確率論基礎

渕野 昌 (Sakaé Fuchino)

2019 年 12 月 15 日

以下のテキストは，2013 年度前期の神戸大学情報知能工学科 2 年生のための「確率論基礎」の講義の講義録（講義ノート）です．このノートの最新版は

https://fuchino.ddo.jp/kobe/probabilityLN2013.pdf

としてダウンロードできます．この講義ノートは，垣内逸郎先生の 2012 年の講義の講義ノートをベースにして作成しています．垣内先生のノートは

https://fuchino.ddo.jp/kobe/probability2012.pdf

として download できます．

あなたがダウンロードしたのは，

2019 年 12 月 15 日 (00:22JST) 版

です．このテキストはまだ暫定版です．学期中に講義の進展に前後して拡張／改良してゆく予定です．すでに書いてある部分についても，改良のための変更をする可能性もあるので注意してください．内容に関するコメントや質問，タイプミスの指摘など何でもあれば遠慮なくしてください．

なお，このテキストを含め，私が神戸大学で行なっている講義に関連する資料へのリンクは，

https://fuchino.ddo.jp/kobe/index.html

からたどって見つけることができます．

目次

1 確率

probability

ある試行 (trial: 実験，測定，観測，アンケートなど) t を行なうことを考えるとき，「 t の結果が ... となることの確率は ... %である」という予測が必要になることがある．ここでの 「結果が ... となること」での 「...」 ような試行の測定結果の属性のことを事象 (event) とよぶ[1)]．たとえば，

prob-Ex-1

例 1.1 (a) 天気予報: 「明日の神戸市の降雨確率は 25 %です.」

(b) サイコロの目: 「サイコロを投げたとき 偶数の目が出る確率は $\frac{1}{2}$ です.」

(c) 街で出逢った日本人の身長:「街中で無作為[2)] に選んだ日本人の身長が 150cm から 160cm の間にある確率は 32.5 %です.」

ある事象 e に関して，「試行 t を行った結果が e となる確率は p である」 とは，試行 t を (同じ条件で) n 回繰り返すとき，e が起っている回数を $m(n)$ とすると，$\lim_{n\to\infty}\frac{m(n)}{n}=p$ となること，と定める．ただし，ここでは，$m(n)$ は試行の繰り返しごとにばらつくかもしれないが，$\lim_{n\to\infty}\frac{m(n)}{n}$ は常に一定値になる，と仮定している．パーセントでは，「t が e となる確率は c %である」と言ったときには，上の p に対して，$c=100p$ である．

例 1.2 (前の例の続き) (a) 「明日の神戸の降水確率は 25 %」については「明日が来る」という「試行」は 1 回しかできないので，上の定義はそのままではうまく解釈できない．しかし，「今日と同じ天気の配置 (地球全体の気圧配置，偏西風の状況, 海水の温度分布 etc.) から出発して 24 時間たったときの (想定される) 状況を n 個 (仮想の parallel worlds として考え (あるいは n 回のシミュレーションの結果として見てみたとき)) n を大きくしてゆくと，これらの n 個の parallel worlds のうち，神戸に雨が降っている状況はそのうちの $0.25\times\ n$ 個程度になる」，というような解釈で理解することができる． prob-Ex-2

(b) いかさまが仕組まれていないサイコロでは，どの目も出る頻度が同じだろう．この想定のもとで，サイコロを n 回投げたときの，偶数の目 (つまり 2 か 4 か 6) が出る回数 $m(n)$ は

$$\lim_{n\to\infty}\frac{m(n)}{n}=\frac{1}{6}+\frac{1}{6}+\frac{1}{6}=\frac{1}{2}$$

[1)] 第 2 節 以降では，ここで事象とよんでいるものを “基本事象” とよび，基本事象の集合として定義される複合的な事象のことを “事象” とよんでいる．

[2)] 無作為 (むさくい) とは意図的な選択をしない，ということだが，意図的であるかないかに関わらず，偏りのない不規則的な選択をすることを「でたらめに選ぶ」とか「ランダムに選ぶ」(choose at random, choose randomly) というように表現することもある．ただし，「でたらめ」という言葉については，「滅茶苦茶」とか「意味をなさないこと」というような否定的なコノーテーションもあることから，たとえば高校の教科書などでは意識的に避けられる場合もあるようである．

となると仮定してよい．これは理論的な考察だが，実際のサイコロを 1 つとったときに，そのサイコロで，本当にどの目が出る確率も同じになっているかどうかは，調べてみなければ分らないだろう．しかし，当然のことながら無限回サイコロを投げて確かめることはできないので，厳密な意味で、このことを確定的に検証することはできない． そこで，サイコロを投げる実験を何回繰り返して，一つ一つの目の出る割合が $\frac{1}{6}$ にどれだけ近い値となったときに，どれくらいの確かさで上の「仮定」が正しいと言えるだろうか？ ということを考察する必要が出てくる[3)]．

(c) 街で出逢った日本人の身長については，日本人の人口は有限の数なので，その数だけ重複のないように試行を繰り返せば，確率について確定的なことが言えるはずである．このような試行のことを全数調査というが，それが実行できたとしても，この例の場合には，(試行を繰り返して調査をしている間に) ひとりひとりの身長は時間とともに変化するはずで，しかもこれらの個体は時間とともに生れたり死んだりもするので，そのことをどう扱うかという問題は残るだろう[4)]．

現実的には，ここでも，試行を実行可能な回数の範囲のうちで何回繰り返してデータをとったときに，そのデータがこの試行の結果の確率的な挙動を十分によく反映するものになっているか，というような指標を与えてくれる理論が必要である．このような理論があれば，それを，(c) のタイプの問題では，(比較的) 少数のデータを用いて全体の状況を推測するための方法論ととらえることも可能だろう．

2　標本空間と事象

「試行 t を行ったときの結果でありえるもの (数値，真偽値など)」 を試行 t に関する基本事象 (elementary events, atomic events) とよぶ．ある試行に関する基本事象を全部集めてできる集合を Ω (ギリシャ文字の大文字のオメガ) で表す[5)]． sample-space

例 2.1 (前の例の続き) (a) 明日の神戸での日降水量を mm 単位で表現することにすると，ここでの基本事象の全体 Ω は半開区間 $[0, 1000)$ に含まれる数値で表すことができるだろう[6)]．このとき，「神戸で明日雨が降る」は明日の神戸での雨量を prob-Ex-3

3) 後でこの問に答えることのできる理論について述べることになる．

4) ここの例では，そこまで考えるのはナンセンスかもしれないが，例えば，確率や統計学を生物学に適用するときには，対象の集まりのそのようなダイナミックな特性まで含めて考察をすることが必要になることもあるだろう．

5) 新約聖書にある 「我はアルファにしてオメガなり」 という言葉 (「天上天下唯我独尊」のキリスト教版?) から，記号 Ω には森羅万象というような意味が付与されることが多い．ここでの記号の選択もそのような連想からと思われる．

6) 日本での日降水量の最大記録は，2011 年 7 月 19 日，高知県魚梁瀬での 851.5mm である．[19.11.05(火 10:55(JST)) の付記]: この脚注は，2013 年に書いたものだが，魚梁瀬の記録は，2019

計測するという試行 t の結果が，$(0, 1000)$ に入っていることである．

(b) サイコロを投げる，という試行 t の結果は，出た目の数だから，この数を基本事象と考えるとそれらの全体は，$\Omega = \{1, 2, 3, 4, 5, 6\}$ である．「偶数の目が出る」という事象は，これの部分集合 $\{2, 4, 6\}$ に対応する．

(c) 街で出逢った日本人の身長の例の続き．ここでの基本事象の全体 Ω は, たとえば cm 単位で考えることにして, 数直線上の区間 $(0, 400)$ に含まれる数と同一視してよいだろう．たとえば,「街で出逢った人の身長を測定した結果が 182.5cm である」というのは, この測定値 (基本事象) が (四捨五入により) 半開区間 $[182.45, 182.55) \subseteq \Omega$ に入っているということと考えられる．

上での例でも分るように，一般的な事象は，Ω の部分集合[7] で表現できる．そこで，Ω の部分集合 (のうちの選ばれた一部のもの) を事象 (event) とよぶことにして，事象の全体を $\mathcal{A}$ であらわすことにする[8]．

特に，

(2.1) Ω 自身: 事象としての解釈は「何か起こる[9]」ということ

(2.2) 空集合[10] $\emptyset$: 事象としての解釈は「何も起こらない」，ということ，

は基本的な 2 つの事象として考察の対象になっている必要があるだろう．

1 つ 1 つの基本事象 e は Ω の要素であって，一般には Ω の部分ではないので，厳密には上の意味での事象にはならないのだが，一つ一つの基本事象 e を $\{\mathsf{e}\} \subseteq \Omega$ と同一視することで，上のような意味での事象の特別な場合と看做すことにする．

事象 A, B,... に対して，次のような事象も考察することが必要となる可能性がある:

事象の和: $A, B \in \mathcal{A}$ のとき[11]，集合

(2.3) $A \cup B = \{x \in \Omega : x \in A$ または $x \in B\}$

年 10 月 12 日に神奈川県箱根町での 922.5mm によって更新されている．

7) Y が集合 X の部分集合であるとは，Y の要素はすべて X の要素になっていることである．「Y が X の部分集合である」を，$Y \subseteq X$ あるいは，$Y \subseteqq X$ で表す．

