システム評価のための確率と統計

Probability and Statistics for System Evaluation

滝根 哲哉*

Tetsuya Takine

目 次

1　確率空間 **2**
1.1　標本空間と事象 . . . 2
1.2　確率測度 . . . 4
1.3　確率変数とその分布 . . . 8
1.4　確率ベクトルと結合分布 . . . 9
1.5　確率変数を引数とする関数と確率変数の畳み込み，混合 . . . 9

2　期待値 **13**
2.1　リーマン-スティルチェス積分 . . . 13
2.2　期待値の定義 . . . 14
2.3　積率，分散，共分散 . . . 16
2.4　母関数 . . . 18
2.5　条件付き期待値と最小二乗法 . . . 20
2.6　代表的な分布の例 . . . 23

3　確率変数と分布の収束 **24**
3.1　事象の列とボレル-カンテリの定理 . . . 24
3.2　確率変数列の概収束と確率収束 . . . 26
3.3　大数の法則 . . . 29
3.4　分布の弱収束と中心極限定理 . . . 30

4　確率過程の基礎 **32**
4.1　確率過程とは . . . 32
4.2　定常過程 . . . 33
4.3　計数過程とポワソン過程 . . . 36

5　離散時間マルコフ連鎖 **39**
5.1　離散時間マルコフ連鎖の遷移確率 . . . 39
5.2　再帰時間と状態の分類 . . . 40
5.3　既約な離散時間マルコフ連鎖の極限確率と定常状態分布 . . . 42
5.4　応用：$M/G/1$ 待ち行列の系内客数分布 . . . 44

6　確率的シミュレーションと統計理論 **49**
6.1　統計的推測と推定量 . . . 50
6.2　標本平均と標本分散 . . . 51
6.3　仮説検定と信頼区間 . . . 55
6.4　シミュレーションによる定常状態での特性評価 . . . 58
6.5　シミュレーションの加速技法：分散減少法 . . . 64

A　定理の証明 **69**
A.1　大数の弱法則の証明 . . . 69
A.2　中心極限定理の証明 . . . 70

*京都大学大学院情報学研究科数理工学専攻（〒 606-8501 京都市左京区吉田本町）
電話：(075)753-4758　　FAX：(075)753-4756　　電子メール：takine@amp.i.kyoto-u.ac.jp
URL：http://www-optima.amp.i.kyoto-u.ac.jp/~takine/

はじめに

授業の目的

この授業の目的は工学，数理科学，オペレーションズ・リサーチ等における問題へ確率モデル (stochastic model) を適用する際に必要となる知識と感覚を身につけることである．特に，システム評価のための確率と統計，すなわち，確率的なシミュレーション実験や確率的な挙動をするシステムの測定を通じてシステムを評価する際に必要となる確率と統計という位置付けで講義を行う．この資料は理論を応用する際に必要となる知識をまとめたものであるが，受講者の興味を勘案して，講義では理論的な内容を詳細に説明することは避け，その代わりに元となる考え方を (多少の不正確さを厭わず) 説明していくつもりである．理論的結果を正しく適用するためには，その元となる考え方を理解することが必須であることに留意されたい．

以下では確率モデルを扱う上で必要となる数学的知識を解説した後，実際の確率的なシミュレーションの出力や確率的な振舞いをするシステムの測定結果の解析手法を解説する．

1 確率空間

1.1 標本空間と事象

確率モデル (あるいは確率的実験) は以下の 3 つの要素から構成されていると考えることが出来る．

- 標本空間 (sample space): 起こり得る結果，すなわち標本 (sample) の集合であり，実験が行われるとこの中の一つだけが実際に起こる
- 事象 (event) の集合：それぞれの事象は標本空間の部分集合
- それぞれの事象へ確率を割り当てる規則：確率は 0 以上 1 以下の実数

注意 **1.1** 標本と事象の違いには特に注意を払う必要がある．標本とは実験の結果そのものであり，一つの実験に対して単一の標本が対応する．一方，事象は実験の結果を部分的に特徴付けるものであり，実験の結果が (その事象に対応する) 標本空間の部分集合に含まれていることを示す． □

実験の最も単純な例は標本の数が有限か，あるいは可算無限個 (個々の標本に対して $\omega_1, \omega_2,$ というふうに番号をつけることができる) の場合である．この場合は，それぞれの標本に対応する事象を考えることで，標本そのものに対して確率を割り当てることができる．また，このとき，任意の事象の確率は，その事象を構成する標本の確率を全て足し合わせたもので与えられる．さらに全ての標本に割り当てられた確率の総和は 1 である．

例 **1.1** コインを 2 回投げる実験を考える．ここで標本空間は四つの標本 {(表，表)(表，裏)(裏，表)(裏，裏)} から成るとし，それぞれの標本点に $1/4$ の確率を割り当てる．標本空間の部分集合はそれぞれ事象である．例えば「最初に表が出る」という事象は {(表，表)(表，裏)} であり，確率 $1/4+1/4=1/2$ をもつ． □

注意 **1.2** 標本空間を上記のように選ぶということは，コインが立つ，あるいはコインが 2 回目に投げられる前になくなってしまうといった現象を無視したことを意味する．また，確率を上記のように選ぶということはコインに歪みがあるという可能性を排除したことになる．このように，実際の物理現象と確率モデルの違いに注意しなければならない． □

例 1.1 をみると事象という概念を持ち出す必要はなく，個々の標本に確率を割り当てれば十分と思えるかも知れない．しかし，次の例では，個々の標本に確率を割り当てるということは意味がなく，なぜ，事象という概念を導入しなければならないかが明らかになる．

例 **1.2** $[0,1]$ の実数からなる標本空間を考え，実験から得られる結果は区間 $[0,1]$ 内で一様に分布しているとする[1]．もし，ある標本に正の確率 $p>0$ が割り当てられているならば，一様に分布しているという仮定より，他の標本も同じ確率 p を持つことになるが，一方，標本の数は無限にあるので，確率の総和は $p\times\infty=\infty$ となってしまい矛盾する．よって各標本に割り当てられる確率は 0 でなければならない．しかし，確率 0 をもつ標本を無限個足し合わせて，個々の事象に対する確率 (例えば標本が区間 $[0,1/2]$ 内に入る確率) を得る合理的な方法はない． □

例 1.2 の場合，任意の区間に対して，区間長に等しい確率を割り当て，互いに交わらない区間の和に対応する確率は，それぞれの区間の確率の和で表すことが合理的であると思われる．例えば標本が区間 $[0,1/2]$ 内に入る確率は $1/2$ であり，区間 $[0,1/2]$ あるいは区間 $[3/4,1]$ のいずれかに含まれる確率は $3/4$ である．

この例から分かるように，確率は，標本ではなく事象に対して割り当てなければならない．さらに，割り当てられた確率は非負の値であり，標本空間全体に対しては 1 であり，かつ，任意の互いに交わらない事象の列 $E_1,E_2,\ldots$ に対してこれらの事象の和集合の確率はそれぞれの事象の確率の和で与えられるようなものでなくては我々が通常の生活で使っている「確率」という概念と一致しない．

標本空間が可算の場合，上の議論は事象に対する確率を事象に含まれる個々の標本に対する確率の和として定義できることを示している．しかし，より一般的な例 1.2 のような場合においては，全ての事象に対して確率が決定できるようにするためには，どのような事象に確率を割り当てておけば十分かという問題に直面することになる．これを解決するため，事象は以下のようにして定義される．

定義 **1.1** (σ-集合体と事象) 標本空間 Ω の部分集合の集まり $\mathcal{F}$ が，条件

1. $\Omega\in\mathcal{F}$ (標本空間 Ω は事象である)
2. $A\in\mathcal{F}\Rightarrow A^c=\Omega-A\in\mathcal{F}$ (A が事象ならば A の余事象 A^c も事象である)
3. $A_i\in\mathcal{F}\ (i=1,2,\ldots)\Rightarrow\cup_{i=1}^{\infty}A_i\in\mathcal{F}$ (事象の無限列に対してそれらの和事象も事象となる[2])

を満たすとき，$\mathcal{F}$ を Ω 上の **σ-集合体** (σ-field) といい，$\mathcal{F}$ の要素を事象と呼ぶ． □

直観的には，標本空間 Ω を分割して得られる集合を集めたものが σ-集合体 $\mathcal{F}$ である．定義より，以下のことが成り立つ．

補題 **1.1** (事象の性質)

1. $\phi\in\mathcal{F}$ (空集合 ϕ も事象である)
2. $A,B\in\mathcal{F}\Rightarrow A\cup B\in\mathcal{F},\ A\cap B\in\mathcal{F},\ A-B\in\mathcal{F}$ (事象 A,B の和事象 $A\cup B$，積事象 $A\cap B$，差事象 $A-B$ も事象である)
3. $A_i\in\mathcal{F}\ (i=1,2,\ldots)\Rightarrow\bigcap_{i=1}^{\infty}A_i\in\mathcal{F}$ (事象の無限列に対してそれらの積事象も事象となる[3])

□

1. $\phi=\Omega^c\in\mathcal{F}$. 2. $A_1=A,A_2=B,A_i=\phi\ (i\geq 3)$ で 3. を適用．$A\cap B=(A^c\cup B^c)^c$．$A-B=A\cap B^c$．3. $\cap_{i=1}^{\infty}A_i=(\cup_{i=1}^{\infty}A_i^c)^c$．

事象 $A_i\ (i=1,\ldots,n)$ が全ての $i,j\ (i\neq j)$ に対して $A_i\cap A_j=\phi$ であるとき，事象 $A_i\ (i=1,\ldots,n)$ は排反 (exclusive) と言われる．標本空間 Ω とその上の σ-集合体 $\mathcal{F}$ の組を $(\Omega,\mathcal{F})$ と表し，可測空間 (measurable space) とよぶ．

[1] 一様に分布しているとは取りうるどの値も同様に確からしいということである．
[2] $\cup_{i=1}^{\infty}A_i=\{\omega\mid$ ある自然数 i が存在して $\omega\in A_i\}$．
[3] $\cap_{i=1}^{\infty}A_i=\{\omega\mid$ 全ての自然数 i に対して $\omega\in A_i\}$．

標本空間 Ω 上の σ-集合体は一意ではない．例えば，$\mathcal{F}_1 = \{\phi, \Omega\}$ や，ある $A \in \Omega$ に対して $\mathcal{F}_2 = \{\phi, A, A^c, \Omega\}$，あるいは $\mathcal{F}_3 = \{\Omega$ の全ての部分集合 $\}$ は全て標本空間 Ω 上の σ-集合体である．以下では

$$2^\Omega = \{\Omega \text{ の全ての部分集合 } \}$$

という記号を用いる．全ての部分集合には空集合 ϕ と全体集合 Ω も含まれていることに注意する．

標本空間の中で応用上特に重要なものに，有限あるいは可算無限個の標本からなる Ω と，実数全体 ($= \mathcal{R}$) からなる Ω がある．有限あるいは可算無限個の標本からなる Ω に対しては，通常，$\mathcal{F} = 2^\Omega$ とする．一方，$\mathcal{R}$ からなる Ω に対しては

$$\mathcal{F} = \mathcal{B}(\mathcal{R}) = \text{全ての半開区間 } (a, b] \text{ を含む最小の } \sigma\text{-集合体}$$

