみたすものとする.

1.1) $A\in\mathcal{B}\Rightarrow 0\leqq P(A)\leqq 1$

1.2) $P(\Omega)=1$

1.3) $A_n\in\mathcal{B}\ (n=1.2,\cdots\cdots),\quad A=\sum_n A_n\Longrightarrow P(A)=\sum_n P(A_n)$

このPを$\Omega(\mathcal{B})$上の“確率測度”(Probability measure)又は“確率分布”(Probability distribution)といい、特にΩがk次元ユークリッド空間R^kで$\mathcal{B}$がR^kのボレル集合の全体$\mathcal{B}^k$のとき,Pを“k次元の分布”又は“R^k一分布”という。そしてΩ, $\mathcal{B}$, Pを組にして考えたものを“確率空間”(probability space)といい$(\Omega,\mathcal{B},P)$又は$\Omega(\mathcal{B},P)$とかく。$\mathcal{B}$に属する集合を“可測集合”(measurable set)という.またΩの元のωに関する条件$\mathcal{E}$を事象といい.特に“$\mathcal{E}$をみたすωの全体”Eが可測集合のとき$\mathcal{E}$を“可測事象”(measurable event)という.確率測度Pが可測集合に対してのみ定義されているので,確率論では可測事象のみを考察の対象とし,可測事象のことを単に事象ともいう.ωに対して条件$\mathcal{E}$がなりたつとき“事象$\mathcal{E}$が起る”又は“事象$\mathcal{E}$がなりたつ”等という.事象$\mathcal{E}$に対し“$\mathcal{E}$のみたすωの全体”Eを考えると,$\mathcal{E}$が可能ならばEは$\mathcal{B}$に属し、逆に$\mathcal{B}$に属するEをとれば“ωがEに属する”という条件即ち可測事象$\mathcal{E}$が考えらるが,$\mathcal{E}$とEがこのような関係にあるとき$\mathcal{E}$をEで表わされる事象という.Eの余集合E^cで表わされる事象を$\mathcal{E}$の“余事象”(Complementary event)とよぶ.$E_n\ (n=1,2,\cdots\cdots)$を有限個または無限個の可測集合とし、それらで表わされる事象を$\mathcal{E}_n\ (n=1,2,\cdots\cdots)$とする.こととき$E_n$の和集合$\bigcup_n E_n$で表わされる事象を$\mathcal{E}_n$の“和事象”といい,$E_n$の交集合$\bigcap_n E_n$で表わされる事象を$\mathcal{E}_n$の“積事象”という.また$E\cap F=\phi$(空集合)なる2つの可測集合$E$,$F$で表わされる事象を$\mathcal{E}$,

$\mathcal{F}$を"排反事象"（exclusive events, disjoint events）といい、ϕで表わされる事象を"空事象"、Ωで表わされる事象を"全事象"という。Eで表わされる事象$\mathcal{E}$に対し、Eにおける確率測度Pの値（P（E)を"事象$\mathcal{E}$の起る確率"といい，$P(\mathcal{E})$とかく。このような立場で確率論を論じると、事象に関することはすべて集合の言葉で表わせるので、可測集合Eと可測事象$\mathcal{E}$を同一視して考える方が便利であり、特に区別を必要とするとき以外は、可測集合Eを事象と見做して"事象E"ともいう。

1.1—1.3）により，P（空事象）＝0，P（全事象）＝1であり

1.4）　$E_1 \subset E_2 \Longrightarrow P(E_1) \leqq P(E_2)$

がなりたつ。これを"確率の単調性"（monotony）という。また有限個または無限個の排反事象 E_n（$n=1, 2, \cdots\cdots$）の和事象を $E=\sum_n E_n$ とすれば

1.5）　$P(E)=\sum_n P(E_n)$

がなりたつ。これを"確率の加法性"（additivity）という。特に$\Omega = E+E^c$であるから，$P(E)+P(E^c)=1$　である。一般に$P(E)=1$のとき，事象Eを"殆んどすべてのωに対しなりたつ事象"という。

§2 確率変数，確率ベクトル及びその分布と諸量

（〔2〕第3章，〔10〕第1章参照）

確率空間をΩ（$\mathcal{B}$，P）とするとき，Ωからk次元ユークリッド空間R^kへの写像 $x(\omega)=(x_1(\omega), x_2(\omega), \text{-----}, x_k(\omega))$ が $\mathcal{B}$－可測のとき，即ちR^kの任意のボレル集合 A に対し

2.1）　$\{\omega ;\ x(\omega)\in A\}\in\mathcal{B}$

のとき x を“k次元確率ベクトル”（random vector）または“R^k－値確率変数”（R^k－valued random variable）といい，特にkが1のときは単に“確率変数”という。このときk次元確率ベクトルxの可測性から，R^kのボレル集合Aに対し“$x(\omega)$が A に属する”という条件は可測事象である．Pは$\mathcal{B}$上の測度であるからPに関する Lebesgue 式積分を考えることができるが，特に確率変数 x が P に関して可積分のとき，その積分値を“xの平均値”（expectation，mean）といい $E(x)$ またはm_xで表わす。即ち

2.2）　$m_x\equiv E(x)\equiv\int_\Omega x(\omega)\,dP(\omega)$

また

2.3）　$E((x-m_x)^2)=\int_\Omega (x(\omega)-m_x)^2\,dP(\omega)$

が存在するとき，これを“xの分散”（variance）といい，$V(x)$または σ_x^2で表わす．x，yが平均値の存在する確率変数ならば，任意の定数α，βに対して（$\alpha x+\beta y$）の平均値も存在して