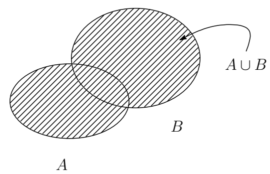
8) 文字 $\mathcal{A}$ を使っているのは，後述の “σ-algebra” という用語に対応させているからである．

9) つまり，試行を行ったときに起きる可能性のある基本事象のうちのどれかが起る．

10) 要素を 1 つも持たない集合を空集合という．空集合はすべての集合の部分集合になる．

11) $x \in X$ で “x は集合 X に属す” を表す．集合の属性は，この要素関係で記述できると考える．たとえば，$X \subseteq Y$ は，「すべての対象 x に対し，$x \in X$ なら $x \in Y$ である」ことであるし，集合 X と Y が等しいというのは 「すべての対象 x に対し $x \in X$ であることと $x \in Y$ であることは同値」だということである．$x \in X$ でないとき， $x \notin X$ と書く．

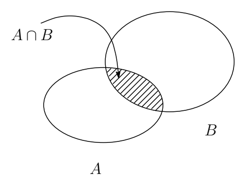
を A と B の和事象 (sum (or join) event) とよぶ．事象としての $A \cup B$ は「A または B (の少なくとも片方が) 起きる」ということである．



　事象の積: A, $B \in \mathcal{A}$ のとき，集合

$$(2.4)\quad A \cap B = \{x \in \Omega : x \in A \text{ かつ } x \in B\}$$

を A と B の積事象 (product (intersection, meet) event) とよぶ．事象としての $A \cap B$ は，「A かつ B (の両方が) 起きる」ということである．



　同様に，m 個の事象 $A_1, A_2, \ldots, A_m$ の和[12]；

$$(2.5)\quad A_1 \cup A_2 \cup \cdots \cup A_m = \{x \in \Omega : x \in A_1 \text{ または}, \ldots, \text{または } x \in A_m\}$$

あるいは可算無限個の[13] 事象 A_n, $n \in \mathbb{N}$ の和[14]

$$(2.6)\quad \bigcup_{n \in \mathbb{N}} A_n = \{x \in \Omega : \text{ある } n \in \mathbb{N} \text{ に対し } x \in A_i\}$$

や積

$$(2.7)\quad A_1 \cap A_2 \cap \cdots \cap A_m = \{x \in \Omega : x \in A_1 \text{ かつ}, \ldots, \text{かつ } x \in A_m\},$$

$$(2.8)\quad \bigcap_{n \in \mathbb{N}} A_n = \{x \in \Omega : \text{すての } n \in \mathbb{N} \text{ に対し } x \in A_i\}$$

も同様に事象として考えなければならなくなる可能性がある．

　事象 $A \in \mathcal{A}$ に対し，集合

$$(2.9)\quad A^c = \{x \in \Omega : x \in A \text{ でない }\}$$

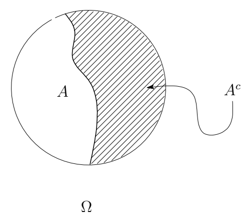
---

[12] $A_1 \cup \cdots \cup A_m$ は，$\bigcup\{A_1, \ldots, A_n\}$ と表されることもある．

[13] 自然数で要素のすべてを添字付けできるような集合は可算 (無限) 集合であるという．ここで可算無限個と言っているのは，考えている集合の族が，可算無限集合となっているということである．

[14] $\bigcup_{n \in \mathbb{N}} A_n$ は $\bigcup\{A_n : n \in \mathbb{N}\}$ と表されることもある．

を余事象 (complement, complementary event) とよぶ．事象としての A^c は「A が起きない」ということである．明らかに $(A^c)^c = A$ である (演習).



3　可測集合族

measurable

基本事象の全体 Ω が区間や $\mathbb{R}^n$ の領域など非可算個の要素を含む場合には，Ω の部分集合のすべてを事象として考察しようとすると，一般には，この上に確率の理論がうまく構築できなくなくなってしまうことが知られている．そのために，事象の全体の族として捉えることのできる，Ω の部分集合族で適当なものを選んでおく必要が出てくる．次の意味での可測集合族は，そのような妥当な集合族として必要最低限の性質を持つものになっている．

$\mathcal{A} \subseteq \mathcal{P}(\Omega)$ が Ω 上の可測集合族 (family of measurable sets, または σ-代数 (σ-algebra)) であるとは，

(3.1)　$\emptyset, \Omega \in \mathcal{A}$; sigma-alg-0

(3.2)　$A_n \in \mathcal{A}$, $n \in \mathbb{N}$ なら[15]，$\bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$ である; sigma-alg-1

(3.3)　$A \in \mathcal{A}$ なら，$A^c \in \mathcal{A}$ である sigma-alg-2

が成り立つこととする．Ω を基本事象全体の集合として，$\mathcal{A}$ が Ω 上の可測集合族のとき，$(\Omega, \mathcal{A})$ を可測空間 (measurable space) とよぶ．

sigma-alg-L-0

補題 3.1 $(\Omega, \mathcal{A})$ を可測空間とする．このとき，

(a) $m \in \mathbb{N}$ で，$A_1, ..., A_m \in \mathcal{A}$ なら，$A_1 \cup \cdots \cup A_m \in \mathcal{A}$ である;

(b) $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ なら，$\bigcap_{n\in\mathbb{N}} A_n \in \mathcal{A}$ である;

(c) $A_1, ..., A_m \in \mathcal{A}$, $m \in \mathbb{N}$ なら，$A_1 \cap \cdots \cap A_m \in \mathcal{A}$ である.

(d) $A, B \in \mathcal{A}$ なら $A - B \in \mathcal{A}$ である[16].

---

[15] $\mathbb{N}$ で自然数の全体をあらわす．ただし，ここでは，自然数は 0 ではなく，1 から始まるものだとする．つまり $\mathbb{N} = \{1, 2, 3, \cdots\}$ である．

[16] 集合 A, B に対し，差集合 $A - B$ は $A - B = \{x \in A : x \notin B\}$ として定義される．この集合は $A \setminus B$ と表されることも多い．事象としては，これは「A だが B でない」に対応する．

証明. (a): $A_n \in \mathcal{A}$ がすべての $1 \le n \le m$ に対して成り立つとき，すべての $n > m$ に対して $A_n = \emptyset$ として，この $\mathcal{A}$ の要素の有限列を，$\mathcal{A}$ の要素の無限列に拡張する．ことのき，(3.1) により，すべての $n \in \mathbb{N}$ に対し $A_n \in \mathcal{A}$ が成り立つ．したがって，(3.2) により，

$$A_1 \cup \cdots \cup A_m = \bigcup_{n\in\mathbb{N}} A_n \in \mathcal{A}$$

である．

(b):

$$\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c = \bigcup_{n\in\mathbb{N}} (A_n)^c \tag{3.4}$$

だから[17]，(3.2), (3.3) により，$\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c \in \mathcal{A}$ である．したがって，(3.3) により，

$$\bigcap_{n\in\mathbb{N}} A_n = \left(\left(\bigcap_{n\in\mathbb{N}} A_n\right)^c\right)^c \in \mathcal{A} \tag{3.5}$$

である．

(c): (b) を用いて (a) と同様に証明できる．

(d): $A - B = A \cap B^c$ だから，(3.3) と (c) から，$A - B \in \mathcal{A}$ がわかる．（証明終り）

4 確率空間

$(\Omega, \mathcal{A})$ を可測空間とするとき，$\mathcal{A}$ から単位区間 $[0,1]$ への写像 probability-sp

$$P : \mathcal{A} \to [0,1] \tag{4.1}$$

sigma-alg-2-0

が $(\Omega, \mathcal{A})$ 上の確率測度 (probability measure) であるとは，以下の (4.2), (4.3) が成り立つことである．

(4.2) $P(\Omega) = 1$; sigma-alg-3

(4.3) $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ が互いに素なら[18]，等式 $P(\dot{\bigcup}_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n)$ が成り立つ[19]． sigma-alg-4

---

[17] 同じ要素を持つ集合は等しいので，この等式を示すには，すべての x に対して，$x \in \left(\bigcap_{n\in\mathbb{N}} A_n\right)^c$ と $x \in \bigcup_{n\in\mathbb{N}} (A_n)^c$ が同値になることを示せばよい．これは，

$$\begin{aligned} x \in \left(\textstyle\bigcap_{n\in\mathbb{N}} A_n\right)^c &\Leftrightarrow\ x \in \Omega \text{ かつ } x \in \textstyle\bigcap_{n\in\mathbb{N}} A_n \text{ でない} \\ &\Leftrightarrow\ x \in \Omega \text{ かつ (すべての } n \in \mathbb{N} \text{ に対し，} x \in A_n) \text{ でない} \\ &\Leftrightarrow\ \text{ある } n \in \mathbb{N} \text{ に対し } x \in (A_n)^c \\ &\Leftrightarrow x \in \textstyle\bigcap_{n\in\mathbb{N}} (A_n)^c \end{aligned}$$

によりよい．

[18] つまり，異る $m, n \in \mathbb{N}$ に対し，$A_m \cap A_n = \emptyset$ が常に成り立つなら．

[19] 互いに素な集合の列 A_n, $n \in \mathbb{N}$ の和集合であることを強調するときには記号 $\bigcup$ の上にドットを付けて $\dot{\bigcup}_{n\in\mathbb{N}} A_n$ と表すことにする．

