統計学への確率論 その先へ　清水 泰隆
ゼロからの測度論的理解と漸近理論への架け橋

σ - 加法族、確率変数、頻度論的確率、公理論的確率、線形変換、a.s.(ほとんど確実に)、歪度、尖度、概収束、弱収束、連続写像定理、「中心極限定理」の名前の由来、

本書は，本格的な数理統計学を目標とする読者向けに，特に統計学で重要となる事柄に重点をおき，速習的に確率論を学ぶことができる**学部生向け教科書**を目指した．微積分，線形代数と集合演算に関する多少の知識のみを前提とし，確率論を学びながら測度論の重要ポイントを押さえることにより，基礎となる理論の内容をおろそかにすることなく，最短経路で「測度論的確率論」の概略をつかみ，その先の「大標本理論」にスムーズに移行できるよう，その橋渡しとなることを目標とした．

このような内容の執筆を思ったのは，早稲田大学基幹理工学部・**応用数理学科**における 3 年生向けの通年講義「確率統計概論」を授業したことがきっかけである．

$\Omega = \mathbb{R}$　実数 $\mathbb{R}$

このように，標本全体の集合を，確率論では通常 Ω という記号で表し，これを**標本空間** (sample space) という．

σ - 加法族

∪

このようにして式 (1.2) の $\mathcal{A}$ から作られた $\sigma(\mathcal{A})$ は，特に，$\mathbb{R}$ 上の**ボレル集合体** (Borel field) といわれ，

$$\mathcal{B} := \sigma(\mathcal{A})$$

などと書かれる．ここで $:=$ は，その右辺を左辺の記号で書く，という意味である．

「確率変数（確率ベクトル）」はその呼び名とは裏腹に，定義に“確率”の概念は不要であることに注意されたい．実際，我々はまだ“確率”をちゃんと定義していない．英語でも「random variable（vector）」であって，“確率”（probability や stochastic）という語は使われない．

～～～～～～～～～～～

確率変数と確率の違い

確率変数は事象を介して確率を付与された変数。

https://www.koubundou.co.jp/files/00087.pdf

〜〜〜〜〜〜〜〜〜〜

頻度論的確率、公理論的確率

1.2.2 “確率”とは何か？：確率と確率空間

日常的に用いられる “確率”や高校で学習した “確率”は多くの場合，

$$(\text{確率}) = \frac{(\text{対象となる場合の数})}{(\text{起こり得るすべての場合の数})}$$

のように，起こり得る頻度の比をもって確率が定義される．これはラプラス[*5]流の “頻度論的確率”といわれる．このような定義は，個々の事象の数を数えられるような場合には広く合意が得られているであろう．本節では，このような確率が持つ自然な性質を失うことなく，より一般の事象に対する確率を考えていこう．

ラプラス流の確率に対して，コルモゴロフ[*6]は確率を以下のような公理を満たす σ-加法族上の写像として定義した．このような定義を用いるのが現代確率論で，“公理論的確率論”ともいう．

定義 1.2.6 (確率の公理 I)**.** 有限加法族 $\mathcal{F}$ が与えられたとき，次の (1)，(2) を満たす写像 $\mathbb{P} : \mathcal{F} \to [0,1]$ を $\mathcal{F}$ 上の (**有限加法的**) **確率**という．

(1) (**全確率**) $\mathbb{P}(\Omega) = 1$.

(2) (**有限加法性**) $A, B \in \mathcal{F}$ が $A \cap B = \emptyset$ を満たせば，

$$\mathbb{P}(A \cup B) = \mathbb{P}(A) + \mathbb{P}(B). \tag{1.5}$$

$\mathbb{P}$ は神が ω を引き抜く際の一種の**法則** (law) であり，この意味で確率のことを**確率法則** (probability law) と呼ぶこともある．

定理 1.2.21 (ベイズの定理)**.** $A_i \in \mathcal{F}\ (i=1,\dots,n)$ に対して，Ω が $(A_i)_{i=1}^n$ の直和分割

$$\Omega = \bigcup_{i=1}^{n} A_i, \quad A_i \cap A_j \neq \emptyset\ (i \neq j)$$