2.4）　$E(\alpha x+\beta y)=\alpha E(x)+\beta E(y)$

がなりたち，また分散が存在すれば

2.5）　$V(\alpha x+\beta)=\alpha^2 V(x)$

がなりたつ。一般に確率変数xのp乗がPに関し可積分ならば

2.4）　$E(|x|^p)\int_\Omega |x(\omega)|^p\,dP(\omega)$，

2.5）　$E(x^p)=\int_\Omega x^p(\omega)\,dP(\omega)$

をそれぞれ x の“p次の絶対能率”（p－th absolute moment）およ

び“p次の能率”（p-th moment）という。また

$$2.6)\quad E(|x-m_x|^p)=\int_\Omega |x(\omega)-m_x|^p\,dP(\omega),$$

$$2.7)\quad E((x-m_x)^p)=\int_\Omega (x(\omega)-m_x)^p\,dP(\omega)$$

をそれぞれのxの“平均値のまわりのp次の絶対能率”および“平均値のまわりのp次の能率”という．$1\leqq p<q$のときのq次の能率が存在すれば，p次の能率も存在して次の不等式がなりたつ．

$$2.8)\quad 1\leqq p<q \Longrightarrow E(|x|^p)^{\frac{1}{p}}\leqq E(|x|^q)^{\frac{1}{q}},$$

$$E(|x-m_x|^p)^{\frac{1}{p}}\leqq E(|x-m_x|^q)^{\frac{1}{q}}$$

また能率が存在すれば任意の$\varepsilon\,(>0)$に対して

$$2.9)\quad P(|x|\leqq\varepsilon)\leqq\frac{E(|x|^p)}{\varepsilon^p}$$

がなりたつ。これを Markov の不等式という．特に$p=2$のときにxの代りに$(x-m_x)$を考えると

$$2.10)\quad P(|x-m_x|\geqq\varepsilon)\leqq\frac{\sigma_x}{\varepsilon^2}$$

がなりたつ。これを Tchebychev の不等式という．

確率ベクトル$x=(x_1, x_2, \cdots\cdots, x_k)$の場合には、$m_j=E(x_j)\ (j=1, 2, \cdots\cdots, k)$が存在するとき$m=(m_1, m_2, \cdots\cdots m_k)$を$x$の“平均値ベクトル”（mean vector）といい，さらに

$$2.11)\quad v_{s,t}=E((x_s-m_s)(x_t-m_t)),\quad s,t=1,2,\cdots,k$$

が存在するとき

$$2.12)\quad v=(v_{s,t}),\quad s,t=1,2,\cdots\cdots,k$$

をxの“分散行列”（variance matrix）という。vが存在すれば、それは（広義の）正定符号行列である。また$\rho_{s,t}=v_{s,t}/\sqrt{v_{s,s}\,v_{t,t}}$を“$x_s$と$x_t$の間の相関係数”（correlation coefficient）という．

確率変数xに対し次式で定義されたR^1上の複素数値函数

$$2.13)\quad \varphi(\xi;x)=E(e^{ix\xi})=\int_\Omega e^{ix(\omega)\xi}\,dP(\omega),\quad（iは虚数単位）$$

を"xの特性函数"（*characteristic function*）という．確率ベクトル $x=(x_1, x_2, \cdots\cdots, x_k)$ の特性函数は、R^k の点 $\xi=(\xi_1, \xi_2, \cdots\cdots, \xi_k)$ に対し $\langle x, \xi\rangle = x_1\xi_1 + x_2\xi_2 + \cdots\cdots + x_k\xi_k$ とおくとき

2．14） $\varphi(\xi; x) = E(e^{i\langle x,\xi\rangle}) \int_{\Omega} e^{i(\sum_{j=1}^{k} x_j(\omega)\xi_j)}\, dP(\omega)$

で定義される。

xをk次元確率ベクトルとし、R^kの任意のボレル集合Aに対し $\Phi(A)$ を

2．15） $\Phi(A) = P(\{\omega; x(\omega) \in A\})$

で定義するとΦはk次元分布になっている．このΦを"xの分布"という．

Φをk次元分布とするとき次のm'_j（$j=1, 2, \cdots\cdots, k$）が

2．16） $m'_j = \int_{R^k} u_j\, d\Phi(u_1, u_2, \cdots\cdots, u_k)$

存在するとき $m' = (m'_1, m'_2, \cdots\cdots, m'_s)$ を"Φの平均値ベクトル"といい，

2．17） $v'_{s,t} = \int_{R^k} (u_s - m'_s)(u_t - m'_t)\, d\Phi(u_1, u_2, \cdots\cdots, u_k)$,

$s, t = 1, 2, \cdots\cdots, k$

が存在するとき，$v' = (v'_{s,t})_{s,t=1,2,\cdots\cdots,k}$ を"Φの分数行列"という。また $\xi = (\xi_1, \xi_2, \cdots\cdots, \xi_k)$ に対し

2．18） $\varphi(\xi; \Phi) = \int_{R^k} e^{i\langle u,\xi\rangle}\, d\Phi(u) = \int_{R^k} e^{i(\sum_{j=1}^{k} u_j\xi_j)}\, d\Phi(u_1, u_2, \cdots\cdots u_k)$,（$i$は虚数単位）

で定義されたR^k上の複素数値函数を"Φの特性函数"というがΦとφは1対1に対応し次の関係がなりたつ．$f(u; a, b)$を(a, b)の定義函数（ただし $f(a; a, b) = f(b; a, b) = \frac{1}{2}$ とする）とすると任意の $a_j < b_j$（$j = 1, 2, \cdots\cdots, k$）に対し

2・19） $\int_{R^k} \prod_{j=1}^{k} f(u_j; a_j, b_j)\, d\Phi(u_1, u_2, \cdots\cdots, u_k)$

$$= \lim_{C\to\infty} \int_{-C}^{C}\cdots\int_{-C}^{C} \prod_{j=1}^{k} \frac{e^{-ib_j\xi_j} - e^{-ia_j\xi_j}}{-i\xi_j}\, \varphi(\xi; \Phi)\, d\xi_1 \cdots d\xi_k$$