とする．この $\mathcal{B}(\mathcal{R})$ はボレル集合体 (Borel field) と呼ばれており，このように事象の集合を選ぶことで，確率を矛盾なく定義することができる．ボレル集合体 $\mathcal{B}(\mathcal{R})$ には 1 点集合 $\{a\}$, 開区間 (a, b), 任意の閉区間 $[a, b]$ や (a, ∞), $(-\infty, b)$, $[a, \infty)$, $(-\infty, b]$ が全て要素として含まれる．

$$\{a\} = \cap_{n=1}^{\infty}(a - 1/n, a] \in \mathcal{B}(\mathcal{R}),\ (a, b) = \cup_{n=1}^{\infty}(a, b - 1/n] \in \mathcal{B}(\mathcal{R}),\ [a, b] = \cap_{n=1}^{\infty}(a - 1/n, b] \in \mathcal{B}(\mathcal{R}) \text{ etc}$$

1.2 確率測度

次に可測空間 $(\Omega, \mathcal{F})$ が与えられたとき，各事象に対して確率測度を定義する．以前に述べたように確率は以下の三つの公理を満足する必要がある．

定義 1.2 (確率の公理) 以下の性質をもつ $\mathcal{F}$ 上で定義された実数値関数 P は確率測度 (probability measure) または確率分布 (probability distribution) と呼ばれる[4]．

1. 全ての $A \in \mathcal{F}$ に対して $P(A) \geq 0$ (確率は非負の値を取る)
2. $P(\Omega) = 1$ (標本空間全体に対する確率は 1)
3. $A_i \in \mathcal{F}$ $(i = 1, 2, \ldots)$ かつ $A_i \cap A_j = \phi$ $(i \neq j)$ ならば

$$P\left(\bigcup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$$

が成立する (排反な事象の無限列の和事象の確率はそれぞれの事象の確率の和に等しい：σ-加法性という)

□

$P(A)$ は事象 A の確率 (probability) と呼ばれる．特に $P(A) = 1$ のとき，事象 A は確率 1 で成り立つ (with probability 1, w.p.1) という．$P(A) = 1$ は $A = \Omega$ を意味しないことに注意する．

例 1.3 例 1.2 において A を標本が開区間 (0,1) 内にある事象とすれば $P(A) = 1$ であるし，B を標本が閉区間 [0,1] から高々可算個の点 $x_1, x_2, \ldots$ を取り除いた残りの区間に含まれる確率とすれば $P(B) = 1$ である． □

言い換えると，ある事象 C に対して $P(C) = 0$ であるからといって，その事象に含まれる標本 $\omega \in C$ が決して起こり得ないかというと，必ずしもそうではない．たとえば，例 1.2 において，実験を一回行えば，必ずある標本 $\omega \in [0, 1]$ が得られるが，$P(\{\omega\}) = 0$ である．このように標本空間が連続な場合，我々が生活で用いている「確率」という概念はそのまま適用することが出来ない．

Ω, $\mathcal{F}$, P の三つ組 $(\Omega, \mathcal{F}, P)$ を確率空間 (probability space) という．確率空間は，標本空間が有限または可算な場合 (離散型確率空間) とそうでない場合に大別することができる．離散型確率空間の場合は，$\Omega = \{\omega_1, \omega_2, \ldots\}$，

[4] P は Pr と書かれることもある．

$\mathcal{F} = 2^\Omega$ に対して，p_i $(i = 1, 2, \ldots)$ を総和が 1 である非負の実数としたとき，

$$P(A) = \sum_{\omega_i \in A} p_i$$

とすれば，$(\Omega, \mathcal{F}, P)$ は確率空間となる．このような p_i は確率関数 (probability function, または probability mass function) と呼ばれる．

一方，標本空間 Ω が実数全体の場合 ($\Omega = \mathcal{R}$)，明らかに，各標本と確率を 1 対 1 に対応させる方法では確率空間を作り出すことは出来ない．このような場合，確率空間は実数値関数を用いて構築される．実数値関数 $F(x)$ $(x \in \mathcal{R})$ が以下の性質をもつと仮定する．

定義 1.3 (分布関数の性質)

1. $\lim_{x \to -\infty} F(x) = 0$ [5]
2. $F(x)$ は非減少 ($x < y \Rightarrow F(x) \leq F(y)$) かつ右連続 ($\lim_{\epsilon \to 0+} F(x + \epsilon) = F(x)$ [6])
3. $\lim_{x \to \infty} F(x) = 1$ [7]

このような性質をもつ関数 $F(x)$ は分布関数 (distribution function) あるいは単に分布 (distribution) と呼ばれる． □

ここで，$\Omega = \mathcal{R}$，$\mathcal{F} = \mathcal{B}(\mathcal{R})$ とする．ある分布関数 $F(x)$ が与えられたとき，

$$F(x) = P((-\infty, x]) \tag{1}$$

を満たすような確率測度 P が一意に定まることが知られている．よって，標本空間が実数全体の場合，確率空間を定めることと分布関数を定めることは等価であり，式 (1) を満たすように確率 P を定めると，$(\Omega, \mathcal{F}, P) = (\mathcal{R}, \mathcal{B}(\mathcal{R}), P)$ は確率空間となる．

ある分布関数 $F(x)$ が与えられたとき

$$F(x) = \int_{-\infty}^{x} f(y)dy \tag{2}$$

が全ての実数 x に対して成り立つような関数 $f(x)$ が存在すれば，$f(x)$ は $F(x)$ の密度関数 (density function) と呼ばれる．定義より，分布関数 $F(x)$ が微分可能ならば密度関数 $f(x)$ が存在し，

$$f(x) = \frac{d}{dx}F(x) = \lim_{\epsilon \to 0} \frac{F(x + \epsilon) - F(x)}{\epsilon}$$

が成立する．式 (1) より，

$$F(x + \epsilon) - F(x) = P((-\infty, x + \epsilon]) - P((-\infty, x]) = P((x, x + \epsilon])$$

であるので，このとき $f(x)$ は

$$f(x)\epsilon \approx P((x, x + \epsilon])$$

と解釈できる．また $F(\infty) = 1$ なので，非負の値を取る関数 $f(x)$ が

$$\int_{-\infty}^{\infty} f(x)dx = 1$$

を満たすならば，式 (2) で定義された関数 $F(x)$ は $f(x)$ を密度関数にもつ分布関数となる．

標本空間が連続な場合と同様に，標本空間 Ω が $\Omega \subset \mathcal{R}$ であるような離散型確率空間 $(\Omega, 2^\Omega, P)$ に対しても分布関数は定義できる．すなわち，可算個の $x_1, x_2, \ldots \in \Omega$, に対して確率関数 $p_i = P(x_i)$ が与えられたとき，

$$F(x) = \sum_{x_i \leq x} p_i$$

[5] 任意の正数 ϵ に対して正数 δ が存在し，$x < -\delta$ ならば $|f(x) - a| < \epsilon$ となるとき $\lim_{x \to -\infty} f(x) = a$ と書く．

[6] $\lim_{\epsilon \to a+}$ は ϵ を正の側から a へ近付けることを意味する．よって，$\lim_{\epsilon \to 0+} F(x + \epsilon)$ は，任意の正数 ϵ に対して正数 δ が存在し，$y \in [x, x + \delta)$ ならば $|f(y) - f(x)| < \epsilon$ となるということである．

[7] 任意の正数 ϵ に対して正数 δ が存在し，$x > \delta$ ならば $|f(x) - a| < \epsilon$ となるとき $\lim_{x \to \infty} f(x) = a$ と書く．

とすれば，$F(x)$ は分布関数となる．

上記では標本空間が 1 次元の分布関数を考えたが，これを多次元に拡張することが出来る．標本空間が n 次元実ベクトル空間 $\mathcal{R}^n$ であるとする．このとき，$\mathcal{R}^n$ 上の全ての長方形領域 $\cap_{i=1}^n (a_i, b_i]$ を含む最小の σ-集合体 $\mathcal{B}(\mathcal{R}^n)$ を $\mathcal{R}^n$ 上のボレル集合体と呼ぶ．1 次元の場合と同様に以下のような性質をもつ関数 $F(x_1, x_2, \ldots, x_n)$ を考える．

1. $\lim_{x_1 \to -\infty} \lim_{x_2 \to -\infty} \cdots \lim_{x_n \to -\infty} F(x_1, x_2, \ldots, x_n) = 0$
2. $F(x_1, x_2, \ldots, x_n)$ は各引数 x_i に対して非増加かつ右連続
3. $\lim_{x_1 \to \infty} \lim_{x_2 \to \infty} \cdots \lim_{x_n \to \infty} F(x_1, x_2, \ldots, x_n) = 1$

このような性質をもつ関数 $F(x_1, x_2, \ldots, x_n)$ は n 次元分布関数と呼ばれる．n 次元分布関数 $F(x_1, x_2, \ldots, x_n)$ が与えられると，可測空間 $(\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n))$ 上において

$$F(x_1, x_2, \ldots, x_n) = P\left(\cup_{i=1}^n (-\infty, x_i]\right), \qquad -\infty < x_i < \infty$$

によって確率測度 P を決定することができる．n 次元分布関数によって決定される確率測度は n 次元確率分布，または，n 次元分布と呼ばれる．

また，

$$F(x_1, x_2, \ldots, x_n) = \int_{-\infty}^{x_1} \int_{-\infty}^{x_2} \cdots \int_{-\infty}^{x_n} f(y_1, y_2, \ldots, y_n) dy_1 dy_2 \cdots dy_n$$

を満たす関数 $f(x_1, x_2, \ldots, x_n)$ を $F(x_1, x_2, \ldots, x_n)$ の密度関数といい，

$$F(x_1, x_2, \ldots, x_n) = \sum_{y_{1,i_1} \leq x_1} \sum_{y_{2,i_2} \leq x_2} \cdots \sum_{y_{n,i_n} \leq x_n} p(y_{1,i_1}, y_{2,i_2}, \ldots, y_{n,i_n})$$

を満たす関数 $p(y_{1,i_1}, y_{2,i_2}, \ldots, y_{n,i_n})$ を $F(x_1, \ldots, x_n)$ の確率関数という．

さて，確率の定義より以下が成立する．

補題 **1.2** (確率の性質) 事象 $A, B \in \mathcal{F}$ に対して

1. $P(\phi) = 0$
2. $A \cap B = \phi$ ならば $P(A \cup B) = P(A) + P(B)$
3. $P(A^c) = 1 - P(A)$
4. $A \subset B$ ならば $P(A) \leq P(B)$

□

性質 1. は $A_i = \Omega$, $A_i = \phi$ $(i = 2, 3, \ldots)$ とすれば公理の 3. より $P(\Omega) = P(\Omega) + \sum_{i=2}^{\infty} P(\phi)$, すなわち $\sum_{i=2}^{\infty} P(\phi) = 0$. 公理の 1. より $P(\phi) \geq 0$ なので $P(\phi) = 0$. 性質 2. は $A_1 = A$, $A_2 = B$, $A_i = \phi$ $(i = 3, 4, \ldots)$ として公理の 3. と性質 1. を用いる．性質 3. は $A_1 = A$, $A_2 = A^c$, $A_i = \phi$ $(i = 3, 4, \ldots)$ として公理の 2., 3. と性質 1. を用いる. 性質 4. は $B = A \cup (B - A)$ かつ $A \cap (B - A) = \phi$ に留意して性質 2. を適用し，公理 1. を用いる．

次に条件付き確率を定義する．

定義 **1.4** (条件付き確率) 事象 $A, B \in \mathcal{F}$ に対して，$P(B) > 0$ のとき，

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

を事象 B が起こったという条件の下での事象 A の条件付き確率 (conditional probability) という．□

定義より，積事象に対する確率 $P(A \cap B)$ は条件付き確率 $P(A \mid B)$ を用いて

$$P(A \cap B) = P(A \mid B) P(B)$$

と表すことができる．そこで事象 B_i $(i = 1, \ldots, n)$ が排反でかつ $\Omega = \cup_{i=1}^{n} B_i$ であるとする．このとき，任意の事象 A に対して

$$P(A) = P(A \cap \Omega) = P\left(A \bigcap \left(\bigcup_{i=1}^{n} B_i\right)\right) = P\left(\bigcup_{i=1}^{n} (A \cap B_i)\right) = \sum_{i=1}^{n} P(A \cap B_i)$$