とする．ただし，$\mathbb{P}(A_i) > 0\ (i=1,\dots,n)$ である．このとき，$B \in \mathcal{F}$ が $\mathbb{P}(B) > 0$ を満たせば，

$$\mathbb{P}(A_j|B) = \frac{\mathbb{P}(A_j)\mathbb{P}(B|A_j)}{\sum_{i=1}^{n} \mathbb{P}(A_i)\mathbb{P}(B|A_i)}.$$

演習 6. 定理 1.2.21 を示せ．

注意 1.2.22. ベイズの定理の特徴は，左辺と右辺で条件が逆転しているところである．「原因 A_j があって事象 B が起こる」という因果関係による確率は通常求めやすいことが多い．これを使って「事象 B が起こったときにその原因が A_j であった確率」を与えるのが，ベイズの定理であり，このような確率を**逆確率** (inverse probability) という．

ベイズの定理の有名な応用例をあげておこう．

例 1.2.23 (感染者問題)**.** 50 代の日本人が大腸がんになる確率は 0.1%といわれている．大腸がん検診を受けた場合，本当にがんである場合，99%の確率で陽性反応が出て，がんでない場合には 98%の確率で陰性反応となることがわかっている．この下で，ある人がこのがん検診を受けて陽性反応が出た場合に，本当に大腸がんにかかっている確率を求めたい．

そこで各事象を以下のように設定する．

- T: 大腸がんにかかっているという事象
- P: 検査で陽性反応が出るという事象
- N: 検査で陰性反応が出るという事象

今わかっていることは，「原因」→「結果」の確率

$$\mathbb{P}(T) = 0.001, \quad \mathbb{P}(P|T) = 0.99, \quad \mathbb{P}(P|T^c) = 0.02, \quad \mathbb{P}(N|T^c) = 0.98$$

などであるが，知りたいのは逆確率 $\mathbb{P}(T|P)$ である．そこでベイズの定理を用いると

$$\mathbb{P}(T|P) = \frac{\mathbb{P}(T)\mathbb{P}(P|T)}{\mathbb{P}(T)\mathbb{P}(P|T) + \mathbb{P}(T^c)\mathbb{P}(P|T^c)}$$
$$= \frac{0.001 \times 0.99}{0.001 \times 0.99 + 0.999 \times 0.02} = 0.04721.$$

つまり，たとえ陽性反応が出たとしても実際にがんである確率はわずか4.7%ほどしかないのである．これは，もともとのがん患者が相当に少ない (だから $\mathbb{P}(T \cap P)$ が小さい) ことが原因である．したがって，信頼できそうな検査で陽性だからといってすぐに悲観しなくてもよいかもしれない．がんなら検査でほぼ確実に引っかかるが，逆はかならずしも真ならず！

例 1.2.24 (ベイズ更新)**.** 「原因」の確率を考える際に，得られた情報 (結果) を基に，ベイズの定理を用いて確率を更新していくという考え方がある．最初に考える (しばしば観測者が仮定する)「原因」の確率を**事前確率** (prior probability) といい，一定の情報 (結果) を得た後に，その情報によって更新された確率を**事後確率** (posterior probability) という．

例1.2.23において，50代の日本人が「大腸がんになる (T)」確率は0.1%であるから，がん検診を受ける前であれば，自分が大腸がんである確率も $\mathbb{P}(T) = 0.001$ だと思うだろう．これが情報を得る前の「事前確率」である．その後，がん検診を受けて「陽性 (P)」という情報を得ることにより，自分ががんである確率が $\mathbb{P}(T|P) = 0.0472103$ に更新され，これが更新された「事後確率」である．

ここで，上記の $\mathbb{P}(T) = 0.001$ という確率が，実はいい加減な調査結果に基づくもので，あまり信頼できない数値であったとしよう．例えば，ある製薬メーカーが自社の薬で大腸がんが減ることを強調したいために，一般的な大腸がんの確率を大きめに表記していたとする．このような疑いがあるが，他にはあまり情報がないとき，例えば確率を少し小さめに見積もって，

$$\mathbb{P}(T) = 0.0007$$

と考えてみる．これは我々の気持ちが入った主観的な事前確率である．このとき事後確率を計算すると，

$$\mathbb{P}(T|P) = 0.03351$$

となり，当初の確率よりもっと小さくなる．自分の期待を込めると，自分が大腸がんであるという確信が弱まることが確率の減少として見える．このように思考のプロセスを，確率を通して表現できるのがベイズ法の特徴である．