ただし $\xi = (\xi_1, \xi_2, \cdots\cdots, \xi_k)$

これを <u>P. Lévy の"反転公式"</u>（inversion formula）という．また X の分布が Ξ のときは、X の平均値ベクトル，分散行列，特性函数はそれぞれ Ξ の平均値ベクトル，分散行列，特性函数と一致する．

§3 独立性 （〔5〕第5章，〔9〕第2章参照）

有限個の事象の系 $\{E_n;\ n=1,2,\cdots\cdots N\}$ の任意の $1\leqq i<j<\cdots\cdots<k\leqq N$ に対して

3・1）　$P(E_i\cap E_j\cap\cdots\cdots\cap E_k)=P(E_i)P(E_j)\cdots\cdots P(E_k)$

をみたしているとき、E_n（$n=1,2,\cdots\cdots,N$）は"<u>互いに独立</u>"（mutually independent）であるといい，また $\{E;\ n=1,2,\cdots\cdots,N\}$ を"<u>有限独立事象系</u>"（fuinte family of independent events）という．無限事象系 $\{E_\lambda;\ \lambda\in\Lambda\}$ の任意の有限部分系が独立のときに $\{E_\lambda;\ \lambda\in\Lambda\}$ を"無限独立事象系"という．このとき有限ヶまたは無限ヶの独立事象系 $\{E_\lambda;\ \lambda\in\Lambda\}$ の部分系は独立であり，一部又は全部の E_λ を E_λ^c で置き換えて得られる事象系も独立である．有限又は可附番無限ヶの独立事象系 $\{E_n;\ n=1,2,\cdots\cdots\}$ の場合には

3・2）　$P(\bigcap_n E_n)=\prod_n P(E_n)$

がなりたつ．特に $P(E_1)=P(E_2)=\cdots\cdots=P(E_n)=\cdots\cdots=p$（$0<p<1$）のとき独立事象系 $\{E_n;\ n=1,2,\cdots\cdots\}$ を Bernoulli 列とよぶ。

可附番ヶの事象系 $\{E_n;\ n=1,2,\cdots\cdots\}$ に対し $\overline{\lim_n}\ E_n=\bigcap_k\bigcup_{n\geqq k}E_n$ を"<u>上極限事象</u>"，$\underline{\lim_n}\ E_n=\bigcap_k\bigcup_{n\geqq k}E_n$ を"<u>下極限事象</u>"という．"上極限事象が起る"ということは無限に多くの事象 E_n が起ることを意味し"下極限事象が起る"というのはある番号以後の事象 E_n がすべて起ることを意味する．

Borel — Cantelli の定理

i) 事象系 $\{E_n ; n=1, 2, \cdots\cdots\}$ が独立であってもなくても

3・3) $\sum_n P(E_n) < +\infty \Longrightarrow P(\overline{\lim_n} E_n = 0$

ii) 事象系 $\{E_n ; n=1, 2, \cdots\cdots\}$ が独立のとき

3・4) $\sum_n P(E_n) = +\infty \Longrightarrow P(\overline{\lim_n} E_n) = 1$

がなりたつ.

3・3), 3・4) の結論の部分は $P(\overline{\lim_n} = 1 - P(\underline{\lim_n} E_n^c)$ を使えば $\underline{\lim_n} E_n^c$ の確率で述べることもできる。この定理から次のことはただちにわかる. サイコロを投げ続けて, "n回目まで連続して1の目が出る" という事象をE_nで表わすと $P(E_n) = \frac{1}{6^n}$ である。従って i) を適用すれば "無限に多くのE_nが起る" 確率は0, 即ち "無限に1の目が出続ける" 確率は0である. また単に "n回目に1の目が出る" という事象をF_nで表わすと $\{F_n ; n=1, 2, \cdots\cdots\}$ は Bernoulli 列となり $P(F_n) = \frac{1}{6}$ であるから. ii) を適用すれば, サイコロを無限に投げ続けたとき "1の目が有限回しか出ない" 確率は0であることがわかる. ii) では事象系の独立性を仮定しているので、Borel — Cantelli の定理 i) ii) では上極限事象, 下極限事象に関するすべての場合が述べられているわけではない. そこで ii) の独立性の仮定をゆるめることが考えられ, 事実それはある程度可能なことであるが (§4 参照) そうだからといって条件を全く取除くわけにはいかない。例えばE_nとして同じ事象E ($P(E) < 1$ とする) をとって考えれば明らかである.

X_n $(n=1, 2, \cdots\cdots, N)$ をそれぞれk_n次元の確率ベクトルとする。R^{k_n} の任意のボレル集合 A_n $(n=1, 2, \cdots\cdots, N)$ に対し

3・5) $$P\left(\bigcap_{n=1}^{N} \{\omega ; X_n(\omega) \in A_n\}\right) = \prod_{n=1}^{N} P(\{\omega ; X_n(\omega) \in A_n\})$$

がなりたつとき, 確率ベクトルX_nは "互いに独立" であるといい, また $\{X_n ; n=1, 2, \cdots\cdots, N\}$ を "独立な有限確率なベクトル系" という. 無限確率ベクトル系 $\{X_\lambda ; \lambda \in \Lambda\}$ の任意の有限部分系が独立のとき $\{X_\lambda ; \lambda \in \Lambda\}$ を "独立な無限確率ベクトル系" という。X_n $(n=1, 2, \cdots\cdots, N)$ が互いに独立なk_n次元の確率ベクトルで $f_n(x)$ $(n=1, 2, \cdots\cdots, N)$ がR^{k_n}上のボレル函数ならば, $f_n(X_n)$ は確率変数となり. 平均値が存在するときには

$$3\cdot6)\quad E\left(\prod_{n=1}^{N} f_n(x_n)\right)=\prod_{n=1}^{N} E\left(f_n(x_n)\right)$$