となるので

$$P(A) = \sum_{i=1}^{n} P(A \mid B_i) P(B_i)$$

を得る．これは全確率の法則 (law of total probabilities) と呼ばれる．さらに，$P(B_j \mid A) = P(B_j \cap A)/P(A)$ であるので

$$P(B_j \mid A) = \frac{P(A \mid B_j) P(B_j)}{\sum_{i=1}^{n} P(A \mid B_i) P(B_i)}$$

が成立する．これはベイズの公式 (Bayes' formula) と呼ばれる．

例 **1.4** (ポリヤの壷) r 個の赤玉と b 個の黒玉の入った壷がある．ここで，この壷から玉を一つ取りだし，取り出した玉と同じ色の玉を c 個加えて壷へ戻すという操作を繰り返す． □

ポリヤの壷において，k 回目に赤玉を取り出す事象を R_k，黒玉を取り出す事象を B_k とする．定義より $P(R_1) = r/(r+b)$ である．また，

$$P(R_1 \cap R_2) = P(R_2 \mid R_1) P(R_1) = \frac{r+c}{r+b+c} \cdot \frac{r}{r+b},$$
$$P(B_1 \cap R_2) = P(R_2 \mid B_1) P(B_1) = \frac{r}{r+b+c} \cdot \frac{b}{r+b}$$

なので全確率の公式より

$$P(R_2) = P(R_2 \mid R_1) P(R_1) + P(R_2 \mid B_1) P(B_1) = \frac{(r+c)r + br}{(r+b)(r+b+c)} = \frac{r}{r+b}$$

となり，$P(R_2) = P(R_1)$ である．

定義 **1.5** (独立) 事象 $A, B \in \mathcal{F}$ に対して，$P(A \cap B) = P(A)P(B)$ が成り立つとき，事象 A と事象 B は独立 (independent) であるといわれる．また，一般に n 個の事象 A_i $(i = 1, \ldots, n)$ が与えられたとき，

$$P\left(\bigcap_{i=1}^{n} A_i\right) = \prod_{i=1}^{n} P(A_i)$$

が成り立つならば，事象 A_i $(i = 1, \ldots, n)$ は互いに独立 (mutually independent) であるといわれる． □

注意 **1.3** もし、事象 A_i $(i = 1, \ldots, n)$ が互いに独立ならば，任意の m $(m = 2, \ldots, n)$ と任意の $i_j \in \{1, \ldots, n\}$ に対して

$$P\left(\bigcap_{j=1}^{m} A_{i_j}\right) = \prod_{j=1}^{m} P(A_{i_j})$$

が成立する。しかし、$m = n$ を除く全ての m $(m = 2, \ldots, n-1)$ と任意の $i_j \in \{1, \ldots, n\}$ に対して上式が成立したとしても，事象 A_i $(i = 1, \ldots, n)$ は互いに独立とは限らない。 □

例 **1.5** 例 1.1 において，A_1 を 1 回目が表，A_2 を 2 回目が裏，A_3 を 1 回目，2 回目ともに表あるいは裏という事象とする．明らかに

$$P(A_1) = P(A_2) = P(A_3) = \frac{1}{2}, \qquad P(A_1 \cap A_2) = P(A_1 \cap A_3) = P(A_2 \cap A_3) = \frac{1}{4}$$

であるので，A_1 と A_2，A_2 と A_3，A_1 と A_3 はそれぞれ独立である．しかし，

$$P(A_1 \cap A_2 \cap A_3) = 0$$

であるので，A_1，A_2，A_3 は互いに独立でない． □

今 $0 < P(B) < 1$ であるとする．もし事象 A と事象 B が独立ならば $P(A \mid B) = P(A \cap B)/P(B) = P(A)$ となる．さらに $P(A \cap B^c) = P(A) - P(A \cap B) = P(A)(1 - P(B)) = P(A)P(B^c)$ となり，事象 A と事象 B^c も独立である．すなわち，$P(A \mid B^c) = P(A)$ となり，事象 A の起こる確率は事象 B に影響されない．さらに $P(A^c \mid B^c) = 1 - P(A \mid B^c) = 1 - P(A) = P(A^c)$ となるので，$P(A^c \cap B^c) = P(A^c \mid B^c)P(B^c) = P(A^c)P(B^c)$ となり，事象 A と事象 B が独立ならば，余事象 A^c と余事象 B^c も独立となる．

1.3 確率変数とその分布

最初に確率変数を定義する．

定義 1.6 (確率変数) 確率空間 $(\Omega, \mathcal{F}, P)$ において，Ω から実数の集合 $\mathcal{R}$ への関数 $X(\omega)$ が任意の実数 x に対して

$$\{\omega \in \Omega \mid X(\omega) \leq x\} \in \mathcal{F}$$

であるならば，$X = X(\omega)$ を確率変数 (random variable) という． □

すなわち，確率変数とは，各標本に対して実数が対応しており，さらに，その実数が取り得る範囲を限定したとき，対応関係を用いて事象を特定することが可能であり，かつ，そのような事象に対する確率が定まるものである．

確率変数の定義において注意すべき点は，各標本に対して実数が定義されていなければならないことである．

例 1.6 例 1.1 において，N を最初に表が出るまでにコインを投げた回数とする．このとき (表，表) なら $N = 1$,(表，裏) なら $N = 1$, (裏，表) なら $N = 2$ であるが (裏，裏) の場合は値が定義されていない．よって N は確率変数ではない[8]． □

特に，事象 A に対して

$$X(\omega) = \begin{cases} 1, & \omega \in A \\ 0, & \text{その他} \end{cases}$$

とすると，$X = X(\omega)$ は確率変数である．このような X は事象 A の定義関数あるいは指示関数 (indicator function) と呼ばれており，$1_{\{A\}}$ と表す．

以下では，誤解を生じない限り確率空間を特に指定せず，単に「X は確率変数である」と書くことにする．また $\{\omega \in \Omega \mid X(\omega) \leq x\}$ も単に $\{X \leq x\}$ と書き，さらに $P(\{X \leq x\})$ は単に $P(X \leq x)$ と書くことにする．

さて，確率変数 X に対して $F(x) = P(X \leq x)$ とおけば，$F(x)$ は分布関数となる．このような $F(x)$ を確率変数 X の分布関数または分布という．さらに，確率変数 X は分布関数 (あるいは分布) $F(x)$ をもつ，または，確率変数 X は分布関数 (あるいは分布) $F(x)$ に従うという．逆に，ボレル集合 $\mathcal{B}(\mathcal{R})$ に含まれる任意の集合 B に対して確率 $P(X \in B)$ は X の分布 $F(x)$ によって決定される．

$$P(X \in B) = \int_B dF(x)$$

ここで，右辺の積分はリーマン-スティルチェス積分 (Riemann-Stieltjes integral) と呼ばれるものである[9]．これについては後に 2.1 節で簡単に解説することにする．特に $F(x)$ の密度関数 $f(x)$ が存在する場合は

$$P(X \in B) = \int_B f(x)dx$$

[8] このような場合 (裏，裏) に対して ∞ を割り当て，確率変数 N の取り得る値に無限大を含めて考えることが多い．このように本来の確率変数の定義を拡張した結果，$P(-\infty < X < \infty) < 1$ となるような確率変数を不完全な (defective) 確率変数という．

[9] 数学的な完璧さを求めるならば，この積分はルベーグ-スティルチェス積分 (Lebesgue-Stieltjes integral) として解釈する必要があるが，この授業で扱う応用においてはリーマン-スティルチェス積分に限って議論しても支障はない．

となる．また，$F(x)$ が離散的で確率関数 $p(x_i)$ を持つ場合は

$$P(X \in B) = \sum_{x_i \in B} p(x_i)$$

となる．

1.4 確率ベクトルと結合分布

n 個の確率変数 $X_1, X_2, \ldots, X_n$ に対して，$\boldsymbol{X} = (X_1, X_2, \ldots, X_n)$ を確率ベクトルという．また，$\boldsymbol{B} \in \mathcal{B}(\mathcal{R}^n)$ に対して，$\boldsymbol{X} \in \boldsymbol{B}$ は事象となり，確率空間 $(\mathcal{R}^n, \mathcal{B}(\mathcal{R}^n), P)$ を作ることが出来る．$P(\boldsymbol{X} \in \boldsymbol{B})$ は $\boldsymbol{X}$ の分布，あるいは 確率変数 $X_1, X_2, \ldots, X_n$ の結合分布 (joint distribution) と呼ばれる．確率変数 $X_1, X_2, \ldots, X_n$ の結合分布は n 次元分布関数 $F(x_1, x_2, \ldots, x_n)$ を用いて一意に定めることが出来る．

$$F(x_1, x_2, \ldots, x_n) = P(X_1 \leq x_1, X_2 \leq x_2, \ldots X_n \leq x_n)$$

このとき，$F(x_1, x_2, \ldots, x_n)$ は $X_1, X_2, \ldots, X_n$ の結合分布関数と呼ばれる．

n 次元分布関数に対して $\mathcal{R}$ 上の関数 $F_1(x)$ を

$$F_1(x) = F(x, \infty, \infty, \ldots, \infty)$$

とすると，$F_1(x)$ は (1 次元の) 分布関数である．このように多次元の分布関数 F に対して幾つかの引数を無限大にして得られる分布関数で決定される分布を F の周辺分布 (marginal distribution) という．

定義 1.7 (確率変数の独立性) 確率変数 $X_1, X_2, \ldots, X_n$ が全ての $B_i \in \mathcal{B}(\mathcal{R})$ $(i = 1, \ldots, n)$ に対して

$$P(X_1 \in B_1, X_2 \in B_2, \ldots, X_n \in B_n) = P(X_1 \in B_1)P(X_2 \in B_2) \cdots P(X_n \in B_n)$$

を満たすならば，確率変数 $X_1, X_2, \ldots, X_n$ は互いに独立であると呼ばれる． □

確率変数 $X_1, X_2, \ldots, X_n$ がそれぞれ周辺分布 $F_i(x)$ $(i = 1, \ldots, n)$ をもっているとき，これらが互いに独立であるということは結合分布が周辺分布の積に等しいことと等価である．

$$F(x_1, x_2, \ldots, x_n) = F_1(x_1)F_2(x_2) \cdots F_n(x_n)$$

また，このとき，結合密度関数と周辺密度関数，あるいは結合確率関数と周辺確率関数に対しても同様の結果を得ることが出来る．

$$f(x_1, \ldots, x_n) = f_1(x_1)f_2(x_2) \cdots f_n(x_n), \qquad p(x_1, \ldots, x_n) = p_1(x_1)p_2(x_2) \cdots p_n(x_n)$$

1.5 確率変数を引数とする関数と確率変数の畳み込み，混合

分布関数 $F(x)$ をもつ確率変数 X と，実数から実数への関数 $h(x)$ が与えられたとき，$Y = h(X)$ が確率変数となるための条件を考える．

定義 1.8 (可測関数) 可測空間 $(\mathcal{R}, \mathcal{B}(\mathcal{R}))$ において $\mathcal{R} \to \mathcal{R}$ への関数 $h(x)$ が任意の $y \in \mathcal{R}$ に対して $\{x \mid h(x) \leq y\} \in \mathcal{B}(\mathcal{R})$ を満たすとき，$h(x)$ は $(\mathcal{R}, \mathcal{B}(\mathcal{R}))$ 上の可測関数 (measurable function) と呼ばれる．特に連続な関数 $h(x)$ は可測関数である[10] ．また，高々可算個の点を除いて連続な関数 $h(x)$ も可測関数である． □