P が可測空間 $(\Omega,\mathcal{A})$ 上の確率測度のとき，三つ組 $(\Omega,\mathcal{A},P)$ を確率空間とよぶ．

prob-Ex-4

例 4.1 (a) Ω を任意の空でない集合として，$\mathcal{A}=\{\Omega,\emptyset\}$ とし，$F:\mathcal{A}\to[0,\infty)$ を $F(\Omega)=1$, $F(\emptyset)=0$ で定義すると，$(\Omega,\mathcal{A},P)$ は確率空間になる．

(b) $\Omega=\{1,2,3,4,5,6\}$, $\mathcal{A}=\mathcal{P}(\Omega)$ として， $E\in\mathcal{A}$ に対し，$P(E)=\frac{1}{6}\#(E)$ とする．ただし，$\#(E)$ で E の要素の個数を表す．このとき，$(\Omega,\mathcal{A},P)$ は確率空間になっている．この $(\Omega,\mathcal{A},P)$ は，例 1.1, (b) や 例 1.2, (b) などでのサイコロを投げるときの事象の確率を表現する確率空間になっていると考えられる．

(c) $\Omega=[0,1000)$ として，$\mathcal{A}$ を $[0,1000)$ の部分開区間の全体から生成される σ-代数とする[20)]．このとき，$f:[0,1000)\to[0,\infty)$ を，$\int_0^{1000}f(x)dx=1$ となるような連続関数として，$A\in\mathcal{A}$ に対し，$P(A)=\int_A f(x)dx$ として P を定義すると，$(\Omega,\mathcal{A},P)$ は確率空間になる．

上の 例 4.1, (a) の確率空間はそれ自身としてはトリビアルであまり役に立ちそうには見えない．

可測空間 $(\Omega,\mathcal{A})$ の基本性質として，

(4.4) すべての基本事象 $\mathsf{e}\in\Omega$ に対し， $\{\mathsf{e}\}\in\mathcal{A}$ が成り立つ sigma-alg-4-0

ことも要請されることが多い．

以下では，表記を簡単にするために，次の集合論での標準的な記法を用いることにする: 自然数全体の “次” に来る数を ω とあらわすことにする．この記号を使うと，$n\in\mathbb{N}$ と，$0<n<\omega$ は同じ意味になる．この記号を使うと，

$w\le\omega$ として，列 A_n, $1\le n<w$ を考える．

というような表現を用いることで，有限または可算無限な列に関する議論がひとまとめにしてできるようになる．

補題 4.1 (a) Ω が有限または可算無限で，$(\Omega,\mathcal{A})$ を (4.4) を満たす可測空間とすると， $\mathcal{A}=\mathcal{P}(\Omega)$ である．

(b) $(\Omega,\mathcal{A})$ を，可測空間として，$1\le w\le\omega$ で，Ω の互いに異る要素の列 e_n, $1\le n<w$ に対し，$p_n\in[0,1]$, $1<n<w$ が $\sum_{1\le n<w}p_n=1$ を満たしているとき，$A\in\mathcal{A}$ に対し，$I_A=\{n:1\le n<w,\,e_n\in A\}$ として，

(4.5) $F(A)=\sum_{n\in I_A}e_n$

[20)] $\mathcal{X}\subseteq\mathcal{P}(\Omega)$ から生成される σ-代数は，$\mathcal{F}=\{\mathcal{B}\subseteq\mathcal{P}(\Omega):\mathcal{X}\subseteq\mathcal{B},\,\mathcal{B}$ は σ-代数 $\}$ として，$\mathcal{F}$ のすべての要素の共通部分として定義することができる (集合族 $\mathcal{F}$ のすべての要素の共通部分は $\bigcap\mathcal{F}$ と表されることもある)．

として $F : \mathcal{A} \to [0,1]$ を定義すると，$(\Omega, \mathcal{A}, F)$ は確率空間になる．

(c) $(\Omega, \mathcal{A}, F)$ を確率空間とする．$w \le \omega$ と $e_n \in \Omega$, $1 \le n < w$ に対して，$\{e_n\} \in \mathcal{A}$ で，$F(\{e_n\}) = p_n$ がすべての $1 \le n < w$ に対し成り立ち，$\sum_{1 \le n < w} p_n = 1$ となるとき，これらの $e_n \in \Omega$, $p_n \in [0,1]$, $1 \le n < w$ に対して (b) でのように定義された確率測度は，F と一致する．

確率空間の基本性質を調べておく:

prob-sp

補題 4.2 $(\Omega, \mathcal{A}, P)$ を確率空間とする．このとき次が成り立つ:

(a) $P(\emptyset) = 0$ である[21)]．

(b) (有限加法性) $m \in \mathbb{N}$ として，$A_1,\dots, A_m \in \mathcal{A}$ が互いに素なとき，補題 3.1, (a) により $A_1 \,\dot{\cup}\, \cdots \,\dot{\cup}\, A_m \in \mathcal{A}$ だが[22)]，このとき

$$P(A_1 \,\dot{\cup}\, \cdots \,\dot{\cup}\, A_m) = P(A_1) + \cdots + P(A_m)$$

が成り立つ．

(c) (単調性) $A, B \in \mathcal{A}$ で $A \subseteq B$ のとき，$P(A) \le P(B)$ である．

(d) (余事象の確率) $A \in \mathcal{A}$ のとき，$P(A^c) = 1 - P(A)$ である．

(e) (加法定理) 任意の (必ずしも互いに素でない) $A, B \in \mathcal{A}$ に対し，$P(A \cup B) = P(A) + P(B) - P(A \cap B)$ が成り立つ．

(f) (連続性) $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ が ($\subseteq$ に関する) 上昇列であるとき[23)]，$P(\bigcup_{n \in \mathbb{N}} A_n) = \lim_{n \to \infty} P(A_n)$ が成り立つ．

(g) (連続性) $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ が ($\subseteq$ に関する) 下降列であるとき[24)]，$P(\bigcap_{n \in \mathbb{N}} A_n) = \lim_{n \to \infty} P(A_n)$ が成り立つ．

証明． (a): すべての $n \in \mathbb{N}$ に対し，$A_n = \emptyset$ とすると，A_n, $n \in \mathbb{N}$ は互いに素だから，

$$P(\emptyset) = P(\dot{\bigcup}_{n \in \mathbb{N}} A_n) = \sum_{n \in \mathbb{N}} P(A_n) \tag{4.6}$$

sigma-alg-5

となり，$P(A_n) = P(\emptyset)$ だから，$P(\emptyset) = 0$ でなくてはならないことがわかる[25)]．

21) これは「$P(A) = 0$ なら $A = \emptyset$ である」ということを主張しているのではないことに注意する．たとえば，例 4.1, (c) で，各 $x \in [0, 1000)$ に対して，$P(\{x\}) = 0$ だが，もちろん $\{x\}$ は $\emptyset$ と異なる．

22) ここでも $\dot{\cup}$ で互いに素な集合の和集合であることを強調して表している．

23) つまり，$A_1 \subseteq A_2 \subseteq A_3 \subseteq \cdots$ が成り立つとき．

24) つまり，$A_1 \supseteq A_2 \supseteq A_3 \supseteq \cdots$ が成り立つとき．

25) もし $P(\emptyset) \ne 0$ なら，(4.6) から $P(\emptyset) = \infty$ とならなくてはならなくなるが，これは，(4.1) に矛盾である．

(b): 補題 3.1, (a) の証明のアイデアによる．$A_1,\ldots,A_m \in \mathcal{A}$ を互いに素として，すべての $n > m$ に対し $A_n = \emptyset$ とすると，$A_n, n \in \mathbb{N}$ は互いに素な $\mathcal{A}$ の要素の列となるから，(4.3) と (a) により，

$$P(A_1 \dot\cup \cdots \dot\cup A_m) = P(\dot\bigcup_{n\in\mathbb{N}} A_n) = \sum_{n\in\mathbb{N}} P(A_n) = P(A_1) + \cdots + P(A_m)$$

である．

(c): $A, B \in \mathcal{A}$ で $A \subseteq B$ なら，$B = A \cup (B - A)$ となり，A と $B - A$ は互いに素だから，(b) により，$P(B) = P(A) + P(B - A)$ となる．したがって，$0 \le P(B - A)$ により，$P(A) \le P(B)$ である．

(d): $\Omega = A \cup A^c$ で，A と A^c は互いに素だから，(4.2) と (b) により，$1 = P(\Omega) = P(A) + P(A^c)$ となるので，移項して $P(A^c) = 1 - P(A)$ である．

(e): $A' = A - (A \cap B)$, $B' = B - (A \cap B)$ とすると，$A = A' \dot\cup (A \cap B)$, $B = B' \dot\cup (A \cap B)$ で，$A \cup B = A' \dot\cup B' \dot\cup (A \cap B)$ である．したがって，(b) から，

$$\begin{aligned} P(A \cup B) &= P(A' \dot\cup B' \dot\cup (A \cap B)) = P(A') + P(B') + P(A \cap B) \\ &= P(A') + P(A \cap B) + P(B') + P(A \cap B) - P(A \cap B) \\ &= P(A) + P(B) - P(A \cap B) \end{aligned}$$