がなりたつ。従ってx，yが互いに独立で分散の存在する確率変数ならば

$$3\cdot7)\quad E(xy)=E(x)E(y)$$

$$3\cdot8)\quad V(x+y)=V(x)+V(y)$$

がなりたつ。また3・6）を使うと確率ベクトル系$\{x_n;n=1,2,\cdots\cdots,N\}$が独立のとき，$x=(x_1,x_2,\cdots\cdots,x_n)$を$(k_1+k_2+\cdots\cdots+k_N)$次元の確率ベクトルと考え，$x$および$x_n$の特性函数を$\varphi((\xi_1,\xi_2,\cdots\cdots,\xi_{k_1+k_2+\cdots\cdots+k_n});x)$，$\varphi((\xi_1,\xi_2,\cdots\cdots,k_n);x_n)$とすると$R^{(k_1+k_2+\cdots\cdots+k_N)}$の任意の点$\xi=(\xi_1,\xi_2,\cdots\cdots,\xi_{k_1+k_2+\cdots+k_N})$で

$$3\cdot9)\quad \varphi\left((\xi_1,\xi_2,\cdots\cdots,\xi_{k_1+k_2+\cdots\cdots+k_N});x\right)$$
$$=\prod_{n=1}^{N}\varphi\left((\xi_{k_1+k_2+\cdots+k_{n-1}+1},\cdots,\xi_{k_1+k_2+\cdots+k_n});x_n\right)$$

がなりたつ．逆に任意のξに対し3・9）がなりたてば$\{x_n;n=1,2,\cdots\cdots,N\}$は独立な確率ベクトル系であることが云える．これをKacの定理という．

§4 条件附平均値，条件附確率

（〔4〕第1章，〔5〕第7章，〔6〕参照）

確率空間$\Omega(\mathcal{B},P)$において$\mathcal{B}_1$を$\mathcal{B}$の部分ボレル集合体とし$\mathcal{B}_1\subset\mathcal{B}$とかくこととする．$x$を平均値の存在する確率変数とし$\mathcal{B}_1(\subset\mathcal{B})$に属する任意の集合$E$に対し

$$4\cdot1)\quad Q(E)=\int_E x(\omega)\,dP(\omega)$$

を考えると$Q(E)$は$\mathcal{B}_1$の元に対し定義された加法的函数でPに関して絶対連続（*absolutely continuous*）になっている．従って測度論における

Radon－Nikodym の定理により $\mathcal{B}_1$ 一可測な函数 $f(\omega)$ が存在して

$$4\cdot2)\quad Q(E)=\int_E f(\omega)\,dP(\omega)$$

と表わせる．このとき $f(\omega)$ は P 一測度 0 を除いて一意に定まる．即ち２つの $\mathcal{B}_1$ 一可測な函数 $f_1(\omega)$，$f_2(\omega)$ があってその何れでも $Q(E)$ が4.2) の形で表わせるならば $P(\{\omega:f_1(\omega)\neq f_2(\omega)\})=0$ である．このように殆んどすべての ω で等しい２つの函数 $f_1(\omega)$，$f_2(\omega)$ を $f_1\simeq f_2$ または $f_1=f_2$ $(a.e.)$ で表わす．上のようにして P 一測度 0 を除いて一意に定まる函数 $f(\omega)$ を"$\mathcal{B}_1$ に関する x の条件付平均値"（conditional expectation）といい $E(x/\mathcal{B}_1)(\omega)$ とかく．従って $E(x/\mathcal{B}_1)(\omega)$ は $\mathcal{B}_1$ 一可測な函数であるが特に ω における値を問題としたり、またそのことを強調するとき以外は ω を省略して $E(x/\mathcal{B}_1)$ ともかく。$\mathcal{B}_2(\subset\mathcal{B})$ が $\mathcal{B}_1$ とは異なるボレル集合体ならば同じように $E(x/\mathcal{B}_2)$ が定義できるが、一般には $E(x/\mathcal{B}_1)(\omega)$ と $E(x/\mathcal{B}_2)(\omega)$ は等しくないばかりでなく $E(x/\mathcal{B}_2)$ は $\mathcal{B}_1$ 一可測にもならない．しかし特に $\mathcal{B}_2\subset\mathcal{B}_1$ で，さらに $E(x/\mathcal{B}_1)$ が $\mathcal{B}_2$ 一可測ならば $E(x/\mathcal{B}_1)\simeq E(x/\mathcal{B}_2)$ がなりたつ．一般に条件付平均値が定義できるときには次のことがなりたつ．P 一測度 0 を除いて

4・3)　$E(1/\mathcal{B}_1)=1$

4・4)　$x\geqq 0\implies E(x/\mathcal{B}_1)\geqq 0$

4・5)　$E(E(x/\mathcal{B}_1))=E(x)$

4・6)　$|E(x/\mathcal{B}_1)|\leqq E(|x|/\mathcal{B}_1)$

4・7)　任意の定数 $c_1,c_2,\cdots\cdots,c_n$ に対し $E\left(\sum_{j=1}^{n}c_jx_j/\mathcal{B}_1\right)$

$$=\sum_{j=1}^{n}c_jE(x_j/\mathcal{B}_1)$$

4・8)　$\lim_{n\to\infty}x_n=x,\quad |x_n|\leq y\ (a.e.)$ かつ $E(|y|)<+\infty$

$$\implies\lim_{n\to\infty}E(x_n/\mathcal{B}_1)=E(x/\mathcal{B}_1)$$

4・9)　$\mathcal{B}_2\subset\mathcal{B}_1\implies E(E(x/\mathcal{B}_1)/\mathcal{B}_2)=E(x/\mathcal{B}_2)$

4・10)　x が $\mathcal{B}$ 一可測 $\implies E(xy/\mathcal{B}_1)=xE(y/\mathcal{B}_1)$

4・11)　y が $\mathcal{B}_1$ 一可測 $\implies E((x-E(x/\mathcal{B}_1))^2)\leqq E((x-y)^2)$