[10] 定義域 D をもつ関数 $h(x)$ が点 $x_0 \in D$ で連続であるとは，任意の正数 ϵ に対して，ある正数 δ が存在し，$|x - x_0| \leq \delta$ かつ $x \in D$ ならば $|h(x) - h(x_0)| \leq \epsilon$ となるということである．関数 $h(x)$ の定義域内の全ての点で連続であるとき，$h(x)$ は D において連続であるという．

可測空間 $(\mathcal{R},\, \mathcal{B}(\mathcal{R}))$ 上の確率変数 X と $\mathcal{R} \to \mathcal{R}$ への可測関数 $h(x)$ が与えられたとき，$Y = h(X)$ は確率変数となる．このとき，確率変数 Y の分布関数 $G(x)$ は

$$G(x) = P(h(X) \le x) = \int_{h(u)\le x} dF(u)$$

で与えられる．もし $F(x)$ が密度関数 $f(x)$ あるいは確率関数 $p(x)$ をもつならば

$$G(x) = \int_{h(u)\le x} f(u)du, \qquad G(x) = \sum_{h(u)\le x} p(u)$$

である．例えば $h(x) = x^2$ とする．このとき $Y = X^2$ であり，分布関数 $G(x)$ は $x \ge 0$ に対しては

$$G(x) = P(X^2 \le x) = P(-\sqrt{x} \le X \le \sqrt{x}) = P(X \le \sqrt{x}) - P(X < -\sqrt{x})$$

であるので

$$G(x) = F(\sqrt{x}) - \lim_{u\to x-} F(-\sqrt{u}), \qquad x > 0$$

となり，$x < 0$ に対しては $G(x) = 0$ である．

この応用として，区間 $(0,1)$ の間の値を取る一様乱数を用いて分布関数 $G(x)$ に従う確率変数 X の標本を生成する方法を考える．まず，分布関数 $G(x)$ の逆写像 $G^{-1}(y)$ を次式で定義する．

$$G^{-1}(y) = \min(z; G(z) \ge y), \qquad 0 < y < 1$$

分布関数 $G(x)$ が連続関数の場合，$G^{-1}(y)$ は通常の逆関数に等しいことに注意する．図 1 は $G^{-1}(y_i) = x_i$ $(i = 1, 2)$ の対応関係を示している．分布関数は不連続点を含むことがあるため，$Y = y_1$ のように $y = G(x)$ となる x の値が定まらない場合は $G(z) \ge y$ である最小の z の値を求め $G^{-1}(y) = z$ とする．

次に，区間 $(0,1)$ の間の値を取る一様乱数 Y と逆写像 $G^{-1}(y)$ によって一意に定まる確率変数 $X = G^{-1}(Y)$ を考える．X の従う分布は

$$\begin{aligned} P(X \le x) &= P(\min(z; G(z) \ge Y) \le x) \\ &= 1 - P(G(z) \ge Y \text{ なる全ての } z \text{ に対して } z > x) \end{aligned}$$

である．ここで $G(z)$ は非減少関数であることに注意すると，事象「$G(z) \ge Y$ なる全ての z に対して $z > x$」は事象「$Y > G(x)$」と等価であることが分かる[11]．よって

$$P(X \le x) = 1 - P(Y > G(x)) = P(Y \le G(x))$$

を得る．さらに Y が区間 $(0,1)$ の間の値を取る一様乱数であるので $P(Y \le G(x)) = G(x)$ となり，X の分布関数は $G(x)$ で与えられる．

定理 1.1 (逆変換法) 区間 $(0,1)$ の間の値を取る一様乱数 Y に対して $X = \min(z; Y \le G(z))$ とすれば，X は分布関数 $G(x)$ をもつ確率変数となる．一様乱数からこのようにして他の確率変数の標本を生み出す方法を逆変換法 (inverse transform method) という． □

次に確率変数の和が従う分布を考える．今，確率変数 X_1, X_2 がそれぞれ分布関数 $F_1(x)$, $F_2(x)$ をもっており，互いに独立であるとする．このとき，新しい確率変数 $Y = X_1 + X_2$ の分布関数 $G(x) = P(Y \le x)$ は

$$\begin{aligned} G(x) = P(X_1 + X_2 \le x) &= \int\int_{u_1+u_2\le x} dF_1(u_1)dF_2(u_2) \\ &= \int_{u_2=-\infty}^{\infty} \int_{u_1=-\infty}^{x-u_2} dF_1(u_1)dF_2(u_2) = \int_{-\infty}^{\infty} F_1(x-u_2)dF_2(u_2) \end{aligned}$$

[11] $G(z) \ge Y$ なる全ての z に対して $z > x$ であるとき，もし $G(x) \ge Y$ ならば $x > x$ となり矛盾する．よって $G(x) < Y$ である．逆に $Y > G(x)$ ならば，$G(z) \ge Y$ なる全ての z に対して $G(z) \ge Y > G(x)$ が成立し，$G(z)$ が非減少関数であるため $z > x$ を得る．

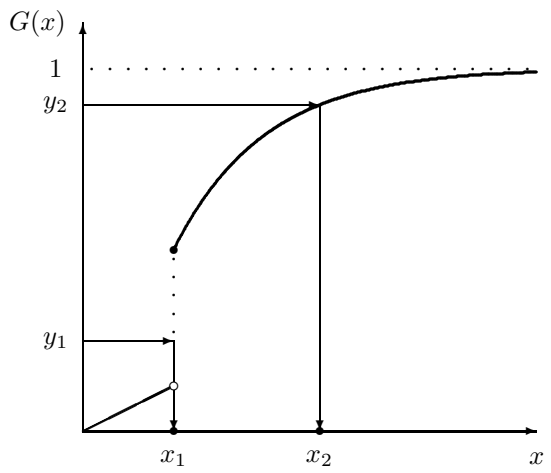

図 1: 逆変換法

となる．左辺の最後の積分は分布関数 $F_1(x)$ と $F_2(x)$ の畳み込み (convolution) と呼ばれ，$F_1 * F_2(x)$ と書かれることが多い．

$$F_1 * F_2(x) = \int_{-\infty}^{\infty} F_1(x-u)dF_2(u)$$

もし，上記において X_1 と X_2 の役割を交換して計算すると

$$G(x) = P(X_1 + X_2 \leq x) = \int\int_{u_1+u_2 \leq x} dF_2(u_2)dF_1(u_1)$$
$$= \int_{u_1=-\infty}^{\infty} \int_{u_2=-\infty}^{x-u_1} dF_2(u_2)dF_1(u_1) = \int_{-\infty}^{\infty} F_2(x-u_1)dF_1(u_1)$$

となる．すなわち，畳み込みは計算の順序に無関係である．

$$G(x) = F_1 * F_2(x) = F_2 * F_1(x)$$

特に独立な非負の確率変数 X_1 と X_2 に対しては

$$G(x) = \int_0^x F_1(x-u_2)dF_2(u_2) = \int_0^x F_2(x-u_1)dF_1(u_1)$$

となる。

もし，分布関数 $F_1(x)$, $F_2(x)$ が密度関数 $f_1(x)$, $f_2(x)$ をもつならば，

$$\frac{d}{dx}G(x) = \frac{d}{dx}\int_{-\infty}^{\infty} F_1(x-u)f_2(u)du = \int_{-\infty}^{\infty} f_1(x-u)f_2(u)du$$

となるので，分布関数 $G(x)$ の密度関数 $g(x)$ は

$$g(x) = \int_{-\infty}^{\infty} f_1(x-u)f_2(u)du$$

で与えられる．左辺の積分は密度関数 $f_1(x)$ と $f_2(x)$ の畳み込みと呼ばれ，$f_1 * f_2(x)$ と書かれる．上式で $v = x-u$ と変数変換すれば

$$g(x) = \int_{-\infty}^{\infty} f_1(v)f_2(x-v)dv = f_2 * f_1(x)$$

となるので，密度関数の畳み込みに関しても計算の順序とは無関係である．特に独立な非負の確率変数 X_1 と X_2 に対しては

$$g(x) = \int_0^x f_1(x-u_2)f_2(u_2)du_2 = \int_0^x f_1(u_1)f_2(x-u_1)du_1$$

となる。

例 **1.7** (正規分布の再現性) パラメタ m, σ (ただし $\sigma > 0$) をもつ密度関数

$$f(x) = \frac{1}{\sqrt{2\pi}\sigma} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

に従う分布を正規分布 (normal distribution) といい，$N(m, \sigma^2)$ と書く．畳み込みの例として，独立な正規分布 $N(m_1, \sigma_1^2)$，$N(m_2, \sigma_2^2)$ の密度関数の畳み込み $g(x)$ を考える．定義より

$$\begin{aligned}
g(x) &= \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}\sigma_1} \exp\left(-\frac{1}{2}\frac{(x-u-m_1)^2}{\sigma_1^2}\right) \frac{1}{\sqrt{2\pi}\sigma_2} \exp\left(-\frac{1}{2}\frac{(u-m_2)^2}{\sigma_2^2}\right) du \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \int_{-\infty}^{\infty} \exp\left(-\frac{1}{2\sigma_1^2\sigma_2^2}\left[(\sigma_1^2+\sigma_2^2)\left(u - \frac{(x-m_1)\sigma_2^2 + m_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2 + \frac{\sigma_1^2\sigma_2^2}{\sigma_1^2+\sigma_2^2}(x-m_1-m_2)^2\right]\right) \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \int_{-\infty}^{\infty} \exp\left(-\frac{\sigma_1^2+\sigma_2^2}{2\sigma_1^2\sigma_2^2}\left(u - \frac{(x-m_1)\sigma_2^2 + m_2\sigma_1^2}{\sigma_1^2+\sigma_2^2}\right)^2\right) du \\
&= \frac{1}{2\pi\sigma_1\sigma_2} \exp\left(-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right) \frac{\sigma_1\sigma_2}{\sqrt{\sigma_1^2+\sigma_2^2}} \int_{-\infty}^{\infty} e^{-y^2/2} dy
\end{aligned}$$

ここで $\int_{-\infty}^{\infty} e^{-y^2/a^2} dy = \sqrt{a\pi}$ を用いると

$$g(x) = \frac{1}{\sqrt{2\pi}\sqrt{\sigma_1^2+\sigma_2^2}} \exp\left(-\frac{(x-m_1-m_2)^2}{2(\sigma_1^2+\sigma_2^2)}\right)$$

を得る．すなわち，独立な正規分布 $N(m_1, \sigma_1^2)$，$N(m_2, \sigma_2^2)$ の密度関数の畳み込み $g(x)$ は正規分布 $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ である．言い換えれば，正規分布 $N(m_1, \sigma_1^2)$，$N(m_2, \sigma_2^2)$ に従う独立な二つの確率変数 X_1, X_2 の和 $Y = X_1 + X_2$ は正規分布 $N(m_1 + m_2, \sigma_1^2 + \sigma_2^2)$ に従う．これは正規分布の再現性と呼ばれている． □

次に整数値をとる離散型分布の畳み込みを考える．独立な二つの確率変数 X_1, X_2 がそれぞれ分布 $\{p_1(k)\}$, $\{p_2(k)\}$ をもつとする．このとき確率変数 $Y = X_1 + X_2$ は整数値をとり，その確率関数 $q(k) = P(Y = k)$ は

$$q(k) = \sum_{i=-\infty}^{\infty} p_1(k-i)p_2(i) = \sum_{i=-\infty}^{\infty} p_1(i)p_2(k-i)$$