である．

(f): $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ を $\subseteq$ に関する上昇列とするとき，$B_1 = A_1$, $B_{n+1} = A_{n+1} - A_n$, $n \in \mathbb{N}$ とすると，$B_n, n \in \mathbb{N}$ は互いに素で，$\bigcup_{n\in\mathbb{N}} A_n = \dot\bigcup_{n\in\mathbb{N}} B_n$ である．したがって，(4.3) と (b) により，

$$\begin{aligned} P(\textstyle\bigcup_{n\in\mathbb{N}} A_n) &= P(\dot\bigcup_{n\in\mathbb{N}} B_n) = \textstyle\sum_{n\in\mathbb{N}} P(B_n) = \lim_{m\to\infty} (\sum_{n=1}^m P(B_n)) \\ &= \lim_{m\to\infty} P(B_1 \dot\cup \cdots \dot\cup B_m) \\ &= \lim_{m\to\infty} P(A_m) \end{aligned}$$

である．

(g): $A_n \in \mathcal{A}$, $n \in \mathbb{N}$ を下降列とすると，$n \in \mathbb{N}$ に対して $B_n = (A_n)^c$ とすると，$B_n, n \in \mathbb{N}$ は上昇列になる．$\bigcap_{n\in\mathbb{N}} A_n = (\bigcup_{n\in\mathbb{N}} B_n)^c$ だから，(d) と (f) により，

$$\begin{aligned} P(\textstyle\bigcap_{n\in\mathbb{N}} A_n) &= 1 - P(\textstyle\bigcup_{n\in\mathbb{N}} B_n) = 1 - \lim_{n\to\infty} P(B_n) = \lim_{n\to\infty}(1 - P(B_n)) \\ &= \lim_{n\to\infty} P(A_n) \end{aligned}$$

である． （証明終り）

5　条件付き確率

condi-prob

試行 t の結果 e が事象 A に属すとき，e が B にも属す確率を考える．ただし，A が起る確率は 0 でないとする．試行 t を n 回繰り返したとき，結果が A に属す回

数を $m(n)$ とし，A かつ B に族す数を $m'(n)$ とすれば，$\frac{m'(n)}{m(n)}$ の $n \to \infty$ での極限が (存在するならこれが)，この確率になるが， $P(A) \neq 0$ のときには，

$$\lim_{n\to\infty} \frac{m'(n)}{m(n)} = \lim_{n\to\infty} \frac{m'(n)/n}{m(n)/n} = \frac{\lim_{n\to\infty} m'(n)/n}{\lim_{n\to\infty} m(n)/n} = \frac{P(A\cap B)}{P(A)} \tag{5.1}$$ condi-prob-0

となるから，実際この極限は存在する．このような確率を，事象 A (が起きたこと) の制限付きの B の確率とよび，この確率を $P(B|A)$ とあらわすことにして，

$$P(B|A) = \frac{P(A\cap B)}{P(A)} \tag{5.2}$$ condi-prob-1

と定義する．試行 t に対応する確率空間が $(\Omega, \mathcal{A}, P)$ のとき，

$$\Omega_A = A,\ \ \mathcal{A}_A = \{B\cap A : B \in \mathcal{A}\},\ \ P_A : \mathcal{A}_A \to [0,1];\ B\cap A \mapsto P(B|A) \tag{5.3}$$ condi-prob-2

とすると，$(\Omega_A, \mathcal{A}_A, P_A)$ は確率空間になる (演習)．$(\Omega_A, \mathcal{A}_A, P_A)$ は，A の制限付きの確率に対応する確率空間である．

(5.2) を変形すると次が得られる:

product-rule

定理 5.1 (a) **(積の法則)** 事象 A, B に対し，$P(A) \neq 0$ なら，

$$P(A\cap B) = P(B|A)P(A) \tag{5.4}$$ condi-prob-3

が成り立つ．同様に $P(B) \neq 0$ なら，

$$P(A\cap B) = P(A|B)P(B) \tag{5.5}$$ condi-prob-4

である．

(b) **(ベイズの定理, Bayes' Theorem)** 事象 A, B に対し，$P(A) \neq 0$, $P(B) \neq 0$ なら，

$$P(B|A) = \frac{P(B)P(A|B)}{P(A)} \tag{5.6}$$ condi-prob-4-0

である．

証明． (a): $P(A) \neq 0$ なら，(5.2) の両辺を $P(A)$ 倍して (5.4) が得られる．(5.5) についても同様である．

(b): (a) により，$P(B)P(A|B) = P(A\cap B)$ だから，

$$P(B|A) = \frac{P(A\cap B)}{P(A)} = \frac{P(B)P(A|B)}{P(A)}$$

である． **(証明終り)**

定理 5.1 の拡張となっている次の定理も成り立つ:

chain-rule

定理 5.2 (連鎖法則) 事象 $A_1, \ldots, A_n$ $(n \in \mathbb{N}, 2 \leq n)$ に対して，$P(\bigcap_{1\leq i\leq n} A_i) \neq 0$ なら，

$$P(\textstyle\bigcap_{1\leq i\leq n} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\cdots P(A_n|\bigcap_{1\leq i<n} A_i) \tag{5.7}$$ condi-prob-5

が成り立つ．

証明. n に関する帰納法で示す．$n = 2$ のときには, (5.7) は 定理 5.1, (a) と一致する．$n = k$ に対して (5.7) が成り立つとして, 事象 $A_1, \ldots, A_{k+1}$ を $P(\bigcap_{1\leq i\leq k+1} A_i) \neq 0$ となるものとすると．$\bigcap_{1\leq i\leq k+1} A_i \subseteq \bigcap_{1\leq i\leq k} A_i$ だから，命題 4.2, (d) により，$P(\bigcap_{1\leq i\leq k} A_i) \neq 0$ である．したがって，帰納法の仮定から，

$$P(\textstyle\bigcap_{1\leq i\leq k} A_i) = P(A_1)P(A_2|A_1)P(A_3|A_1\cap A_2)\cdots P(A_k|\bigcap_{1\leq i<k} A_i) \tag{5.8}$$ condi-prob-6

である．一方，再び 定理 5.1, (a) により，

$$\begin{aligned} P(\textstyle\bigcap_{1\leq i\leq k+1} A_i) &= P((\textstyle\bigcap_{1\leq i\leq k} A_i)\cap A_{k+1}) \\ &= P(\textstyle\bigcap_{1\leq i\leq k} A_i)P(A_{k+1}|\textstyle\bigcap_{1\leq i\leq k} A_i) \end{aligned} \tag{5.9}$$ condi-prob-7

である．したがって，(5.8) と (5.9) から，

$$P(\textstyle\bigcap_{1\leq i\leq k+1} A_i) = P(A_1)P(A_2|A_1)\cdots P(A_k|\bigcap_{1\leq i<k} A_i)P(A_{k+1}|\bigcap_{1\leq i\leq k} A_i) \tag{5.10}$$

となるがこれが示したいことだった． **（証明終り）**

上のベイズの定理で, 更に $P(B) \neq 1$ のときには, $P(B^c) \neq 0$ だが, $B \,\dot{\cup}\, B^c = \Omega$ だから，$(A\cap B) \,\dot{\cup}\, (A\cap B^c) = A\cap(B \,\dot{\cup}\, B^c) = A\cap\Omega = A$ である．したがって，$P(A\cap B) + P(A\cap B^c) = P((A\cap B) \,\dot{\cup}\, (A\cap B^c)) = P(A)$ となるが，補題 5.1, (a) により $P(A\cap B) = P(B)P(A|B)$, $P(A\cap B^c) = P(B^c)P(A|B^c)$ だから，

$$P(A) = P(B)P(A|B) + P(B^c)P(A|B^c) \tag{5.11}$$ condi-prob-7-0

となる．これを (5.6) に代入すると，

$$P(B|A) = \frac{P(B)P(A|B)}{P(B)P(A|B) + P(B^c)P(A|B^c)} \tag{5.12}$$ condi-prob-8

が得られる．この等式がベイズの定理と呼ばれることもある．

Ex-0

例 5.1 ある症状 S を示す患者のうち 3% が疾患 D に罹っている．症状 S を持っている患者に検査 T を行なうと，D に罹っている者のうち 92% が陽性を示し，D にかかっていないもののうち 9% が陽性を示す．

症状 S を持っている患者が検査 T で陽性になったとき，D に罹っている確率を求める．

(S を示す患者のうちの 1 人無作為に選ぶという試行を行ったとき) 選ばれた患者が D に罹っているという事象を A とし，選ばられた患者が T で陽性を示すという事象を B とすると，患者が検査 T で陽性になったとき D に罹っている確率 $P(A|B)$ は，

$$P(A|B) = \frac{P(A)P(B|A)}{P(A)P(B|A)+P(A^c)P(B|A^c)} = \frac{0.03\times 0.92}{0.03\times 0.92 + 0.97\times 0.09} \approx 0.24$$

となる[26)]．

今もう 1 つの検査 T_1 に対し，症状 S を持っている患者に検査 T_1 を行なうと，D に罹っている者のうち 90% が陽性を示し，D にかかっていないもののうち 7% が陽性を示すとしてみる．

この場合にも T_1 の結果が陽性であるとき (この事象を B_1 と表すことにする) の患者が D に罹っている確率も，

$$P(A|B_1) = \frac{P(A)P(B_1|A)}{P(A)P(B_1|A)+P(A^c)P(B_1|A^c)} = \frac{0.03\times 0.9}{0.03\times 0.9 + 0.97\times 0.07} \approx 0.28$$