4・11)は x を $\mathcal{B}_1$ 一可測な函数で近似するとき２乗平均の意味で最良なも

のは$E(x/\mathcal{B}_1)$であることを示している。

xが可測集合Eの定義函数のとき$E(x/\mathcal{B}_1)$を"$\mathcal{B}_1$に関するEの条件付確率"(*conditional probability*)といい，$P(E/\mathcal{B}_1)$とかく．4.3) 4.4)によれば$1 \geqq P(E/\mathcal{B}_1) \geqq 0$ (a.e) であって

$$4.12)\quad F \in \mathcal{B}_1 \Longrightarrow P(E \cap F) = \int_F P(E/\mathcal{B}_1)(\omega)\, dP(\omega)$$

がなりたつ．特に$\mathcal{B}_1 = \{F, F^c, \phi, \Omega\}$のときには，$P(E/\mathcal{B}_1)$の$\mathcal{B}_1$-可測性から二つの定数$a$, bがあって

$$\omega \in F \Longrightarrow P(E/\mathcal{B}_1)(\omega) = a,$$

となっている。特に$a > 0$ならばFに属するωに対しては

$$P(E/\mathcal{B}_1)(\omega) = \frac{P(E \cap F)}{P(F)}$$

がなりたつ。この左辺をFに関するEの条件付確率といい$P(E/F)$とかく。即ち$P(F) > 0$のときFに関するEの条件付確率$P(E/F)$は

$$4.13)\quad P(E/F) \equiv \frac{P(E \cap F)}{P(F)}$$

で定義される．この条件付確率を使うと§3の*Borel—Cantelli*の定理 ii)の部分は次のように改良される。

<u>*Chung—Erdös*の定理</u>　事象系$\{E_n ; n = 1, 2, \cdots\cdots\}$が次の2条件をみたすとする。

i)　任意の自然数h及びk $(h \leqq k)$に対し正数$c(h)$及び$n(h, k)$が定まり

$$4.14)\quad n \geqq n(h, k) \Longrightarrow P(E_n / E_h^c \cap E_{h+1}^c \cap \cdots\cdots \cap E_k^c) > c(h) P(E_k)$$

ii)　定数C_1, C_2を適当にとれば，任意のnに対し$\{n_j ; j = 1, 2, \cdots\cdots, s(n)\}$が選べて

$$4.15)\quad \sum_{j=1}^{s(n)} P(E_n \cap E_{n_j}) < C_1 P(E_n)$$

$$4.16)\quad m > n \text{ で } m \neq n_j \ (j = 1, 2, \cdots\cdots, s(n)) \Longrightarrow P(E_n \cap E_m) < C_2 P(E_n) P(E_m)$$

がなりたつ。このとき

4・17)　$\sum_n P(E_n)=+\infty \Longrightarrow P(\overline{\lim_n}\ E_n)=1$

がなりたつ．独立事象列のときには i)，ii) はみたされるからこれは Borel — Cantelli の定理 ii) の部分の改良である．

確率変数系 $\{x_\lambda;\lambda\in\Lambda\}$ の各変数 x_λ を可測にするような最小のボレル集合体を $\mathcal{B}_\Lambda$ とするとき，$E(x/\mathcal{B}_\Lambda)$ を $\{x_\lambda;\lambda\in\Lambda\}$ に関する x の条件付平均値といい $E(x/x_\lambda,\lambda\in\Lambda)$ ともかく．また可測集合 E の $\mathcal{B}_\Lambda$ に関する条件付確率を $P(E/x_\lambda,\lambda\in\Lambda)$ ともかく．Λ がどんな無限集合であっても x に応じてたかだか可附番無限の部分集合 $\{\lambda_j;\ j=1,2,\text{-----}\}$ を適当にとれば

4・18)　$E(x/x_\lambda,\lambda\in\Lambda)\simeq E(x/x_{\lambda_j},j=1,2,\text{----})$

がなりたつ．x を可測集合 E の定義函数とすれば同じように $\{\lambda_j;\ j=1,2,$ -----$\}$ が存在して

4・19)　$P(E/x_\lambda,\lambda\in\Lambda)\simeq P(E/x_{\lambda_j},j=1,2,\text{-----})$

がなりたつ．また x と $\{x_\lambda;\lambda\in\Lambda\}$ が独立ならば

4・20)　$E(x/x_\lambda;\lambda\in\Lambda)\simeq E(x)$

がなりたつ。

条件付確率 $P(E/\mathcal{B}_1)$ は E を固定すれば P ー測度 0 を除いて一意的に定まるが E が変れば除外集合も変る。しかしたかだか可附番無限ケの事象だけを考えるときには P ー測度 0 の集合を除いて

4・21)　$0\leqq P(E/\mathcal{B}_1)\leqq 1$

4・22)　$P(\Omega/\mathcal{B}_1)=1$

4・23)　$E=\sum_n E_n \Longrightarrow P(E/\mathcal{B})=\sum_n P(E_n/\mathcal{B}_1)$

がなりたつ．しかし可附番無限ケ以上の事象を同時に考えると除外集合の和集合が可測になっているか或いは可測であっても P ー測度 0 になっているかどうかがわからない．この問題が肯定的に解決されるとき，即ち E と ω の函数 $P_{\mathcal{B}_1}(E,\omega)$ があって

i)　$E\in\mathcal{B}\Longrightarrow P_{\mathcal{B}_1}(E,\omega)$ は ω の函数として $\mathcal{B}$ー可測

ii)　ω を固定すれば $P_{\mathcal{B}_1}(E,\omega)$ は $\Omega(\mathcal{B})$ 上の確率分布

iii)　$E\in\mathcal{B}\Longrightarrow P_{\mathcal{B}_1}(E,\omega)\simeq P(E/\mathcal{B}_1)(\omega)$