で与えられる．特に独立な非負の整数値を取る確率変数 X_1 と X_2 に対しては，確率変数 $Y = X_1 + X_2$ は非負の整数値をとり，

$$q(k) = \sum_{i=0}^{k} p_1(k-i)p_2(i) = \sum_{i=0}^{k} p_1(i)p_2(k-i)$$

となる。

例 **1.8** (ポワソン分布の再現性) パラメタ λ をもつ確率関数

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

に従う分布をポワソン分布 (Poisson distribution) という．ここでは，独立なパラメタ λ_1, λ_2 をもつポワソン分布に従う確率変数 X_1, X_2 の和 Y を考える．定義より，確率変数 Y の確率関数 $q(k)$ は

$$\begin{aligned}
q(k) &= \sum_{i=0}^{k} e^{-\lambda_1}\frac{\lambda_1^{k-i}}{(k-i)!} e^{-\lambda_2}\frac{\lambda_2^i}{i!} \\
&= e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!} \sum_{i=0}^{k} \frac{k!}{(k-i)!i!}\left(\frac{\lambda_1}{\lambda_1+\lambda_2}\right)^{k-i}\left(\frac{\lambda_2}{\lambda_1+\lambda_2}\right)^i
\end{aligned}$$

となり，右辺の和が $(\lambda_1/(\lambda_1+\lambda_2) + \lambda_2/(\lambda_1+\lambda_2))^k$ を二項展開したものになっていることに注意すると

$$q(k) = e^{-(\lambda_1+\lambda_2)}\frac{(\lambda_1+\lambda_2)^k}{k!}$$

を得る．これは，独立なパラメタ λ_1, λ_2 をもつポワソン分布に従う確率変数 X_1, X_2 の和 Y はパラメタ $\lambda_1 + \lambda_2$ をもつポワソン分布になることを示している． □

次に，複数個の確率変数の中からある確率法則に従って一つの確率変数を選んだときの分布を考える．n 個の確率変数 $X_1, X_2, \ldots, X_n$ がそれぞれ分布関数 $F_i(x)$ をもっており，確率 $p(i)$ $(i = 1, \ldots, n, p(1)+p(2)+\cdots+p(n) = 1)$ で確率変数 X_i を選んだときの値を Y とする．Y の分布 $G(x)$ は

$$G(x) = p(1)P(X_1 \leq x) + p(2)P(X_2 \leq x) + \cdots + p(n)P(X_n \leq x) = p(1)F_1(x) + p(2)F_2(x) + \cdots + p(n)F_n(x)$$

である．この $G(x)$ は分布関数 $F_1(x), F_2(x), \ldots, F_n(x)$ の混合 (mixture) と呼ばれる．

例 **1.9** 高々可算個の点 x_i $(i = 1, 2, \ldots)$ でのみ不連続であるような分布関数 $F(x)$ を考える．まず $q_i = F(x_i) - \lim_{x \to x_i -} F(x)$ $(i = 1, 2, \ldots)$ を定義する．このとき $q_i = P(X = x_i) > 0$ であることに注意する．さらに q_i の総和を q とし，$0 < q < 1$ であると仮定する．このとき，$p_i = q_i/q$ $(i = 1, 2, \ldots)$ とおくと，これは

$$p_i = P(X = x_i \mid X \text{ は } x_1, x_2, \ldots \text{のいずれかに等しい})$$

という条件付き確率であり，p_i の総和は 1 となる．このような p_i は条件付き確率関数 (conditional probability function) と呼ばれる．よって関数 $G_1(x)$ を

$$G_1(x) = \sum_{\{i | x_i \leq x\}} p_i$$

で定義すると，$G_1(x)$ は分布関数である．なお，このような分布関数は条件付き分布関数 (conditional distribution function) と呼ばれる．

一方，関数 $G_2(x)$ を

$$G_2(x) = \frac{F(x) - \displaystyle\sum_{\{i | x_i \leq x\}} q_i}{1 - q}$$

で定義すると，$G_2(x)$ は単調非減少な関数であり，$\lim_{x \to -\infty} G_2(x) = 0$ を満たす．また $\lim_{x \to \infty} F(x) = 1$ であるので $\lim_{x \to \infty} G_2(x) = 1$ となり，$G_2(x)$ も分布関数，すなわち X が x_i $(i = 1, 2, \ldots)$ のいずれにも等しくないという条件付き分布関数となる．もし $G_2(x)$ が密度関数 $g_2(x)$ をもつならば，$g_2(x)$ は条件付き密度関数 (conditional density function) と呼ばれる．

以上の議論より，高々可算個の点 x_i $(i = 1, 2, \ldots)$ でのみ不連続であるような分布関数 $F(x)$ は，離散的な条件付き分布関数 $G_1(x)$ と連続な条件付き分布関数 $G_2(x)$ の混合

$$G(x) = qG_1(x) + (1 - q)G_2(x)$$

で一意に表現される． □

問 **1.1** [5], [8] など参考にして，各自の計算機環境で，区間 (0,1) で定義された一様疑似乱数が生成できるように準備せよ[12]．さらに定理 1.1 を用いて分布関数 $F(x) = 1 - \exp(-x)$ $(x \geq 0)$ に従う標本を 300 万個生成し，値が 1 より大きな標本数の割合と理論値 $\exp(-1)$ を比較せよ． □

2 期待値

2.1 リーマン-スティルチェス積分

期待値を考えるには，まずリーマン-スティルチェス積分 (Riemann-Stieltjes Integral) に対する知識が必要となる．ここでは [12] に沿って，リーマンースティルチェス積分の概略を紹介する．ある区間 $[a, b]$ で定義された関数 $u(x)$ を考える．まず，区間 $[a, b]$ を n 個の部分区間 $[x_{i-1}, x_i]$ ($i = 1, \ldots, n$) に分ける．ただし $a = x_0 \leq x_1 \leq$

[12] もし [8] を参考にするなら ran1 あるいは ran2 を実装せよ (良い疑似乱数は一度，実装しておけば一生使うことが出来る数少ないプログラムの一つである)．

$\cdots \leq x_{n-1} \leq x_n = b$ である．さらに，M_i と m_i ($i = 1, \ldots, n$) をそれぞれ区間 $[x_i, x_{i-1}]$ における関数 $u(x)$ の上界値，下界値とする．

ここで，区間 $[a, b]$ で定義された右連続単調非減少関数 $F(x)$ に対して，$\Delta F_i = F(x_i) - F(x_{i-1})$ とする．このとき，上界，下界の和

$$\sum_{i=1}^{n} m_i \Delta F_i, \qquad \sum_{i=1}^{n} M_i \Delta F_i$$

が，$n \to \infty$ (すなわち $x_i - x_{i-1} \to 0$) の極限において一致する時，$u(x)$ の $F(x)$ に関するリーマン-スティルチェス積分 $\int_a^b u(x)dF(x)$ が以下のように定義される．

$$\int_a^b u(x)dF(x) = \lim_{n\to\infty} \sum_{i=1}^{n} m_i \Delta F_i = \lim_{n\to\infty} \sum_{i=1}^{n} M_i \Delta F_i$$

以下のことに注意する．ある点 $\tau \in [a, b]$ で $u(x)$ は連続であるが $F(x)$ が不連続な場合，$n \to \infty$ の極限において点 τ における上界あるいは下界の積和への貢献は $u(\tau)(F(\tau) - F(\tau-))$ となる．ただし，$F(\tau-)$ は $F(x)$ の $x = \tau$ における左極限 (x を左から τ に近付けた時の極限) である．よって，$F(x)$ が x_i でジャンプをもつ階段関数ならば，$x = x_i$ において定義可能な $u(x)$ に対して

$$\int_a^b u(x)dF(x) = \sum_{i=1}^{n} u(x_i)\Delta F_i$$

が成立する．

通常のリーマン積分 (Riemann Integral: 区分求積法による積分の定義) は $F(x) = x$ の場合に対応する．もし，区間 $[a, b]$ で $F(x)$ が微分可能であれば，リーマン-スティルチェス積分はリーマン積分で表現できる．

$$\int_a^b u(x)dF(x) = \int_a^b u(x)\frac{dF(x)}{dx}dx$$

なお，異常積分はリーマン積分と同様に極限として定義される．

$$\int_a^\infty u(x)dF(x) = \lim_{b\to\infty} \int_a^b u(x)dF(x)$$

2.2 期待値の定義

最初に分布関数 $F(x)$ をもつ非負の確率変数 X の期待値を定義する．

定義 2.1 (非負の確率変数の期待値) 分布関数 $F(x)$ をもつ非負の確率変数 X の期待値 (expectation) $E(X)$ は

$$E(X) = \int_0^\infty xdF(x)$$

で与えられる．$E(X)$ は無限大となることもある． □

次に分布関数 $F(x)$ をもつ任意の確率変数 X の期待値を考える．まず，$X^+ = \max(0, X)$，$X^- = -\min(0, X)$ によって新しい二つの確率変数を導入する．定義より $X^+ \geq 0$，$X^- \geq 0$ であり，$X = X^+ - X^-$ である．

定義 2.2 (確率変数の期待値) もし，$E(X^+)$ と $E(X^-)$ の少なくとも一方が有限であるならば，X の期待値は

$$E(X) = E(X^+) - E(X^-)$$

で与えられる．$E(X^+) = E(X^-) = \infty$ の場合は期待値は存在しない． □

もし，期待値 $E(X)$ が存在するならば，定義 2.2 より，

$$E(X) = \int_{-\infty}^{\infty} x dF(x)$$

となる．右辺は分布 $F(x)$ の平均 (mean) と呼ばれる．上式は分布 $F(x)$ が離散的ならば

$$E(X) = \sum_{i=1}^{\infty} x_i p(x_i)$$

と等価であり，分布 $F(x)$ が密度関数 $f(x)$ をもつならば

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx$$

と等価である．

期待値は以下の性質をもつ．

定理 2.1 (期待値の性質) 確率変数 X と Y の期待値 $E(X)$，$E(Y)$ が共に有限であるとする．このとき

1. $X \leq Y$ ならば $E(X) \leq E(Y)$
2. a を定数としたとき $E(aX) = aE(X)$
3. $E(X+Y) = E(X) + E(Y)$
4. X と Y が互いに独立であるならば $E(XY) = E(X)E(Y)$

が成立する． □

ある確率変数の列 $\{X_n;\ n = 1, 2, \ldots\}$ が全ての $\omega \in \Omega$ に対して $n \to \infty$ としたとき，$X_n(\omega) \to X(\omega)$ へ収束するとする．しかしながら，

$$\lim_{n\to\infty} E(X_n) = E(\lim_{n\to\infty} X_n) \tag{3}$$

は本質的に積分と極限の交換操作であるので，必ずしも成り立たない．

例 2.1 分布関数が $F(x) = 1 - \exp(-\lambda x)$ (パラメタ $\lambda > 0$) であるような分布を指数分布 (exponential distribution) という．確率空間 $(R, \mathcal{B}(\mathcal{R}), P)$ において，P がパラメタ λ をもつ指数分布に従うとする．すなわち，標本 ω が x 以下である確率は $P(\{\omega \leq x\}) = 1 - \exp(-\lambda x)$ で与えられる．この確率空間上で，次のような確率変数 X_n $(n = 1, 2, \ldots)$ を考える．

$$X_n(\omega) = \begin{cases} e^{\lambda n}, & \omega > n \\ 0, & \omega \leq n \end{cases}$$

定義より $P(\{\omega > n\}) = \exp(-\lambda n)$ であることに注意する．一方，確率変数 X は全ての ω に対して $X(\omega) = 0$ とする．明らかに $E(X) = 0$ である．$n \to \infty$ とすると $X_n \to X$ となるが，全ての $n = 1, 2, \ldots$ に対して $E(X_n) = \exp(-\lambda n)\exp(\lambda n) + \{1 - \exp(-\lambda n)\} \cdot 0 = 1$ であるので $E(X_n) \to E(X)$ は成立しない． □