となって，あまり高いものになはらない．一方，T と T_1 を同じ患者に対して試行したとき，両方と陽性だった場合を考えてみる．ここで，T と T_1 の結果は，同じ患者に対して試行したときにも，A および A^c の制限のもとで，以下に述べるような意味で独立であると仮定すると[27)]，$P(B\times B_1|A) = 0.92\times 0.9$, $P(B\times B_1|A^c) = 0.09\times 0.07$ だから，

$$\begin{aligned} P(A|B\times B_1) &= \frac{P(A)P(B\times B_1|A)}{P(A)P(B\times B_1|A)+P(A^c)P(B\times B_1|A^c)} \\ &= \frac{0.03\times 0.92\times 0.9}{0.03\times 0.92\times 0.9 + 0.97\times 0.09\times 0.07} \approx 0.8 \end{aligned}$$

となり，この患者が疾患 D に罹っている確率は格段に高くなることがわかる．

これは，2 つの独立な検査の結果が陽性のときには，疾患に罹っている可能性はきはめて高い，という我々の直観に一致するものとなっている．

6　事象の独立性

indep

以下では，ある試行に対応する確率空間 $(\Omega, \mathcal{A}, P)$ を固定して考える．

L-indep-0

---

26) 「D に罹っている者のうち 92% が陽性を示し，D にかかっていないもののうち 9% が陽性を示す」という文章を読んだとき，T は D の患者を見出す検査としてそれほど悪くないのではないか，という印象を持つ人が多いのではないだろうか． この例では，ある症状 S を示す患者のうち疾患 D に罹っている者の割合が 3% と小さいことが， $P(A|B)$ が小さな値になることに寄与している．試しに，これを 40% としてみると，$P(A|B) = \frac{0.4\times 0.92}{0.4\times 0.6+0.97\times 0.09} \approx 0.87$ となる．

27) もちろん，この仮定が妥当なものかどうかは，T や T_1 の内容等に依存する事柄である．

補題 6.1 事象 $A, B \in \mathcal{A}$ で，$0 < P(A), P(B) < 1$ となるものに対し，次は同値である:

(6.1) $P(B|A) = P(B)$; indep-a-0

(6.2) $P(A \cap B) = P(A)P(B)$; indep-a-1

(6.3) $P(A|B) = P(A)$; indep-a-2

(6.4) $P(A^c|B) = P(A^c)$; indep-a-3

(6.5) $P(A|B^c) = P(A)$; indep-a-4

(6.6) $P(A^c|B^c) = P(A^c)$. indep-a-5

証明.

$$P(B|A) = \frac{P(B \cap A)}{P(A)} = \frac{P(A \cap B)}{P(A)}$$

だから，

$$P(B|A) = P(B) \Leftrightarrow \frac{P(A \cap B)}{P(A)} = P(B) \Leftrightarrow P(A \cap B) = P(A)P(B)$$
$$\Leftrightarrow \frac{P(A \cap B)}{P(B)} = P(A)$$

となる．したがって，(6.1) $\Leftrightarrow$ (6.2) $\Leftrightarrow$ (6.3) である．

一方，$P(A^c \cap B) = P(B) - P(A \cap B)$ だから，

$$\frac{P(A^c \cap B)}{P(B)} = \frac{P(B) - P(A \cap B)}{P(B)} = 1 - \frac{P(A \cap B)}{P(B)}$$

である．したがって，

$$P(A \cap B) = P(A)P(B) \Leftrightarrow 1 - \frac{P(A \cap B)}{P(B)} = 1 - P(A) \Leftrightarrow P(A^c|B) = P(A^c)$$

により， (6.2) $\Leftrightarrow$ (6.5) がいえた．(6.5) と (6.6) の同値性も同様に示せる (演習).

（証明終り）

事象 $A, B \in \mathcal{A}$ で，$0 < P(A), P(B) < 1$ となるものが (互いに) 独立とは，

(6.1) $P(B|A) = P(B)$

が成り立つこととする．補題 6.1 により，これは，(6.1) ∼ (6.6) のどれか (またはすべて) が成り立つということと同値である．特に，(6.1) と (6.3) の同値性から，(6.1) は A と B の間の対称な関係になっているので，「互いに独立」という表現は妥当である．[28)]

---

28) 集合 X 上の二項関係 R が対称であるとは，すべての a, b に対して，$a \, R \, b \Leftrightarrow b \, R \, a$ が成り立つことである．

7 確率変数

randomv

ある試行 t に対応する確率空間を $(\Omega, \mathcal{A}, P)$ とする．試行 t を行った結果 (基本事象) e の特性値の 1 つを返す "変数" を確率変数 (random variable) とよぶ．確率変数は X, Y,..., X_1, X_2,... などの大文字のアルファベットやそれに添字をつけたもので表すことにする．試行 t の結果の基本事象 e の X の値を $X(\mathsf{e})$ と表すことにして，実数 x に対し，事象[29] $A = \{\mathsf{e} \in \Omega : X(\mathsf{e}) \leq x\}$ の確率 $P(A)$ を $P(X \leq x)$ とあらわす．$P(X < x)$, $P(X = x)$ なども同様に定義する．

$P(X \leq x)$ で $x \in \mathbb{R}$ を変数と考えると，実数 x に実数 $P(X \leq x) \in [0, 1]$ を対応させる関数が得られる．x の関数と見たときの $P(X \leq x)$ を X の累積分布関数 (cumulative distribution function, c.d.f. と略) とよぶ．

X の c.d.f. は次の性質を満たす:

L-randomv-1

補題 7.1 (1) すべての $x \in \mathbb{R}$ に対し，$0 \leq P(X \leq x) \leq 1$ である．

(2) $\lim_{x\to-\infty} P(X \leq x) = 0$, $\lim_{x\to\infty} P(X \leq x) = 1$.

(3) $x_1 < x_2$ なら，$P(x \leq x_1) \leq P(X \leq x_2)$ である (= もあり得る).

(4) $P(X \leq x)$ は右連続である．つまり，すべての $a \in \mathbb{R}$ に対し，$\lim_{x\to a+0} P(X \leq x) = P(X \leq a)$ が成り立つ．

以下では，(7.1) と (7.2) の次の 2 つの場合を考える[30]．

(7.1) $P(X \leq x)$ は有限個または可算個の値しかとらない． discr

このときには，$D_X = \{a \in \mathbb{R} : \lim_{r\to a-0} P(X \leq r) \neq P(X \leq a)\}$ は $\mathbb{R}$ の可算部分集合になる．各 $d \in D_X$ に対し，$p_X(d) = P(x \leq a) - \lim_{r\to a-0} P(X \leq r)$ とすると，$p_X(d) > 0$ で，$\sum\{p_X(d) : d \in D_X\} = 1$ となることがわかる．$p_X(d) = P(X = d)$ である．

したがって，このときには，確率測度 P の可算加法性 (4.3) から，すべての $A \in \mathcal{A}$ に対し，$P(X \in A) = \sum\{p(d) : d \in A\}$ となる．このようなとき，X は離散的確率変数 (discrete random variable) であるといい，p_X を X の確率関数 (probability function) とよぶ．

ここで考えるもう 1 つの場合は次のものである:

(7.2) $P(X \leq x)$ は C_1-級である． conti

[29] X が確率変数と言ったときには，ここで考察しているような Ω の部分集合 A はすべて事象 $(\in \mathcal{A})$ になっているものとする．

[30] 測度論 (measure theory) と呼ばれる数学理論を駆使すると，ここで述べている，(7.1) と (7.2) の二分律 (dichotomy) — 特に (7.2) の場合 — はもっと改良されたものにできる．また，実際には (7.1) と (7.2) の混合を考察することが必要になる場合もある．

このときには，$f_X = \frac{d}{dx}P(X \leq x)$ とすると．微分積分学の基本定理から，

$$\int_{-\infty}^{a} f_X(x)dx = \int_{-\infty}^{a} \frac{d}{dx}P(X \leq x)dx = P(X \leq a) \tag{7.3}$$

となる．このことから，任意の区間 $[a,b]$ に対し，

$$\begin{aligned} P(a \leq X \leq b) &= P(X \leq b) - \lim_{x \to a-0} P(X \leq x) \\ &= \int_{-\infty}^{b} f_X(x)dx - \int_{-\infty}^{a} f_X(x)dx = \int_{a}^{b} f_X(x)dx \end{aligned} \tag{7.4}$$

となる．特に可測集合族 $\mathcal{A}$ が妥当に設定されているときには，このことから，すべての $A \in \mathcal{A}$ に対して，$P(A \in \mathcal{A}) = \int_A f_X(x)dx$ が成り立つ．このようなとき，X は連続確率変数 (continuous random variable) であるといい，f_X は X の確率密度関数 (probability dencity function, p.d.f.) である，という．