がなりたつとき $P_{\mathcal{B}_1}(E,\omega)$ を“$\mathcal{B}_1$ に関する条件付確率法則”（conditional probability law）という．条件付確率法則が存在するときには

$$4.24)\quad E(x/\mathcal{B}_1)(\omega) \simeq \int_{\Omega} x(\omega')\, dP_{\mathcal{B}_1}(\omega', \omega)$$

がなりたつ．k次元確率ベクトル x を可測にするような最小ボレル集合体を $\mathcal{B}_x$（$\subset \mathcal{B}$）とするとき，$\mathcal{B}_x$ に属する集合 E に対し定義された E と ω の函数 $P'_{\mathcal{B}_1}(E, \omega)$ があって $\mathcal{B}_x$ 上で上の i)，ii)，iii) をみたしているとき，$P_{\mathcal{B}_1}(E, \omega)$ を "$\mathcal{B}_1$ に関する x の条件付確率法則" という。x の条件付確率法則がある場合には 4.24）と同様に

$$4.25)\quad E(x/\mathcal{B}_1)(\omega) \simeq \int_{\Omega} x(\omega')\, dP'_{\mathcal{B}_1}(\omega', \omega)$$

がなりたつ．確率ベクトル x の場合には x の条件付確率法則は存在する。（それは x の値域空間を R^k としているからである．一般に x の値域空間が局所コンパクトな位相空間で第二附可番性公理をみたすときは x の条件付確率法則が存在する。）特に x が $\{x_\lambda ; \lambda \in \Lambda\}$ と独立ならば，各 x_λ を可測にする最小のボレル集合体 $\mathcal{B}_\Lambda$ に関する x の条件付確立法則は ω に関係しない．即ち

$$4.26)\quad x \text{ と } \{x_\lambda ; \lambda \in \Lambda\} \text{ が独立} \Longrightarrow$$

$$\mathcal{B}_x \text{ の任意の元 } E \text{ に対し } P'_{\mathcal{B}_\Lambda}(E, \omega) \simeq P(E)$$

§5 確率ベクトルの収束

（〔9〕第2章参照）

x，x_n（$n = 1, 2, \cdots\cdots$）を k次元確率ベクトルとする．

1．概収束

$E = \{\omega ; x_n(\omega) \xrightarrow[n\to\infty]{} x(\omega)\}$ は可測事象であるがこの確率が1のとき即ち

$$5.1)\quad P(x_n \longrightarrow x) = 1$$

のとき "x_n は x に概収束（convergence almost everywhere, convergence almost sure）する" といい，$x_n \longrightarrow x$ (a. e.) とかく．R^k の点 ξ と原点との距離を $|\xi|$ で表わし ε_n（$n = 1, 2, \cdots\cdots$）を0に収

束する正数列とする．Borel − Cantelli の定理 i) を使えば

5・2)　$\sum_n P(|X_n - X| > \varepsilon_n) < +\infty \Longrightarrow X_n \longrightarrow X\ (a.e.)$

が証明される．また任意の $\varepsilon\ (>0)$ に対し，$m > n \to \infty$ のとき

5・3)　$P(\max_{n+1 \leq l \leq m} |X_n - X_l| > \varepsilon) \longrightarrow 0$

ならば，ある X があって $X_n \longrightarrow X\ (a.e.)$ がなりたつ。

2. 確 率 収 束

任意の $\varepsilon\ (>0)$ に対し

5・4)　$\lim_{n\to\infty} P(|X_n - X| > \varepsilon) = 0$

がなりたつとき"X_n は X に確率収束 (convergence in probability, stochastic convergence) する"といい $X_n \longrightarrow X$ (in prob) とかく。$X_n \longrightarrow X\ (a.e.)$ ならば $X_n \longrightarrow X$ (in prob) であるが逆は一般にはなりたたない．しかし $X_n \longrightarrow X$ (in prob) のときには適当な部分列 X_{n_j} をとれば $X_{n_j} \longrightarrow X\ (a.e.)$ となる．また任意の $\varepsilon\ (>0)$ に対し $m > n \to \infty$ のとき

5・5)　$P(|X_m - X_n| > \varepsilon) \longrightarrow 0$

ならばある X が存在して，$X_n \longrightarrow X$ (in prob) となっている．

3. 平 均 収 束

$p > 0$ に対し

5・6)　$\lim_{n\to\infty} E(|X_n - X|^p) = 0$

のとき"X_n は X に p 次平均収束 (convergence in p-th mean) する"といい，$X_n \longrightarrow X$ (in p-th mean) とかく．2・8) から $q > 1$ のとき $X_n \to X$ (in q-th mean) ならば，$1 \leq p < q$ なる任意の p で $X_n \longrightarrow X$ (in p-th mean) であり，また 2・9) から，$X_n \longrightarrow X$ (in p-th mean) ならば $X_n \longrightarrow X$ (in prob) はでるが，逆は一般になりたたない。

4. 法 則 収 束

X_n の分布を Φ_n，X の分布を Φ とするとき，台 (support, carrier) がコンパクトな R^k 上の任意の連続函数 $f(X)$ に対し

5・7) $\displaystyle\lim_{n\to\infty}\int_{R^k} f(x)\,d\Phi_n(x)=\int_{R^k} f(x)\,d\Phi(x)$

となるとき"x_nはxに法則収束(convergence in law)する"といい，$x_n \longrightarrow x$ (in law) とかく。これはまた次のように云ってもよい．R^kのボレル集合の全体を$\mathcal{B}^k$とするとき

5・8) $E\in\mathcal{B}^k,\ \Phi(E^i)=\Phi(E^a)\Longrightarrow \displaystyle\lim_{n\to\infty}\Phi_n(E)=\Phi(E)$,