極限と期待値の順序交換が可能な条件として，以下のような定理が知られている．

定理 2.2 (単調収束定理 (monotone convergence theorem)) $\{X_n;\ n = 1, 2, \ldots\}$ は非負確率変数列で，$0 \leq X_1 \leq X_2 \leq \cdots$ が確率 1 で成り立つとき，式 (3) が成立する． □

定理 2.3 (優収束定理 (dominated convergence theorem)) 確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ が全ての $\omega \in \Omega$ に対して $\lim_{n\to\infty} X_n(\omega) = X(\omega)$ であり，全ての $\omega \in \Omega$ と n に対して $|X_n(\omega)| \leq Y(\omega)$ となるような有限の平均をもつ非負確率変数 Y が存在するならば，式 (3) が成立する． □

注意 2.1 定数は確率変数なので，定理 2.3 において，$|X_n(\omega)| \leq M$ となるような正定数 M が存在するならば式 (3) が成立し，これは有界収束定理 (bounded convergence theorem) と呼ばれる． □

一般に，確率変数の関数に対しては次の定理が成り立つ．

定理 2.4 分布関数 $F(x)$ をもつ確率変数 X と $\mathcal{R}$ から $\mathcal{R}$ への可測関数 $u(x)$ に対して期待値 $E(u(X))$ が存在するならば，それは

$$E(u(X)) = \int_{-\infty}^{\infty} u(x)dF(x)$$

で与えられる． □

例 2.2 非負の確率変数 X の期待値を別の方法で求める．X が非負ならば $X = \int_0^X dx$ と形式的にかけるので

$$E(X) = E\left(\int_0^X dx\right) = E\left(\int_0^\infty 1_{\{x \le X\}} dx\right)$$

である．ここまで見てきたように期待値は積分と等価であるので，被積分関数が非負ならば，重積分と同様に，期待値と積分の順序を交換することが出来る．よって

$$E(X) = \int_0^\infty E(1_{\{x \le X\}})dx = \int_0^\infty (1 - F(x))dx$$

が成立する[13]． □

多次元確率変数の関数に対しても同様の定理が成り立つ．

定理 2.5 結合分布関数 $F(x_1, \ldots, x_n)$ をもつ n 次元確率変数 $\boldsymbol{X} = (X_1, \ldots, X_n)$ と $\mathcal{R}^n$ から $\mathcal{R}^n$ への可測関数 $g(x_1, \ldots, x_n)$ に対して期待値 $E(g(\boldsymbol{X}))$ が存在するならば，それは

$$E(g(\boldsymbol{X})) = \int_{x_1=-\infty}^{\infty} \cdots \int_{x_n=-\infty}^{\infty} g(x_1, \ldots, x_n)F(dx_1, \ldots, dx_n)$$

で与えられる． □

最後に分布と期待値の間に成立する不等式を導く．分布関数 $F(x)$ をもつ非負の確率変数 X に対して

$$P(X \ge x) = \int_{x-}^{\infty} dF(y) \le \int_{x-}^{\infty} \frac{y}{x} dF(y) \le \frac{1}{x}\int_0^\infty ydF(y) = \frac{E[X]}{x}$$

となるので，次の定理を得る．

定理 2.6 (マルコフ不等式) 非負の確率変数 X に対して

$$P(X \ge x) \le \frac{E(X)}{x}, \qquad x > 0$$

が成立する．これはマルコフ不等式 (Markov's Inequality) と呼ばれている． □

もし、$x \le E(X)$ ならば $E(X)/x \ge 1$ となるので，マルコフ不等式は $x > E(X)$ の場合にのみ意味がある。

2.3 積率，分散，共分散

X を分布関数 $F(x)$ をもつ確率変数とする．このとき，$\mathcal{R}$ から $\mathcal{R}$ への可測関数 $g(x)$ に対する期待値 $E(g(X))$ は，分布関数 $F(x)$ の関数と見なすことが出来る[14]．それゆえ，期待値 $E(g(X))$ を調べることで分布関数 $F(x)$ の特徴をある程度把握することが出来る．このような目的で良く用いられるものには $g(x) = x^n$ や $g(x) = (x - E(X))^n$ などがある．

定義 2.3 分布関数 $F(x)$ をもつ確率変数 X に対して，$E(X^n)$ を確率変数 X (あるいは分布 $F(x)$) の $\boldsymbol{n}$ 次積率 (nth moment) という．$n = 1$ のとき，これは確率変数 X の期待値，あるいは分布 $F(x)$ の平均に等しい．また，$E(\{X - E(X)\}^n)$ を確率変数 X (あるいは分布 $F(x)$) の $\boldsymbol{n}$ 次中心積率 (nth central moment) という．特に，$n = 2$ の場合，$E(\{X - E(X)\}^2)$ を分散 (variance) と呼び，$V(X)$ と書く． □

[13] $\int_0^\infty (1 - F(x))dx = [x(1 - F(x))]_0^\infty + \int_0^\infty xdF(x)$ であり，$x(1 - F(x)) = x\int_x^\infty dF(u) \le \int_x^\infty udF(u)$ が $x \to \infty$ のとき 0 へ収束することからも示すことができる．

[14] このような，関数の関数を汎関数 (functional) という．

分散 $V(X) = E(\{X - E(X)\}^2)$ は確率変数 X の値が期待値からどの程度離れた値を取るかという，バラツキを表す指標として用いられる．分散には次のような性質がある．

定理 **2.7** $E(X^2) < \infty$ ならば

1. $V(X) = E(X^2) - \{E(X)\}^2$
2. 定数 a, b に対して $V(aX + b) = a^2V(X)$
3. $V(X) = 0$ ならば $P(X = E(X)) = 1$

が成立する． □

さて，分散 $V(X)$ をもつ確率変数 X に対して $Y = (X - E(X))^2$ により，新しい確率変数 Y を定義する．Y に定理 2.6 のマルコフ不等式を適用すると

$$P(Y \geq x^2) \leq \frac{E(Y)}{x^2}$$

を得る．ここで $x > 0$ に対して $P(Y \geq x^2) = P((X-E(X))^2 \geq x^2) = P(|X-E(X)| \geq x)$ であり，$E(Y) = V(X)$ に注意すると

$$P(|X - E(X)| \geq x) \leq \frac{V(X)}{x^2} \tag{4}$$

を得る．これはチェビシェフの不等式 (Chebyshev's inequality) と呼ばれている．マルコフ不等式が分布の上側の裾野の上界値のみを与え，また，非負の確率変数のみに適用できるのに対し，チェビシェフの不等式は分布の両側の裾野の上界値を与えていることに注意する．また，マルコフ不等式の上界値は 0 からの距離の逆数に比例する形で 0 へ向かうのに対し，チェビシェフの不等式の上界値は平均からの距離の二乗の逆数に比例する形で 0 へ向かう．

確率変数 X に対して $0 < V(X) < \infty$ のとき

$$Y = \frac{X - E(X)}{\sqrt{V(X)}} \tag{5}$$

によって新しい確率変数 Y を定義すると，Y の期待値は 0，分散は 1 となる．このような変換は標準化 (standardization) と呼ばれており，分母に現れる $\sqrt{V(X)}$ は標準偏差 (standard deviation) と呼ばれる．また，非負の確率変数 X に対しては，バラツキを表す無次元の量として変動係数 (coefficient of variation) $C(X)$ がある．

$$C(X) = \frac{\sqrt{V(X)}}{E(X)}$$

次に二つの確率変数の間の依存性を表す指標を導入する．

定義 **2.4** (共分散) 平均をもつ二つの確率変数 X と Y に対して定義された

$$Cov(X, Y) = E(\{X - E(X)\}\{Y - E(Y)\}) \tag{6}$$

を確率変数 X と Y の共分散 (covariance) という． □

共分散 $Cov(X, Y)$ は，期待値の性質を用いると

$$Cov(X, Y) = E(XY) - E(X)E(Y)$$

と書くことが出来る．もし $X = Y$ ならば共分散は分散に等しい．また，X と Y が独立ならば $E(XY) = E(X)E(Y)$ となるので，共分散は 0 である．

一般に n 個の確率変数 $X_1, X_2, \ldots, X_n$ の和の分散は共分散を用いて

$$V(X_1 + \cdots + X_n) = \sum_{i=1}^{n} V(X_i) + 2\sum_{i<j} Cov(X_i, X_j)$$

で与えられる．よって，n 個の確率変数 $X_1, X_2, \ldots, X_n$ が互いに独立ならば，それらの和の分散はそれぞれの分散の和で与えられる．

補題 **2.1** (コーシー - シュワルツの不等式) $E(X^2)$，$E(Y^2)$ が有限で正の値を取るならば

$$\{E(XY)\}^2 \leq E(X^2)E(Y^2)$$

が成立する．等号成立はある定数 a に対して $Y = aX$ となるときのみである．これはコーシー - シュワルツの不等式 (Cauchy-Schwarz inequality) と呼ばれている． □

任意の実数 t に対して $E((X-tY)^2) \geq 0$ であるので，$E((X-tY)^2) = E(X^2) - 2E(XY)t + E(Y^2)t^2 \geq 0$ が成立する．これを t の 2 次関数とみれば判別式は非正でなければならない．よって $\{E(XY)\}^2 - E(X^2)E(Y^2) \leq 0$ を得る．また，等号が成り立つのは $X - tY = 0$ の場合に限る．

共分散の $Cov(X,Y)$ の取る値は，例えば X を a 倍すると a 倍となる．このように共分散は単位の取り方に依存する．そこでこのような単位の取り方に依存しないように正規化した指標が相関係数である．

定義 **2.5** (相関係数) 確率変数 X, Y が有限でかつ 0 でない分散を持つとする．このとき

$$\rho_{X,Y} = \frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}} \tag{7}$$

で定義される $\rho_{X,Y}$ を確率変数 X と Y の相関係数 (correlation coefficient) という． □

補題 2.1 より $-1 \leq \rho_{X,Y} \leq 1$ である[15]．特にある正数 a に対して $Y = aX$ ならば $\rho_{X,Y} = 1$ となり，$Y = -aX$ ならば $\rho_{X,Y} = -1$ となる．

2.4 母関数

確率変数の特性を分布関数以外で表す方法に母関数がある．以下で見るように母関数を用いれば，積率の計算や独立な確率変数の和に関する操作が簡単化される．

定義 **2.6** 非負の整数値をとる確率変数 X が確率関数 $p_i = P(X = i)$ をもつとする．このとき確率変数 X の確率母関数[16] (probability generating function) $X(z)$ は

$$X(z) = E\left(z^X\right) = \sum_{i=0}^{\infty} z^i p_i, \qquad |z| \leq 1$$

で与えられる． □

確率母関数は以下の性質をもつ．

1. 確率母関数 $X(z)$ は $|z| \leq 1$ で収束し，$|z| < 1$ で何回でも微分可能である．
2. $X^{(n)}(z)$ を確率母関数 $X(z)$ の n 階導関数とすると，$E[X^n] < \infty$ ならば $X^{(n)}(1) = E[X(X-1)\cdots(X-n+1)]$ であり，これは $\boldsymbol{n}$ 次階乗積率 (nth factorial moment) と呼ばれる．特に X の平均 $E[X]$ は $X^{(1)}(1)$ で与えられ，2 次積率 $E[X^2]$ は $X^{(2)}(1) + X^{(1)}(1)$ で与えられる．
3. 非負の確率分布と確率母関数は 1 対 1 に対応している．
4. 二つの確率変数 X，Y の確率母関数をそれぞれ $X(z)$，$Y(z)$ とすると，これらの和 $Z = X + Y$ の確率母関数 $Z(z)$ は $Z(z) = E(z^{X+Y}) = E(z^X z^Y)$ となるので，二つの確率変数 X，Y が独立ならば，これらの和 $Z = X + Y$ の確率母関数 $Z(z)$ は積 $Z(z) = X(z)Y(z)$ で与えられる．