8 確率変数の期待値と分散

expected

試行 t を独立に n 回行ったときの X の返す n 個の値 $x_1,\ldots,x_n$ の平均値を考えてみる．X が離散的で，可算または有限な $D_X \subseteq \mathbb{R}$ に対する $p_X(d)$, $d \in D_X$ が X の確率関数のとき，n が十分に大きなときには，n 個の値のうち X が値 $d \in D_X$ をっている回数は，$n \cdot p_X(d)$ 回に近い回数になっている，と考えられる．したがって，$x_1,\ldots,x_n$ の平均は，$n \to \infty$ のとき $\frac{\sum_{d \in D_X} d \cdot n \cdot p_X(d)}{n} = \sum_{d \in D_X} d \cdot p_X(d)$ に近づくと考えてよいだろう．そこで，X の平均値 (mean) あるいは期待値 (expected value) を，

$$E(X) = \sum_{d \in D_X} d \cdot p_X(d) \tag{8.1}$$ e-1

と定義することにする．

同様な議論から，X が連続的なときには，f_X を X の p.d.f. として，

$$E(X) = \int_{-\infty}^{\infty} x f_X(x)dx \tag{8.2}$$ e-2

として X の平均値あるいは期待値を定義するのが妥当であることがわかる．X の平均値 (<u>m</u>ean) は μ または μ_X と表されることも多い．

連続関数 $\varphi : \mathbb{R} \to \mathbb{R}$ に対し[31]，$\varphi(X)$ で，X の返す値を x とするとき，$\varphi(x)$ を返すような確率変数をあらわす．

X が離散的で，$p_X(d)$, $d \in D_X$ が X の確率関数のとき，$\varphi''D_X = \{\varphi(d) : d \in D_X\}$ として，$p_{\varphi(X)}(e) = \sum\{p_X(d) : d \in D_X, \varphi(d) = e\}$ とすると，$p_{\varphi(X)}(e)$, $e \in \varphi''D_X$ が $\varphi(X)$ の確率関数となる．したがって，

$$(8.3) \quad E(\varphi(X)) = \textstyle\sum_{e \in \varphi'' D_X} e \cdot p_{\varphi(X)}(e) = \sum_{d \in D_X} \varphi(d) p_X(d)$$ e-3

である．

確率変数 X が連続的で f_X が X の p.d.f. のときには，$\varphi(X)$ の c.d.f. は $r \in \mathbb{R}$ に対し，$A_r = \{x \in \mathbb{R} : \varphi(x) \leq r\}$ として，

$$(8.4) \quad P(\varphi(X) \leq r) = \int_{A_r} f(x)dx$$

となるから，これに対応する p.d.f. $f_{\varphi(X)}$ によって，$E(\varphi(X)) = \int_{-\infty}^{\infty} x f_{\varphi(X)}(x)dx$ が $\varphi(X)$ の期待値となるが，これは，

$$(8.5) \quad E(\varphi(X)) = \int_{-\infty}^{\infty} \varphi(x) f_X(x)dx$$ e-4

によって計算できることがわかる[32)]．

L-e-1

補題 8.1 X を確率変数として，a, b, $c \in \mathbb{R}$ として，φ, $\psi : \mathbb{R} \to \mathbb{R}$ を連続関数とするとき，

$$(8.6) \quad E((a\varphi + b\psi + c)(X)) = aE(\varphi(X)) + bE(\psi(X)) + c$$ e-4-0

が成り立つ．

証明． X が連続的確率変数で，f_X が X の p.d.f. である場合について証明する (X が離散的な場合については演習とする)．

このときには，(8.5) から，

$$\begin{aligned} E((a\varphi + b\psi)(X)) &= \int_{-\infty}^{\infty} (a\varphi(x) + b\varphi(x) + c) f_X(x)dx \\ &= a\int_{-\infty}^{\infty} \varphi(x) f_X(x)dx + b\int_{-\infty}^{\infty} \psi(x) f_X(x)dx + c\int_{-\infty}^{\infty} f_X(x)dx \\ &= aE(\varphi(X)) + bE(\psi(X)) + c \end{aligned}$$

となるからよい． （証明終り）

確率変数 X に対し，$(X - \mu_X)^2$ の期待値 $E((X - \mu_X)^2)$ を X の分散 (variance) とよびこれを $V(X)$ または σ^2, ${\sigma_X}^2$ などで表す．以下の Chebychev の定理 (定理 8.2) が示すように，分散は X の帰す値の散らばりの程度を示す指標になっていると考えることができる．

補題 8.1 により，以下が成り立つ．

$$
\begin{aligned}
V(X) &= E((X-E(X))^2) \\
&= E(X^2 - 2E(X)X + (E(X))^2) \\
&= E(X^2) - 2(E(X))^2 + (E(X))^2 \\
&= E(X^2) - (E(X))^2.
\end{aligned}
\tag{8.7}
$$ e-5

e-Ex-1

例 8.1 再び例 4.1, (b) の例を考えてみる．X をサイコロの出た目の値を返す確率変数を考えることにすると (フェアなサイコロの場合のモデルで考えると) $P(X=1)=P(X=2)=\cdots=P(X=6)$ としてよく，事象 $X=1,\ldots,X=6$ は Ω の分割になっているから，例 4.1, (b) で与えた確率空間でのように，$P(X=1)=P(X=2)=\cdots=P(X=6)=\frac{1}{6}$ としてよい．

このときには，

$$
E(X) = 1\cdot\frac{1}{6} + 2\cdot\frac{1}{6} + \cdots + 6\cdot\frac{1}{6} = \frac{7}{2}
$$

である．一方

$$
E(X^2) = 1^2\cdot\frac{1}{6} + 2^2\cdot\frac{1}{6} + \cdots + 6^2\cdot\frac{1}{6} = \frac{91}{6}
$$

となるから，(8.7) から，$V(X)=E(X^2)-(E(X))^2=\frac{91}{6}-\frac{49}{4}=\frac{35}{12}\approx 2.92$ となる．

確率変数 X に対しては，次のような指標もよく用いられる:

分散の平方 $\sigma=\sigma_X=\sqrt{V(X)}$ を標準偏差 (standard deviation) とよぶ．

$$
\gamma_1 = \frac{E(X-\mu_X{}^3)}{\sigma_X{}^3}
\tag{8.8}
$$ e-6

を X の歪度 (わいど，skewness) とよぶ．X の歪度は，X の確率密度関数の平均値 μ_X を中心として見たときの非対称度 (歪み) の指標となっている (値の絶対値が大きいほど歪みが大きい)，と考えられる．

$$
\gamma_2 = \frac{E(X-\mu_X{}^4)}{\sigma_X{}^4}
\tag{8.9}
$$ e-7

を X の尖度 (せんど，kurtosis) とよぶ．X の尖度は，X の確率密度関数の平均値 μ_X を中心として見たときの裾野の広さの指標となっている，と考えられる (「尖度」という日本語の命名からの連想とは逆に値が大きいほど裾野が広い)．

補題 8.1 により，(8.8) と (8.9) の分子はそれぞれ

$$
E((X-(E(X))^3)) = E(X^3 - 3E(X)X^2 + 3(E(X))^2X - (E(X))^3)
\tag{8.10}
$$ e-8

[31] ここでは，連続関数という条件は，(8.2) の右辺の計算ができることの保証として用いられているだけなので，実際にはもっと弱めることができる．

[32] これはたとえば次のようにして見ることができる: $\varphi(X)$ の p.d.f. を離散的な確率関数で近似することを考えると，(8.5) はこの近似での (8.3) に対応することから，この近似の極限として，を考えると，(8.5) が得られる．

$$= E(X^3) - 3E(X)E(X^2) + 3(E(X))^2E(X) - (E(X))^3$$
$$= E(X^3) - 3E(X)E(X^2) + 2(E(X))^3,$$

e-9

$$\begin{aligned}(8.11)\quad & E((X - E(X))^4) \\ &= E(X^4 - 4E(X)X^3 + 6(E(X))^2X^2 - 4(E(X))^3X + (E(X))^4) \\ &= E(X^4) - 4E(X)E(X^3) + 6(E(X))^2E(X^2) - 4(E(X))^3E(X) + (E(X))^4 \\ &= E(X^4) - 4E(X)E(X^3) + 6(E(X))^2E(X^2) - 3(E(X))^4\end{aligned}$$

として計算できる．

chbyshev

定理 8.2 (チェビシェフの不等式，Chebyshev's inequality) X を任意の確率変数として，$k > 0$ とするとき．

e-9-a

$$(8.12)\quad P(|X - \mu_X| \geq k\sigma_X) \leq \tfrac{1}{k^2}$$

が成り立つ．

証明． X が連続的な場合について示す．f_X を X の p.d.f. として，$I = \{x \in \mathbb{R} : |x - \mu_x| \geq k\sigma_X\}$ とすると，

e-9-a-0

$$\begin{aligned}(8.13)\quad \sigma_X^2 &= \int_{-\infty}^{\infty} (x - \mu_X)^2 f(x)dx \\ &\geq \int_I (x - \mu_X)^2 f(x)dx \\ &\geq \int_I k^2{\sigma_X}^2 f(x)dx \\ &= k^2{\sigma_X}^2 \int_I f(x)dx \\ &= (k\sigma_X)^2 P(|X - \mu_X| \geq k\sigma_X)\end{aligned}$$

である．(8.13) の両辺を $(k\sigma_X)^2$ で割ると，$\frac{1}{k^2} \geq P(|X - \mu_X| \geq k\sigma_X)$ を得るが，これが求めていた不等式である．

X が離散的のときにも同様にして証明できる (演習)． (証明終り)

9 多変量確率分布[33)]

multi

X と Y を (同一の試行に対応する同一の確率空間での) 確率変数とする．このとき，実数 x, $y \in \mathbb{R}$ に対する事象 $A = \{t \in \Omega : X(t) \leq x$ かつ $Y(t) \leq y\}$ の確率 $P(A)$ を $P(X \leq x, Y \leq y)$ と書くことにする．ここでも，$P(X \leq x, Y \leq y)$ を x

33) multivariate probability distibution.