このようにベイズの考え方では，事前確率の定め方によって事後確率が異なってくる．事前確率はその人が思う主観的な確率であり**主観確率**ともいわれる[*10]．このことは一見普遍性がないようにも見えるが，一方，事前の情報 (あるいは思い込み) によって期待の程度が変わるという点で，人間の思考に近いプロセスを実現しているとも考えられる．

注意 1.2.25. 本書の目標の 1 つは「大標本理論」への接続であるが，これが統計学への確率論的正当性を与えたことで現代数理統計学が確立された．これらはしばしば**頻度論**と呼ばれ，教養などで学ぶ初等統計はこの頻度論的手法が主である．一方，近年の統計学ではベイズの定理を基礎にした**ベイズ推定**の手法が強力なツールとして盛んに用いられるようになってきた．ベイズ流は人の直感に素直な一面があり，機械学習や AI にも用いられるなど，実用上もその有用性が認められている．古くから「**頻度主義** (frequentist) **vs ベイズ主義** (Baysian)」といった哲学的論争もあるが，ベイズ法の正当化の手段として頻度論が用いられることもあり，頻度論的手法とベイズ法は現代統計学の両輪としていずれも不可欠といえる．

例 1.3.15 (線形変換)**.** $X \sim N(\mu, \sigma)$ $(\mu \in \mathbb{R}, \sigma > 0)$ とし，

$$Z = \frac{X - \mu}{\sigma}$$

なる線形変換を考える．定理 1.2.5 によれば Z はまた確率変数である．Z の分布を求めてみよう．そのためには，分布関数を計算すればよい．

$$\begin{aligned}
\mathbb{P}(Z \leq x) &= \mathbb{P}((X-\mu)/\sigma \leq x) = \mathbb{P}\left(X \leq \sigma x + \mu\right) \\
&= \int_{-\infty}^{\sigma x + \mu} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(y-\mu)^2}{2\sigma^2}} \, \mathrm{d}y \quad (y = \sigma z + \mu \text{ と置換すると}) \\
&= \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}} \, \mathrm{d}z
\end{aligned}$$

となり，$f_Z(y) = \dfrac{1}{\sqrt{2\pi}} e^{-\frac{z^2}{2}}$ である．したがって，

$$X = \sigma Z + \mu \sim N(\mu, \sigma^2) \quad \Leftrightarrow \quad Z = \frac{X - \mu}{\sigma} \sim N(0, 1).$$

である．このような線形変換については後述の定理 2.3.2, (3)も参照のこと．

〜〜〜〜〜〜〜

線形変換

線形変換（線型写像）とは、簡単に表現すると「行列によって空間（線形空間）を変形させること」です。

https://www.headboost.jp/what-is-linear-transformation/

〜〜〜〜〜〜〜

1.3.5 ほとんど確実に？

確率空間 $(\Omega, \mathcal{F}, \mathbb{P})$ において，根元事象 $\omega \in \Omega$ に関する命題 $\mathcal{P}(\omega)$ が

$$\mathbb{P}(\{\omega \in \Omega \,|\, \mathcal{P}(\omega) \text{ が真である}\}) = 1$$

を満たすとき，「**ほとんど確実に** (almost surely) 命題 $\mathcal{P}$ が成り立つ」などと表現する．

例えば，$X \sim N(0,1)$ であるような確率変数 X に対しては，

$$\mathbb{P}(X^2 > 0) = \int_{\{x \neq 0\}} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, \mathrm{d}x = 1$$

であるので

$$\text{ほとんど確実に } X^2 > 0 \text{ である}$$

と表現する．これを記号で以下のように表す．

$$X^2 > 0 \quad a.s.$$

$a.s.$ とは “almost surely” の頭文字である．

なぜ「確実に」ではなく，「ほとんど確実に」なのか？この意味は，$X^2 = 0$ となることもあるにはあるのだが，そのような確率は 0 でほとんど起こり得ない，という気持ちである．もう少し正確に書くと，ある $\mathbb{P}$-零集合 $N \in \mathcal{F}$ が存

◆歪度、尖度

2.3.1 積率 (モーメント)