連続な確率変数に対して用いられる母関数には次のようなものがある．

[15] 補題 2.1 において，X の代わりに $X - E(X)$，Y の代わりに $Y - E(Y)$ とすると $Cov(X,Y)^2 \leq V(X)V(Y)$ を得る．
[16] 確率母関数は Z 変換とも呼ばれる．

定義 2.7 分布関数 $F(x)$ をもつ連続な確率変数 X に対して

$$\psi(\eta) = E\left(e^{\eta X}\right) = \int_{-\infty}^{\infty} e^{\eta x} dF(x) < \infty,$$

となる正数 η が存在するならば，$\psi(\eta)$ を確率変数 X (あるいは分布 $F(x)$) の積率母関数 (moment generating function) という．□

もし，確率変数 X の積率母関数 $\psi(\eta)$ が存在し，かつ，n 次積率 $E(X^n)$ が存在するならば

$$E(X^n) = \psi^{(n)}(0)$$

である．ただし $\psi^{(n)}(\eta)$ は $\psi(\eta)$ の n 階導関数である．積率母関数をもつ確率変数に対しては以下のような限界式が知られている．

定理 2.8 (シャーノフの限界式) 積率母関数をもつ確率変数 X に対して，シャーノフの限界式 (Chernoff's bound)

$$P(X \geq x) \leq \exp(-I(x)) \tag{8}$$

が成立する．ただし，$I(x)$ は確率変数 X の対数-積率母関数 (log-moment generating function) $\ln E\left(e^{\eta X}\right)$ のルジャンドル変換 (Legendre transform) であり

$$I(x) = \sup_{\eta \geq 0} \left(\eta x - \ln E\left(e^{\eta X}\right)\right)$$

で与えられる．□

任意の $\eta \geq 0$ に対して

$$P(X \geq x) \leq P\left(e^{\eta X} \geq e^{\eta x}\right)$$

ただし，等号は $\eta > 0$ のとき成立する．さらに，定理 2.6 のマルコフ不等式より

$$P\left(e^{\eta X} \geq e^{\eta x}\right) \leq \frac{E\left(e^{\eta X}\right)}{e^{\eta x}} = e^{-\eta x} \exp\left(\ln E\left(e^{\eta X}\right)\right) = \exp\left(-\left\{\eta x - \ln E\left(e^{\eta X}\right)\right\}\right)$$

が任意の η に対して成立する．よって，η をうまく選んで上界を出来るだけ小さくすれば

$$P(X \geq x) \leq \inf_{\eta \geq 0} \exp\left(-\left\{\eta x - \ln E\left(e^{\eta X}\right)\right\}\right)$$

となり，シャーノフの限界式が得られる．

注意 2.2 $f(x)$ ($-\infty < x < \infty$) を非負，非減少な関数とする．このとき

$$P(X \geq x) \leq P(f(X) \geq f(x))$$

となり，定理 2.6 のマルコフ不等式を上式の右辺に用いると

$$P(X \geq x) \leq \frac{E(f(X))}{f(x)}, \qquad f(x) > 0$$

を得る．これもマルコフ不等式と呼ばれることがある．シャーノフ限界式は $f(x) = e^{\eta x}$ の場合に対応する． □

注意 2.3 シャーノフの限界式は $x > E(X)$ のときだけに意味がある．もし，そうでないとすると，これは単に $P(X \geq x) \leq 1$ となってしまうかも知れない． □

注意 2.4 マルコフ不等式、チェビシェフの不等式がそれぞれ、平均、分散のみを情報として用いているのに対し、シャーノフの限界式は分布がもつ全ての情報を用いて構築されている。それ故、マルコフ不等式やチェビシェフの不等式と比較して、精度が高いと考えられる。 □

さらに，連続な確率変数 X に対してはラプラス・スティルチェス変換 (Laplace-Stieltjes transform) がある．非負の確率変数 X の分布関数を $F(x)$ としたとき，確率変数 X (あるいは分布 $F(x)$) に対するラプラス・スティルチェス変換 $X^*(s)$ は

$$X^*(s) = E\left(e^{-sX}\right) = \int_0^{\infty} e^{-sx} dF(x), \qquad \mathrm{Re}(s) > 0$$

で与えられる．X の密度関数 $f(x) = dF(x)/dx$ が存在する場合は，

$$X^*(s) = \int_0^\infty e^{-sx} f(x) dx$$

であり，密度関数に対する通常のラプラス変換と等価である．ラプラス・スティルチェス変換は密度関数が存在しないような場合に通常のラプラス変換を拡張したものと見ることができる．ラプラス・スティルチェス変換は次の性質をもつ．

1. $X^*(s)$ は $\mathrm{Re}(s) \geq 0$ で収束し，$\mathrm{Re}(s) > 0$ で何回でも微分可能である．
2. $X^{*(n)}(s)$ を $X(s)$ の n 階導関数とすると，$E(X^n) < \infty$ ならば $X^{*(n)}(0) = (-1)^n E(X^n)$ となる．特に X の平均 $E(X)$ は $-X^{*(1)}(0)$ で与えられ，2 次積率 $E(X^2)$ は $X^{*(2)}(0)$ で与えられる．
3. 確率分布とラプラス・スティルチェス変換は 1 対 1 に対応している．
4. 二つの確率変数 X，Y のラプラス・スティルチェス変換をそれぞれ $X^*(s)$，$Y^*(s)$ とすると，これらの和 $Z = X + Y$ のラプラス・スティルチェス変換 $Z^*(s)$ は $Z^*(s) = E[e^{-s(X+Y)}] = E[e^{-sX}e^{-sY}]$ となるので，二つの確率変数 X，Y が独立ならば，これらの和 $Z = X + Y$ のラプラス・スティルチェス変換 $Z^*(s)$ は積 $Z^*(s) = X^*(s)Y^*(s)$ で与えられる．

また，実数 η を引数とする関数

$$\varphi(\eta) = E\left(e^{i\eta X}\right) = E(\cos(\eta X)) + iE(\sin(\eta X))$$

を確率変数 X の特性関数 (characteristic function) という．ただし i は虚数単位である ($i^2 = -1$)．特性関数は常に存在し，確率変数と 1 対 1 に対応している．

2.5 条件付き期待値と最小二乗法

確率変数 X に対して，$F(x \mid A)$ を $P(A) > 0$ であるような事象 A が起こったという条件の下での条件付き分布関数とする．

$$F(x \mid A) = P(X \leq x \mid A)$$

可測関数 $h(x)$ が与えられたとき，$F(x)$ の代わりに $F(x \mid A)$ を用いて定義された $h(X)$ の期待値

$$E(h(X) \mid A) = \int_{-\infty}^{\infty} h(x) dF(x \mid A) \tag{9}$$

を A に関する $h(X)$ の条件付き期待値 (conditional expectation) という．もし $A = \Omega$ ならば条件付き期待値 $E(h(X) \mid \Omega)$ は期待値 $E(h(X))$ に等しい．

$P(A_i) > 0$ なる事象 A_i $(i = 1, \ldots, n)$ を互いに排反で $\Omega = \cup_{i=1}^n A_i$ を満たすとする．このとき

$$E(X) = \sum_{i=1}^{n} E(X \mid A_i) P(A_i)$$

が成立する．ここで n は無限大であっても良い．

式 (9) において，条件に対応する事象は，例えば，ある事象 $B \in \mathcal{B}(\mathcal{R})$ に対して $A = \{X \in B\}$ といった確率変数 X に対する事象の場合もあれば，他の確率変数 Y に対して $A = \{Y \in B\}$ という場合もある．以下では後者の場合について考える．

二つの確率変数 X と Y に対して，$Y = y$ に関する $h(X)$ の条件付き期待値を必要とする場合がしばしば現れる．もし，Y が離散的な確率変数であり，$P(Y = k) > 0$ ならば，定義より，確率変数 X の条件付き分布関数 $F(x \mid Y = k)$ は

$$F(x \mid Y = k) = \frac{P(X \leq x, Y = k)}{P(Y = k)} \tag{10}$$

で与えられるので，これを式 (9) へ適用すれば良い．もし，X と Y が独立ならば $P(X \leq x, Y = k) = P(X \leq x)P(Y = k)$ となるので $F(x \mid Y = k) = P(X \leq x) = F(x)$ となり，条件付き期待値 $E(h(X) \mid Y = y)$ は期待値 $E(h(X))$ に等しい．また，$E(|h(X)|) < \infty$ ならば

$$E(h(X)) = \sum_{k \in \Omega} E(h(X) \mid Y = k)P(Y = k)$$

が成立する．

次に，Y は連続な確率変数であるとする．このとき任意の y に対して $P(Y = y) = 0$ となり，式 (10) は意味をなさない．このような困難を克服するにはラドン-ニコディムの定理 (Radon-Nikodym Theorem) が必要となる．しかし，これは本講義の範囲を越えるので，以下では確率変数 X と Y の連続な結合密度関数 $f(x, y)$ が存在すると仮定して話を進める．

まず，微小な正数 ϵ に対して事象 $B_\epsilon = \{y - \epsilon \leq Y \leq y + \epsilon\}$ を定義する．$\lim_{\epsilon \to 0} B_\epsilon = \{Y = y\}$ であるので，$Y = y$ に関する $h(X)$ の期待値 $E(h(X) \mid Y = y)$ を

$$E(h(X) \mid Y = y) = \lim_{\epsilon \to 0} E(h(X) \mid B_\epsilon) \tag{11}$$

で定義する．ここでの目的はこの条件付き期待値を連続な結合密度関数 $f(x, y)$ を用いて表現することである．準備として，確率変数 Y の周辺密度関数 $g(y)$ を定義する．

$$g(y) = \int_{-\infty}^{\infty} f(x, y)dx, \quad -\infty < y < \infty$$

このとき $P(B_\epsilon) > 0$ なので式 (9) が適用でき，

$$E(h(X) \mid B_\epsilon) = \frac{\displaystyle\int_{x=-\infty}^{\infty} h(x) \int_{u=y-\epsilon}^{y+\epsilon} f(x, u)dxdu}{\displaystyle\int_{y-\epsilon}^{y+\epsilon} g(u)du} \tag{12}$$

である．仮定より $f(x, u)$ は連続なので，分子に現れる積分 $\int_{y-\epsilon}^{y+\epsilon} f(x, u)du$ へ平均値の定理を適用すると，ある ξ $(y - \epsilon \leq \xi \leq y + \epsilon)$ が存在して

$$\int_{y-\epsilon}^{y+\epsilon} f(x, u)du = 2\epsilon f(x, \xi)$$

となる．同様に分母に対してもある η $(y - \epsilon \leq \eta \leq y + \epsilon)$ が存在して

$$\int_{y-\epsilon}^{y+\epsilon} g(u)du = 2\epsilon g(\eta)$$

となる．$\epsilon \to 0$ のとき，$\xi \to y$, $\eta \to y$ であるので式 (11)，(12) より

$$E(h(X) \mid Y = y) = \lim_{\epsilon \to 0} \frac{\displaystyle\int_{x=-\infty}^{\infty} h(x) \int_{u=y-\epsilon}^{y+\epsilon} f(x, u)dxdu}{\displaystyle\int_{y-\epsilon}^{y+\epsilon} g(u)du} = \int_{-\infty}^{\infty} h(x) \cdot \frac{f(x, y)}{g(y)}dx$$