と y の関数と見たとき，これが離散的な場合と連続的な場合，およびその混合の場合 (たとえば X が離散的で Y が連続的な場合，または X と Y の 1 つ以上がそれ自身として離散的なものと連続的なものの混合となっている場合) が考えられるが，この最後の場合については，ここでは考察しないことにする．また，連続的な場合には，簡単のために，$P(X \leq x, Y \leq y)$ は C_2-級であることを仮定することにする[34)]．前者の場合，X と Y の同時確率分布は離散的である，といい，後者の場合には，X と Y の同時確率分布は連続的であるということにする．

X と Y の同時確率分布 $P(X \leq x, Y \leq y)$ が離散的な場合，特にこのときには X も Y も離散的となるので，$D_X \times D_Y = \{(x,y) : x \in D_X,\, y \in D_Y\}$ は有限または可算になる．各 $(x,y) \in D_X \times D_Y$ に対し，

$$p(x,y) = P(X = x, Y = y) \tag{9.1}$$ e-9-0

とするとき，$p(x,y)$, $(x,y) \in D_X \times D_Y$ を同時確率関数 (joint probability function) とよぶ．

このときには，$a, b, c, d \in \mathbb{R}$ で $a < b$, $c < d$ となるものに対して，

$$P(a \leq X \leq b, \leq Y \leq d) = \sum\{p(x,y) : a \leq x \leq b,\, c \leq x \leq d\} \tag{9.2}$$ e-10

となる．特に

$$\begin{aligned} P(X = x) &= P(X = x, -\infty < Y < \infty), \\ P(Y = y) &= P(-\infty < X < \infty, Y = y) \end{aligned} \tag{9.3}$$

だから，$x \in D_X$, $y \in D_Y$ に対して，

$$\begin{aligned} p_X(x) &= \sum\{p(x,y) : y \in D_Y\}, \\ p_Y(y) &= \sum\{p(x,y) : x \in D_X\} \end{aligned} \tag{9.4}$$ e-11

とすれば，$p_X(x)$, $x \in D_X$, と $p_Y(y)$, $h \in D_Y$ はそれぞれ確率変数 X と Y の確率関数になる．このような $p_X(x)$, $x \in D_X$, と $p_Y(y)$, $h \in D_Y$ を同時確率関数 $p(x,y)$, $(x,y) \in D_X \times D_Y$ の周辺確率関数 (marginal probability functions) とよぶ．

X と Y の同時確率分布が連続的な場合 — これは，ここでは，x, y を変数とする 2 変数関数 $P(X \leq x, Y \leq y)$ が C_2-級の場合のことと考えることにするが，このときには，

$$f_{X,Y}(x,y) = \frac{\partial^2}{\partial x \partial y} P(X \leq x, Y \leq y) \tag{9.5}$$ e-12

とすると，微分積分学の基本定理から，

$$P(X \le x, Y \le y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)du\,dv \tag{9.6}$$ e-13

となる．上のような $f_{X,Y}(x,y)$ を X と Y の同時確率密度関数 (joint probability density function) とよぶ．またこのときには，(9.4) でと同様に，

$$\begin{aligned} f_X(x) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dy \\ f_Y(y) &= \int_{-\infty}^{\infty} f_{X,Y}(x,y)dx \end{aligned} \tag{9.7}$$ e-14

が，それぞれ X と Y の確率密度関数となる． $f_X(x)$, $f_Y(y)$ はそれぞれ X と Y の周辺確率密度関数 (marginal probaility density function, m.p.d.f.) とよばれる．2 変数の連続関数 $\varphi : \mathbb{R}^2 \to \mathbb{R}$ に対し，試行 t の結果確率変数 X と Y が値 $X(\mathsf{t})$, $Y(\mathsf{t})$ を返したとき，$\varphi(X(\mathsf{t}), Y(\mathsf{t}))$ を返すような確率変数を考えて，この確率変数を，

$$\varphi(X,Y) \tag{9.8}$$ e-14-0

とあらわす．X, Y が離散的な場合には，$p(x,y)$ を (9.1) のようにとると，$D_{\varphi(X,Y)} = \{\varphi(x,y) : x \in D_X,\, y \in D_Y\}$ として，$d \in D_{\varphi(X,Y)}$ に対し，

$$p_{\varphi(X,Y)}(d) = \sum\{p(x,y) : x \in D_X,\, y \in D_Y,\, \varphi(x,y) = d\} \tag{9.9}$$ e-15

とすると，$p_{\varphi(X,Y)}(d)$, $d \in D_{\varphi(X,Y)}$ が $\varphi(X,Y)$ の確率関数となる．したがって，(8.1) から，

$$E(\varphi(X,Y)) = \textstyle\sum_{d \in D_{\varphi(X,Y)}} d \cdot p_{\varphi(X,Y)}(d) = \sum_{(x,y) \in D_X \times D_y} \varphi(x,y)p(x,y) \tag{9.10}$$ e-16

となることがわかる．

X と Y の同時確率分布が連続的な場合，$f_{X,Y}$ を X と Y の同時確率密関数とすると，$\varphi(X,Y)$ の c.d.f. は，$r \in \mathbb{R}$ に対して $A_{\varphi(X,Y) \le r} = \{(x,y) \in \mathbb{R}^2 : \varphi(x,y) \le r\}$ として，

$$P(\varphi(X,Y) \le r) = \int_{A_{\varphi(X,Y) \le r}} f_{X,Y}(x,y)dxdy \tag{9.11}$$ e-17

となる．これに対応する p.d.f. を求めて，これを用いて期待値 $E(\varphi(X,Y))$ を計算することができるが，(8.5) でと同様に，これはもっと直接的に

$$E(\varphi(X,Y)) = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x,y) f_{X,Y}(x,y)dxdy \tag{9.12}$$ e-18

---

[34] ここでも，測度論を用いると C_2-級の仮定を，もっと弱くてエレガントな条件で置き換えることができる．

によって計算できる．

確率変数 X と Y は，同時確率関数，または，同時確率密度関数が変数分離形をしているとき，つまり $p_{X,Y}(d,d') = p_X(d)p_Y(d')$, $(d,d') \in D_X \times D_Y$ と書けるとき，または，$f_{X,Y}(x,y) = f_X(x)f_Y(y)$, $(x,y) \in \mathbb{R}^2$ と書けるとき，独立 (independent) であるという．

L-e-2

定理 9.1 確率変数 X, Y が独立なら，すべての $a, b, c, d \in \mathbb{R} \cup \{-\infty, \infty\}$, $a < b$, $c < d$ に対し，

$$P(a \le X \le b,\, c \le Y \le d) = P(a \le X \le b)P(c \le Y \le d). \tag{9.13}$$ e-19

が成り立つ．

逆に，すべての $a, b, c, d \in \mathbb{R} \cup \{-\infty, \infty\}$, $a < b$, $c < d$ に対し，(9.13) が成り立つなら，X と Y は独立である，

証明. 同時確率変数 X, Y が連続的な場合について示す (離散的な場合は読者の演習とする)．

$$\begin{aligned} P(a \le X \le b,\, c \le Y \le d) &= \int_c^d \int_a^b f_{X,Y}(x,y)dxdy \\ &= \int_c^d \int_a^b f_X(x)f_Y(y)dxdy \\ &= \int_c^d \Big(f_Y(y)\int_a^b f_X(x)dx\Big)dy \\ &= \int_a^b f_X(x)dx \cdot \int_c^d f_Y(y)dy \\ &= P(a \le X \le b)P(c \le Y \le d) \end{aligned}$$

である．

逆に， 逆に，すべての $a, b, c, d \in \mathbb{R} \cup \{-\infty, \infty\}$, $a < b$, $c < d$ に対し，(9.13) が成り立つとすると，特に，すべての $a, b \in \mathbb{R}$ に対し

$$P(X \le a,\, Y \le b) = P(X \le a)P(Y \le b). \tag{9.14}$$ e-20

だから，$F_{X,Y}(x,y)$ を X と Y の同時累積分布関数として，$F_X(x)$, $F_Y(y)$ をそれぞれの周辺累積分布関するとすると，$F_{X,Y}(x,y) = F_X(x)F_Y(y)$ が成り立つ．このことから，$f_X(x)$, $f_Y(y)$ をそれぞれの周辺確率密度関数とすると， $f_X(x)f_Y(y)$ は X と Y の同時確率密度関数となることがわかるが，同時確率密度関数の一意性から[35]，$f_{X,Y}(x,y) = f_X(x)f_Y(y)$ がわかる． **(証明終り)**

[35] ここでは，確率密度関数や同時確率密度関数が連続になるように連続的な (同時) 確率変数の定義が設定されているので，一意性が成り立っている．

2 個以上の確率変数についての同時確率分布や独立性についても同様に定義する．例えば，連続的な場合には，$X_1,...X_n$ を確率変数とするとき，これらが (互いに) 独立であるとは，これらの同時確率密度関数 $f_{X_1,\ldots,X_n}(x_1,\ldots,x_n)$ が，

$$f_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = f_{X_1}(x_1)f_{X_2}(x_2)\cdots f_{X_n}(x_n) \tag{9.15}$$ e-21

と変数分離できること，とする．

L-e-3

補題 9.2 X, Y 確率変数として a, b, c を実数とするとき，

(1) $E(aX+bY+c) = aE(X)+b(Y)+c$ が成り立つ．

(2) $\varphi(x,y)$ が多項式 $a_{m,n}x^m y^n + \cdots$ のとき，$E(\varphi(X,Y)) = a_{m,n}E(X^mY^n)+\cdots$ である．