ただしE^i，E^{a}はそれぞれEの内点の全体，Eの閉包を表わす。(5.8)の条件$\Phi(E^i)=\Phi(E^a)$をみたすEをΦの"連続集合(continuity set)という.)法則収束は分布函数で定義されているので，他の収束とは異なった意味を持っているが，$x_n \longrightarrow x$ (in prob) ならば $x_n \longrightarrow x$ (in law) となっている．従って$x_n \longrightarrow x$ (a, e) または$x_n \longrightarrow x$ (in p-th mean) ならば $x_n \longrightarrow x$ (in law) である。$x_n \longrightarrow x$ (in law) から$x_n \rightarrow x$ (in prob) は一般にはでないが，特にR^kの一点aに対し $x(\omega)\simeq a$のときには$x_n \longrightarrow a$ (in law) から $x_n \longrightarrow a$ (in prob) がでる． $x_n \longrightarrow x$ (in law) なるための必要かつ十分な条件は次の特性函数の収束で与えられる。R^kの任意の点ξで

5・9) $\displaystyle\lim_{n\to\infty}\varphi(\xi;x_n)=\varphi(\xi;x)$

これを V. Glivenko の定理という．また

5・10) $\varphi(\xi;x_n)$がR^kで広義一様収束

ならばあるxが存在して $x_n \longrightarrow x$ (in law) がなりたつ。これを P. Lévy の定理という.

上に述べた4種の収束の間には次の関係がなりたち、概収束と平均収束とはどちらが強いとも云えない。

5・11) $\left.\begin{array}{l}\text{概収束}\\ \text{平均収束}\end{array}\right\rangle\Longrightarrow$ 確率収束 $\Longrightarrow$ 法則収束

§6　独立確率変数列とその和

（〔1〕第6章，〔2〕第4章，〔7〕第6章，〔8〕第3．4章，〔10〕第2章参照）

$\{x_n ; n=1,2,\cdots\}$ を独立確率変数列とする．§2で Tchebychev の不等式を述べたが独立性の決定のもとではこれは次のように拡張できる．

6・1)　$E(x_j)=0,\ V(x_j)<\infty\ (j=1,2,\cdots,n) \Longrightarrow$

任意の $\varepsilon(>0)$ に対し，$P\left(\max_{1\leq j\leq n}|x_1+x_2+\cdots+x_j|\geqq\varepsilon\right)\leqq\frac{1}{\varepsilon^2}\sum_{j=1}^{n}V(x_j)$

これを Kolmogorov の不等式という．6・1）で特に $n=1$ の場合が Tchebychev の不等式である．また $\mathcal{B}_n$ を $x_{n+1},\ x_{n+2},\ \cdots$ をすべて可測にする最小ボレル集合体とする．

Kolmogorov の 0－1 定理　$E\in\mathcal{B}_n\ (n=1,2,\cdots)\Longrightarrow P(E)=0$ 又は 1 がなりたつ．（この定理は x_n を k_n 次元確率ベクトルとしても正しい。）従って独立確率変数列の場合には

6・2)　$E=\{\omega ; \sum_n x_n(\omega)$ が収束する$\}$

とおくと E は定理の仮定をみたし（収束するか否かは最初の有限ヶの x_n の値には関係ない）$P(E)$ は 0 または 1 である．これが 1 になる場合は $S_m=\sum_{n=1}^{m}x_n$ は $m\longrightarrow\infty$ のとき概収束しているわけである．そのための条件としては 6・1）を使うと

6・3)　$\sum_{n=1}^{\infty}E(x_n)$，$\sum_{n=1}^{\infty}V(x_n)$ が共に収束

$$S_m=\sum_{n=1}^{m}x_n \text{ は } m\longrightarrow\infty \text{ のとき 概収束}$$

ができる．また次の定理もある．

Kolmogorov — Khintchine の 3 級数定理．$|x_n(\omega)|\leqq 1$ のとき $x'_n(\omega)=x_n(\omega)$，$|x_n(\omega)|>1$ ならば $x'_n(\omega)=0$ とするとき $S_m=\sum_{n=1}^{m}x_n$ が概収束するための必要かつ十分な条件は次の 3 級数がすべて収束することである。

6・4)　$\sum_{n=1}^{\infty}E(x'_n)$，$\sum_{n=1}^{\infty}V(x'_n)$，$\sum_{n=1}^{\infty}P(x_n\neq x'_n)$

各種の収束の間には 5・11）の関係にあるが，独立確率変数列の部分和 $S_m=\sum_{n=1}^{m}x_n$ の収束に関しては次のことが云える。

(A1〜18)

6・5)　法則収束 $\Longrightarrow$ 確率収束 $\Longrightarrow$ 概収束

従ってこの場合 5.11) と合せると上の3種の収束は同等なことがわかるから $S_m=\sum_{n=1}^{m}\chi_n$ が概収束するためには法則収束または確率収束すること水必要かつ十分である。6.3)，6.4)は何れも確率変数の分布に関する量で述べられているが，上に述べた理由からこれは次のように云いかえてもよい。

χ_n の分布を Φ とすると，χ_n の独立性から $S_m=\sum_{n=1}^{m}\chi_n$ の分布 Ψ_m は Φ_1, Φ_2, ----- Φ_m の重畳（convolution）即ち $\Psi_m=\Phi_1*\Phi_2*----*\Phi_m$ となるが，S_m が法則収束従って概収束するための必要かつ十分な条件は Ψ_m の特性函数

$$6\cdot 6)\quad \varphi(\xi;\Psi_m)=\prod_{n=1}^{m}\int_{-\infty}^{\infty}e^{i\xi x}\,d\Phi_n(x)$$

が $m\longrightarrow\infty$ のとき $-\infty<\xi<\infty$ で広義一様収束することである。

また次のような概念を導入して S_m の収束を少し拡張した形で述べることが出来る。1次元確率分布 Φ と $\ell>0$ に対し

$$6\cdot 7)\quad Q_\Phi(\ell)=\sup_{-\infty<a+\infty}\Phi([a,a+\ell])$$

を"Φの最大濃度函数"（maximal concentrations function）という。これは P. Lévy により導入された量であるが国沢清典は次のように"平均値濃度函数"を定義した。$\mu(\xi)$ を非負の偶函数で

$$\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mu(\xi)\,d\xi=1$$

かつそのフーリエ変換

$$6\cdot 8)\quad m(x)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}e^{i\xi x}\mu(\xi)\,d\xi$$