定義 2.3.1. 実数値確率変数 X と $k \in \mathbb{N}$ に対して，

$$\mu_k := \mathbb{E}[X^k], \qquad \alpha_k := \mathbb{E}[(X - \mu_1)^k]$$

とおく．このとき，μ_k を k 次の**積率** (**モーメント**, moment)，α_k を k 次の**中心積率** (central moment) という．特に，以下のような呼び名がある．

- $\mu := \mu_1$: **平均** (mean)
- $\sigma^2 := \mathrm{Var}(X) := \alpha_2$: **分散** (variance)
- $\sigma := \sqrt{\mathrm{Var}(X)}$: **標準偏差** (standard deviation)
- $\gamma_1 := \alpha_3/\sigma^3$: **歪度** (skewness)
- $\gamma_2 := \alpha_4/\sigma^4$: **尖度** (kurtosis)
  ($\gamma_2' = \alpha_4/\sigma^4 - 3$ とすることもある：演習 14)

平均は分布の “中心” (密度関数の重心) を表す量であり，分散はその “中心” からのばらつき (広がり) を表す量である．平均と分散がわかれば，その分布のおおよその形がイメージできる．

分布の形をさらに詳しく検討するには歪度，尖度を見ることも重要である．密度関数が平均の軸に関して対称ならば $\alpha_3 = 0$ となるので，γ_1 が 0 に近いほど，直感的に分布は左右対称に近く，$\gamma_1 > 0$ となると右に裾を引く形状になる．また，密度関数 f の面積は 1 と決まっているので，密度関数の中心付近が尖って細くなれば，その分密度関数の左右の裾が厚くなり，α_4 の値は大きくなるであろう．つまり，γ_2 が大きいほど，直感的に密度関数は尖って見えると考えられる (**図 2.4**)．ただ，この説明は直感的なもので図も典型的なものを描いており，必ずしもこの説明に合わないこともあるので，あくまで目安の量と思って

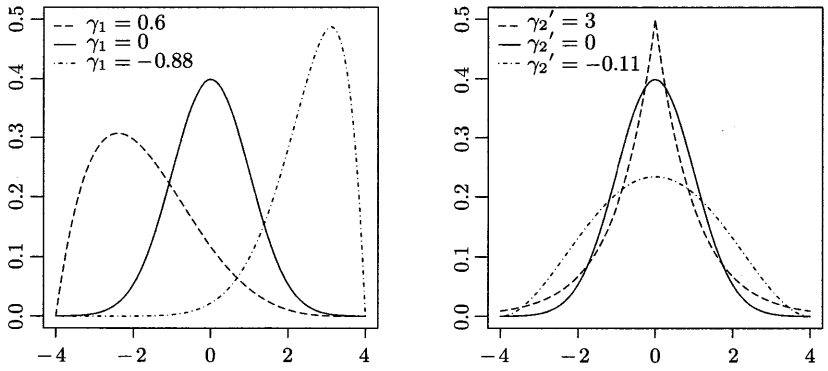

図 2.4 密度関数の歪度 (左図)，尖度 (右図) による違い.

おくべきである.

(3) $Z := \dfrac{X-\mu}{\sigma}$ とおくと,

$$\mathbb{E}[Z] = 0, \quad \mathrm{Var}(Z) = 1.$$

X から Z への変換を**正規化** (normalization)[*5]という.

[*5] **標準化** (standardization) ともいう.

3.2.1 相関と相関係数

定義 3.2.1. 実数値確率変数 X, Y の平均をそれぞれ μ_X, μ_Y とするとき,

$$\mathrm{Cov}(X,Y) := \mathbb{E}\left[(X-\mu_X)(Y-\mu_Y)\right]$$

を X, Y の**共分散** (covariance) という. 特に, $\mathrm{Cov}(X,X) = \mathrm{Var}(X)$ である. このとき,

$$\rho(X,Y) := \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)}\sqrt{\mathrm{Var}(Y)}}$$

を X, Y の**相関係数** (correlation)[*3]という. ただし, $\mathrm{Var}(X)\mathrm{Var}(Y) = 0$ のときは $\rho(X,Y) = 0$ と定める. 特に, $\rho(X,Y) \neq 0$ のとき, X と Y には**相関がある**といい, $\rho(X,Y) > 0$ なら**正の相関**, $\rho(X,Y) < 0$ なら**負の相関**, $\rho(X,Y) = 0$ のときは**無相関**といわれる.