(3) X と Y が独立なら，任意の $m, n \in \mathbb{N}$ に対し，$E(X^mY^n) = E(X^m)E(Y^n)$ が成り立つ．

(4) X と Y が独立なら，任意の $a, b \in \mathbb{R}$ に対し，$V(aX+bY) = a^2V(X)+b^2V(Y)$ が成り立つ．

証明． (1): X, Y の同時確率分布が連続的な場合，この同時確率密度関数を $f_{X,Y}(x,y)$ とすると，

$$\begin{aligned}
E(aX+bY+c) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(ax+by+c)f_{X,Y}(x,y)dxdy \\
&= a\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}xf_{X,Y}(xy)dxdy \\
&\quad + b\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}yf_{X,Y}(xy)dxdy + c\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}f_{X,Y}(xy)dxdy \\
&= a\int_{-\infty}^{\infty}x\left(\int_{-\infty}^{\infty}f_{X,Y}(xy)dy\right)dx \\
&\quad + b\int_{-\infty}^{\infty}y\left(\int_{-\infty}^{\infty}f_{X,Y}(xy)dx\right)dy + c \\
&= aE(X)+bE(Y)+c
\end{aligned}$$

である．X, Y の同時確率分布が離散的のときにも，同様である (演習)．

(2): (1) から明らかである．

(3): 再び X と Y の同時確率分布が連続的なときについて考える．X と Y の同時確率密度関数が $f_X(x)f_Y(y)$ と表されているとして，

$$\begin{aligned}
E(X^mY^n) &= \int_{-\infty}^{\infty}\int_{-\infty}^{\infty}x^my^nf_X(x)f_Y(y)dxdy \\
&= \int_{-\infty}^{\infty}y^nf_Y(y)\left(\int_{-\infty}^{\infty}x^mf_X(x)dx\right)dy
\end{aligned}$$

$$
\begin{aligned}
&= \int_{-\infty}^{\infty} x^m f_X(x)dx \cdot \int_{-\infty}^{\infty} y^n f_Y(y)dy \\
&= E(X^m)E(Y^n)
\end{aligned}
$$

である．

(4): (8.7) と (3) から，

$$
\begin{aligned}
V(aX+bY) &= E((aX+bY)^2) - (E(aX+bY))^2 \\
&= a^2E(X^2) + 2abE(XY) + b^2E(Y^2) - (aE(X)+bE(Y))^2 \\
&= a^2E(X^2) + b^2E(X^2) - a^2(E(X))^2 - b^2(E(Y))^2 \\
&= a^2(E(X^2) - (E(X))^2) + b^2(E(Y^2) - (E(Y))^2) \\
&= a^2V(X) + b^2V(Y)
\end{aligned}
\tag{9.16}
$$

e-22

である． （証明終り）

10 共分散と相関係数

covariance

X と Y を (同一の試行に対応する同一の確率空間での) 2 つの確率変数とするとき， X と Y の共分散 (covariance) $Cov(X,Y)$ ($\sigma_{X,Y}$ とも表す) を

$$
Cov(X,Y) = \sigma_{X,Y} = E((X-\mu_X)(Y-\mu_Y))
\tag{10.1}
$$

で定義する[36)]．

L-cov-1

補題 10.1 X と Y を確率変数とする．このとき

(1) $Cov(X,Y) = E(XY) - E(X)E(Y)$ である．

(2) X と Y が独立なら，$Cov(X,Y) = 0$ である．

(3) ある実数 a, b, $a \neq 0$ に対し，$Y = aX+b$ なら，$Cov(X,Y) = \dfrac{a}{|a|}\sqrt{V(X)V(Y)}$ である．

証明． (1): 補題 9.2 を用いると，

$$
\begin{aligned}
Cov(X,Y) &= E((X-E(X))(Y-E(Y))) \\
&= E(XY - E(X)Y - E(Y)X + E(X)E(Y)) \\
&= E(XY) - E(X)E(Y) - E(Y)E(X) + E(X)E(Y) \\
&= E(XY) - E(X)E(Y)
\end{aligned}
$$

---

36) ここでの $E((X-\mu_X)(Y-\mu_Y))$ は，(9.8) の意味での確率変数 $(X-\mu_X)(Y-\mu_Y)$ の期待値を表していることに注意する．

である．

(2): 補題 9.2, (3) と 上の (1) から導かれる．

(3): 補題 9.2, (1) から，

$$
\begin{aligned}
Cov(X,Y) &= E((X-E(X))(aX+b-(aE(X)+b)) \\
&= E(aX^2 - aE(X)X - (aE(X)X - a(E(X))^2) \\
&= aE(X^2) - a(E(X))^2 - a(E(X))^2 + a(E(X))^2 \\
&= a((E(X^2)-(E(X))^2) = \frac{a}{|a|}\sqrt{V(X)V(Y)}
\end{aligned}
$$

である． **(証明終り)**

X と Y の相関係数 (correlation coefficient) $\rho_{X,Y}$ を

$$
\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)V(Y)}} \quad \left(= \frac{\sigma_{X,Y}}{\sigma_X \sigma_Y}\right) \tag{10.2}
$$

と定義する．

L-cov-2

補題 10.2 確率変数 X, Y に対し，次が成り立つ:

(1) $-1 \le \rho_{X,Y} \le 1$;

(2) X と Y が独立なら，$\rho_{X,Y} = 0$;

(3) ある実数 $a, b \in \mathbb{R}$, $a \neq 0$ に対し $Y = aX + b$ なら，$|\rho_{X,Y}| = 1$ で，このとき $a > 0$ なら $\rho_{X,Y} = 1$ となり，$a < 0$ なら $\rho_{X,Y} = -1$ となる．

証明． (2) と (3) はそれぞれ，補題 10.1, (2), (3) によりよい．

(1) の証明は略す[37]． **(証明終り)**

11 大数の法則

LLN

定理 11.1 (大数の法則, Law of large numbers) X_1, X_2,... が独立な確率変数で，ある $\mu \in \mathbb{R}$ と $\sigma^2 > 0$ に対し，$E(X_1) = E(X_2) = \cdots = \mu$, $V(X_1), V(X_2), \cdots \le \sigma^2$ となるとき， Th-large-n-1

$$
\lim_{n\to\infty} P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \ge \varepsilon\right) = 0 \tag{11.1}
$$

large-n-0

がすべての $\varepsilon > 0$ に対して成り立つ．つまり，n が大きくなるほど，$\dfrac{X_1 + \cdots + X_n}{n}$ の値が μ から大きくはずれる確率は小さくなる．

[37] 証明は，Schwarz の不等式 $(E(XY))^2 \le E(X^2)E(Y^2)$ から導ける．

証明. $n \in \mathbb{N}$ に対し，$Y_n = \dfrac{X_1 + \cdots + X_n}{n}$ と置く．このとき 補題 9.2, (1), (4) から，$E(Y_n) = \mu$, ${\sigma_n}^2 = V(Y_n) \leq \dfrac{\sigma^2}{n}$ となることがわかる．

したがって，チェビシェフの不等式から，

$$P\left(\left|\frac{X_1 + \cdots + X_n}{n} - \mu\right| \geq \varepsilon\right) = P\left(|Y_n - \mu| \geq \frac{\varepsilon}{\sigma_n} \cdot \sigma_n\right) \leq \frac{{\sigma_n}^2}{\varepsilon^2} \leq \frac{1}{n} \cdot \frac{\sigma^2}{\varepsilon^2} \tag{11.2}$$ large-n-1

となるが，不等式の右辺は $n \to \infty$ のとき $\to 0$ となる． （証明終り）

12 正規分布

normal

連続関数 $f : \mathbb{R} \to [0, \infty)$ が分布関数であるとは，f が $\int_{-\infty}^{\infty} f(x)dx = 1$ を満たすこととする．実数の組 $(\mu, \sigma^2) \in \mathbb{R} \times (0, \infty)$ に対して分布関数 $f = \mathcal{D}(\mu, \sigma^2)$ で，$\int_{-\infty}^{\infty} f(x)xdx = \mu$, $\int_{-\infty}^{\infty}(f(x) - \mu)^2 dx = \sigma^2$ となるようなものを対応させる写像 $\mathcal{D}$ を分布 (distribution) とよぶ．実際に扱われる分布は，更にいくつかの良い条件を満たすものとなっていることが多いが，ここでは，これに関する議論は省略する．

確率変数 X が分布 $\mathcal{D}$ に従う，とは $\mu = E(X)$, $\sigma^2 = V(X)$ として，X の確率密度関数が，$\mathcal{D}(\mu, \sigma^2)$ と一致することとする．

X が $\mathcal{D}$ に従い，$\mu = E(X)$, $\sigma^2 = V(X)$ となっていることを $X \sim \mathcal{D}(\mu, \sigma^2)$ と表す．

確率論全体で中心的な役割をはたす分布に，正規分布 (normal distribution) $\mathcal{N}(\mu, \sigma^2)$ がある．これは，

$$\mathcal{N}(\mu, \sigma^2) = \frac{1}{\sqrt{2\pi}\,\sigma} exp(-(x - \mu)^2/2\sigma^2) \tag{12.1}$$ nd-0

として与えられる．

13 中心極限定理

CLT