は $(0,\infty)$ で単調減少（非増加）であるとする，このとき Φの特性函数 $\varphi(\xi;\Phi)$ を使って作った

$$6\cdot 9)\quad g_\Phi(\ell)=\frac{1}{\sqrt{2\pi}}\int_{-\infty}^{\infty}\mu(\ell\xi)\,|\varphi(\xi;\Phi)|^2\,d\xi$$

を一般な場合の Φ の"平均濃度函数"（mean concentration function）という。$Q_\Phi(\ell)$ と $g_\Phi(\ell)$ の間には $a>0$ に対して

$$6\cdot 10)\quad 2\left(1+\int_{-\infty}^{\infty}m(x)\,dx\right)Q_\Phi(\ell)\geq g_\Phi(\ell)\geq m(a)\,Q_\Phi^2(a\ell)$$

なる関係がなりたつ。6・9）の式で特に$\mu(\xi)=\sqrt{\frac{2}{\pi}}\left(\frac{\sin\xi}{\xi}\right)^2$とおくと河田竜夫による平均濃度函数

6・11) $C_{\Phi}(l)=\frac{1}{2l}\int_{-\infty}^{\infty}\{\Phi((x-l,x+l])\}^2dx$

が導かれる．また$\mu(\xi)=\sqrt{\frac{\pi}{2}}e^{-|\xi|}$とおき，$\Phi(dx)$と$\{1-\Phi(-dx)\}$の重畳を$\tilde{\Phi}(dx)$で表わすと国沢清典の平均濃度函数

6・12) $\varphi_{\Phi}(l)=\int_{-\infty}^{\infty}\frac{l^2}{l^2+x^2}\,d\tilde{\Phi}(x)$

が導かれる。2・10）によれば分布のちらばり状態はその分散によって評価出来るが，分散は常に存在するとは限らない．これに反し上記諸量はどんなΦに対しても定義出来て，ある意味で分布の散らばり或いは滑らかさを示している．さらにこれらの量は何れも重畳により減少する．即ち

6・13) $Q_{\Phi_1*\Phi_2}(l)\leqq Q_{\Phi_j}(l)$, $g_{\Phi_1*\Phi_2}(l)\leqq g_{\Phi_j}(l)$ $(j=1,2)$

そこで独立確率変数列$\{x_n;n=1,2,\cdots\}$に対し$S_{n,m}=\sum_{j=n+1}^{m}x_j$の分布の濃度函数をそれぞれ$Q_{n,m}(l)$, $C_{n,m}(l)$, $\varphi_{n,m}(l)$とすると6・13）から$Q(l)=\lim_{n\to\infty}\lim_{m\to\infty}Q_{n,m}(l)$, $C(l)=\lim_{n\to\infty}\lim_{m\to\infty}C_{n,m}(l)$, $\varphi(l)=\lim_{n\to\infty}\lim_{m\to\infty}\varphi_{n,m}(l)$が存在するがこれは何れも$l\,(>0)$の値に無関係に0または1になる。そして適当な定数列$\{a_n;n=1,2,\cdots\}$をとって$S'_m=\sum_{n=1}^{m}(x_n-a_n)$が$m\to\infty$のとき概収束するようにできるための必要かつ十分な条件は

6・14) $Q(l)=1,\quad C(l)=1,\quad \varphi(l)=1$

で与えられる．$\varphi_{\Phi_n}(l)$を使って6・14）を書き直せば

6・15) $\sum_{n=1}^{\infty}\{1-\varphi_{\Phi_n}(l)\}<+\infty$

となる。また6・9）の式で特に$\mu(\xi)=\sqrt{\frac{2}{\pi}}\frac{1}{1+\xi^2}$とおいて$\frac{1}{g_{\Phi}(1)}$の対数をとると伊藤清の“$\Phi$の散布度”（*dispersion*）

6・16) $\delta_{\Phi}=-\log\left[\int_{-\infty}^{\infty}e^{-|x|}\,d\tilde{\Phi}(x)\right]$

が導かれる。これについては

6・17) $\delta_{\Phi_1*\Phi_2}\geq\delta_{\Phi_j}\quad(j=1,2)$

(A1~20)

がなりたち，従って $S_m=\sum_{n=1}^{m} x_n$ の分布の散布度を δ_m とすると $\delta=\lim_{m\to\infty}\delta_m$ が存在する，これを使うと 6・14)の必要十分条件は

6・18) $\delta<+\infty$

で表わせる．

なお本項目の責任者・原稿執筆者は白尾，討論の段階で近藤が協力した。

文　　献

〔1〕 A. Kolmogorov : Grundbegriffe der Wahrscheinlichkeitsrechnung, Erg. der Math (Berlin), 1933

〔2〕 P. Lévy : Theorie de l'addition des variables aleatoires (Paris), 1937

〔3〕 W. Feller : An Introduction to probability theory and its application (New york), 1957

〔4〕 J.L. Doob : Stochastic process (New York), 1952

〔5〕 M. Loéve : Probability theory (New York), 1955

〔6〕 K.L. Chung & P. Erdös : On the application of the Borel-Cantelli's lemma, Trans. Amer. Math. Soc, vol 72, 1952

〔7〕 河田竜夫 : フーリエ解析と確率論（中文館），1947

〔8〕 国沢清典 : 確率論に於ける極限定理（中文館），1949

〔9〕 伊藤　清 : 確率論（岩波），1952

〔10〕 〃 : 確率過程（岩波，現代応用数学講座），1957

〔11〕 丸山儀四郎 : 確率論（共立，現代数学講座），1957

索　　引

M

N

R

S

T

Y

Z