[*3] **ピアソン** (Peason) **の積率相関係数**ともいわれる. 統計学には複数の「相関係数」があり, 他にも, スピアマンの順位相関係数 (スピアマンの ρ), ケンドールの順位相関係数 (ケンドールの τ), などと呼ばれるものがある. 単に「相関係数」という場合, 通常はピアソンの相関係数を指す.

定義 4.1.3. d 次元確率ベクトルの列 $\{X_n\}$, X に対して,

$$\mathbb{P}\left(\lim_{n\to\infty} X_n = X\right) = 1$$

となるとき, X_n は X に**概収束** (almost-sure convergence) するといい,

$$X_n \xrightarrow{a.s.} X \quad (n \to \infty)$$

と書く. 一般には, 1.3.5 節の表記に従って $\lim_{n\to\infty} X_n = X$ $a.s.$ とか, $X_n \to X$ $a.s.$ などと書くことが多い.

注意 4.2.6. X_n, X の分布をそれぞれ $P_n := \mathbb{P} \circ X_n^{-1}$, $P := \mathbb{P} \circ X^{-1}$ として, $X_n \xrightarrow{d} X$ の定義を書き換えると,

$$\lim_{n\to\infty} \int_{\mathbb{R}} f(x)\, P_n(\mathrm{d}x) = \int_{\mathbb{R}} f(x)\, P(\mathrm{d}x)$$

であり, ここから「分布 P_n が P へ収束する」という雰囲気が理解されるだろう. これを $P_n \Rightarrow P$ などと書いて, 確率測度 (分布) P_n の**弱収束** (weak convergence) という. つまり, $X_n \xrightarrow{d} X$ とは分布の弱収束である.

4.3.3 連続写像定理

以下の定理は**連続写像定理** (continuous mapping theorem) といわれ, 統計学では頻繁に用いられる.

定理 4.3.5 (連続写像定理)**.** d 次元確率ベクトルの列 $\{X_n\}, X$ が

$$X_n \xrightarrow{*} X, \quad n \to \infty$$

を満たすとする. ただし, 収束の意味は $* = a.s.$, p, または d である. このとき, $\mathbb{R}^d$ 上の任意 $f \in C(\mathbb{R}^d)$ に対して[*7], 同じ意味での収束

$$f(X_n) \xrightarrow{*} f(X), \quad n \to \infty$$

が成り立つ.

独立同一分布 (i.i.d.)

independent and identically distributed

$$I_N := \frac{1}{N}\sum_{i=1}^{N} (b-a) f(X_i) \xrightarrow{a.s.} I, \quad N \to \infty$$

である. この概収束によって, N を十分大きくとることによって I の近似値を求めることができる. このような乱数を用いた積分計算を**モンテカルロ法** (Monte Carlo simulation) という. 上の原理は多次元の図形の面積・体積を求めるのにも利用できる (演習 30).

「中心極限定理」の名前の由来

注意 5.2.4. IID 列の和の収束率が $\sqrt{n}$ であることから派生して，統計学で現れる “よい”統計量 (推定量) というものは $\sqrt{n}$ の収束率を持つことが多く，この性質を **$\sqrt{n}$-一致性**などといったりする．実は多くの場合，この $\sqrt{n}$ が統計量の最適な (最も速い) 収束率であることが導かれる．そのため，統計学では $\sqrt{n}$-一致性を持つ統計量をつくることを目標にすることが多い．

中心極限定理は，2 項分布 $Bin(n,p)$ の確率関数が標本数 n の増加に伴い正規分布の密度関数で近似できるという，いわゆる**ド・モアブル=ラプラスの定理**[*6]に始まる．その後，ラプラスやチェビシェフらの仕事を経て定理 5.2.1 のような結果が得られることとなり，この結果が “分散を持つ IID 列ならばなんでもよい”という普遍性を持っていたため,「確率論における “中心的な (central)” 極限定理 (limit theorem)」として広く知られるようになった．これが “中心極限定理 (central limit theorem)” という名前の由来となったようだ[*7].

確率的(stochastic)

stochastic　確率的な、確率論的な、推計学的な