を得る．ここで現れた $f(x, y)/g(y)$ は $Y = y$ という条件の下での X の条件付き密度関数である．

もし，X と Y が独立ならば，確率変数 X の周辺密度関数を $f(x)$ としたとき $f(x, y) = f(x)g(y)$ となり，条件付き期待値 $E(h(X) \mid Y = y)$ は期待値 $E(h(X))$ に等しい．また，$E(|h(X)|) < \infty$ ならば

$$E(h(X)) = \int_{-\infty}^{\infty} E(h(X) \mid Y = y)g(y)dy$$

が成立する．

さて，連続な二つの確率変数 X と Y が独立でない場合に，ある関数 $h(x)$ を用いて $Y = h(X)$ というように関連づけようとする場合がある．特に $E(Y^2)$ ならびに $E(\{h(X)\}^2)$ が有限であるという条件の下で

$$L = E(\{Y - h(X)\}^2)$$

なる L が最小となるような $h(x)$ を求めるのが最小二乗法 (method of least squares) である．以下では，条件付き期待値を用いて最小二乗法の原理を説明する．

$X = x$ であるという条件の下での Y の条件付き期待値 $E(Y \mid X = x)$ は x の (可測) 関数となっている．そこで，この関数を $E(Y \mid X)$ と書くことにする．X の分布関数を $F(x)$ としたとき，

$$E(Y) = E(E(Y \mid X)) = \int_{-\infty}^{\infty} E(Y \mid X = x) dF(x)$$

である．よって

$$L = E(\{Y - h(X)\}^2) = E(E(\{Y - h(X)\}^2 \mid X))$$

であるので，全ての $X = x$ に対して，$E(\{Y - h(X)\}^2 \mid X = x)$ が最小となるように $h(x)$ を定めると，L も最小となる．さらに $E(\{Y - h(X)\}^2 \mid X = x) = E(\{Y - h(x)\}^2 \mid X = x)$ に注意すると

$$\begin{aligned} E(\{Y - h(X)\}^2 \mid X = x) &= E(Y^2 \mid X = x) - 2h(x)E(Y \mid X = x) + h(x)^2 \\ &= (h(x) - E(Y \mid X = x))^2 + E(Y^2 \mid X = x) - \{E(Y \mid X = x)\}^2 \\ &= (h(x) - E(Y \mid X = x))^2 + E((Y - E[Y])^2 \mid X = x) \end{aligned}$$

であるので次の定理を得る．

定理 2.9 (最小二乗法の原理) $E(Y^2)$ ならびに $E(\{h(X)\}^2)$ が有限であるとき，L を最小にする関数 $h(x)$ は $h(x) = E(Y \mid X = x)$ で与えられる． □

上では一般の関数 $h(x)$ を考えたが，特に $h(x)$ が 1 次関数の場合が応用上，重要である．以下では $h(x) = ax + b$ と仮定して L を最小とするような定数 a，b を求める．まず，

$$L = E(\{Y - aX - b\}^2) = E(Y^2) + a^2E(X^2) + b^2 - 2aE(XY) - 2bE(Y) + 2abE(X)$$

であるので，L は a, b の 2 次関数と見ることが出来る．よって L を最小とする a, b は

$$\frac{\partial L}{\partial a} = 0, \qquad \frac{\partial L}{\partial b} = 0$$

によって決定される．実際に計算すると

$$\frac{\partial L}{\partial a} = 2aE(X^2) - 2E(XY) + 2bE(X) = 0, \qquad \frac{\partial L}{\partial b} = 2b - 2E(Y) + 2aE(X) = 0$$

となり，この連立方程式を解くことにより，L を最小とする a, b は

$$a = \frac{E(XY) - E(X)E(Y)}{E(X^2) - \{E(X)\}^2} = \frac{Cov(X,Y)}{V(X)}, \qquad b = E(Y) - E(X)\frac{Cov(X,Y)}{V(X)}$$

を得る．すなわち直線 $y = h(x)$ は

$$y - E(Y) = \frac{Cov(X,Y)}{V(X)}(x - E(X))$$

で与えられる．この直線を Y の X 上への回帰直線 (regression line) といい，傾き

$$\frac{Cov(X,Y)}{V(X)} = \rho_{X,Y}\sqrt{\frac{V(Y)}{V(X)}}$$

を回帰係数 (regression coefficient) という．

2.6 代表的な分布の例

ここで代表的な分布を紹介する．まず，連続型分布の代表的なものには以下のようなものがある．

例 **2.3** (連続型分布の例)

- 区間 $[a,b]$ 上の一様分布 (uniform distribution) :

$$f(x)=\begin{cases}\frac{1}{b-a}, & a\le x\le b\\ 0, & \text{その他}\end{cases}\qquad F(x)=\begin{cases}0, & x<a\\ \frac{x-a}{b-a}, & a\le x\le b\\ 1, & x>b\end{cases}$$

  平均 $(a+b)/2$，分散 $(b-a)^2/12$．
- 指数分布 (exponential distribution): パラメタ λ (>0)

$$f(x)=\begin{cases}\lambda e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}\qquad F(x)=\begin{cases}1-e^{-\lambda x}, & x\ge 0\\ 0, & x<0\end{cases}$$

  平均 $1/\lambda$，分散 $1/\lambda^2$．変動係数が 1 であり，バラツキの程度と評価する際の標準的な分布として用いられる．
- ガンマ分布 (gamma distribution): パラメタ α (>0), β (>0)

$$f(x)=\begin{cases}\frac{\beta^\alpha x^{\alpha-1}}{\Gamma(\alpha)}e^{-\beta x}, & x\ge 0\\ 0 & x<0\end{cases}$$

  ただし $\Gamma(\alpha)=\int_0^\infty y^{\alpha-1}e^{-y}dy$ はガンマ関数である．平均 α/β，分散 α/β^2．特にある自然数 k に対して $\alpha=k$，$\beta=k\lambda$ のとき，$\Gamma(\alpha)=\alpha!$ となり、これをパラメタ λ をもつ k 次のアーラン分布 (Erlang distribution) という．1 次のアーラン分布は指数分布と等価である．
- 正規分布 (normal distribution, Gaussian distribution): パラメタ 平均 m, 標準偏差 σ (>0)

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)$$

  通常 $N(m,\sigma^2)$ と表記される．特に $m=0$，$\sigma=1$ のとき，標準正規分布 (standard normal distribution) と呼ばれる．
- ワイブル分布 (Weibull distribution): パラメタ η (>0), β (>0)

$$f(x)=\begin{cases}\frac{\beta x^{\beta-1}}{\eta^\beta}\exp\left(-\left(\frac{x}{\eta}\right)^\beta\right), & x\ge 0\\ 0, & x<0\end{cases}\qquad F(x)=\begin{cases}1-\exp\left(-\left(\frac{x}{\eta}\right)^\beta\right) & x\ge 0\\ 0, & x<0\end{cases}$$

  平均 $\eta\Gamma((\beta+1)/\beta)$，分散 $\eta^2[\Gamma((\beta+2)/\beta)-\{\Gamma((\beta+1)/\beta)\}^2$．

□

代表的な離散型分布には次のようなものがある．

例 **2.4** (離散型分布の例)

- 二項分布 (binomial distribution): パラメタ n (自然数), α $(0<\alpha<1)$

$$p(k)=\binom{n}{k}\alpha^k(1-\alpha)^{n-k},\qquad k=0,1,\dots,n$$

  平均 $n\alpha$，分散 $n\alpha(1-\alpha)$．
- 幾何分布 (geometric distribution): パラメタ α $(0<\alpha<1)$

$$p(k)=(1-\alpha)\alpha^k,\qquad k=0,1,\dots$$

  平均 $\alpha/(1-\alpha)$，分散 $\alpha/(1-\alpha)^2$．

- ポワソン分布 (Poisson distribution): パラメタ 平均 λ (> 0)

$$p(k) = e^{-\lambda}\frac{\lambda^k}{k!}, \qquad k = 0, 1, \ldots$$

分散 λ .

□

代表的な多次元分布には次のようなものがある .

例 **2.5** (多次元分布の例)

- 多項分布 (multinomial distribution): パラメタ n (自然数) , k (自然数) , p_i $(i = 1, \ldots, n)$

$$p(k_1, k_2, \ldots, k_n) = \frac{k!}{k_1!k_2!\cdots k_n!}p_1^{k_1}p_2^{k_2}\cdots p_n^{k_n} \qquad k_i = 0, 1, \ldots, n, k_1 + \cdots + k_n = k$$

- $\boldsymbol{n}$ 次元正規分布 (multnormal distribution): パラメタ 平均ベクトル $\boldsymbol{m} = (m_1, m_2, \ldots, m_n)$, 非負対称正定値行列 (共分散行列) $\boldsymbol{\Sigma}$

$$f(x_1, x_2, \ldots, x_n) = \frac{1}{\sqrt{|\boldsymbol{\Sigma}|}(\sqrt{2\pi})^n}\exp\left(-\frac{1}{2}(\boldsymbol{x} - \boldsymbol{m})\boldsymbol{\Sigma}^{-1}(\boldsymbol{x} - \boldsymbol{m})^T\right)$$

ただし $\boldsymbol{x} = (x_1, x_2, \ldots, x_n)$, T は転置を表す .

□

3 確率変数と分布の収束

この章では確率変数ならびに分布の様々な収束の概念について解説する . これらの収束概念に対して正確な感覚を身につけることは , 確率・統計における最も重要な定理である大数の法則や中心極限定理の意味するところを正しく理解するために必須である .

3.1 事象の列とボレル-カンテリの定理

確率空間 $(\Omega, \mathcal{F}, P)$ に対して事象の列 $\{A_n;\ n = 1, 2, \ldots\}$ が $A_1 \subset A_2 \subset A_3 \subset \cdots$ であるとき , 増加列という . 増加列 $\{A_n;\ n = 1, 2, \ldots\}$ の極限を $\lim_{n\to\infty} A_n = \cup_{k=1}^{\infty} A_k$ で定義する . 同様に , 事象の列 $\{A_n;\ n = 1, 2, \ldots\}$ が $A_1 \supset A_2 \supset A_3 \supset \cdots$ であるとき , 減少列という . 減少列 $\{A_n;\ n = 1, 2, \ldots\}$ の極限を $\lim_{n\to\infty} A_n = \cap_{k=1}^{\infty} A_k$ で定義する . 事象の列 $\{A_n;\ n = 1, 2, \ldots\}$ が増加列または減少列であるとき , $\{A_n;\ n = 1, 2, \ldots\}$ は単調列と呼ばれる .

任意の事象の列 $\{A_n;\ n = 1, 2, \ldots\}$ に対して $B_n = \cup_{k=n}^{\infty} A_k$ である新しい事象の列 $\{B_n;\ n = 1, 2, \ldots\}$ を定義する . このとき $\{B_n;\ n = 1, 2, \ldots\}$ は減少列となる . そこで , その極限を上極限 (superior limit) と呼び , $\limsup_{n\to\infty} A_n$ と書く .

$$\lim_{n\to\infty} B_n = \cap_{n=1}^{\infty} B_n = \cap_{n=1}^{\infty} \cup_{k=n}^{\infty} A_k = \limsup_{n\to\infty} A_n$$

$\limsup_{n\to\infty} A_n$ は , 無限に多くの A_n に含まれている要素の全体からなる集合である . すなわち

$$\begin{aligned} \omega \in \limsup_{n\to\infty} A_n &\Leftrightarrow \text{全ての } n \text{ に対して } \omega \in B_n \\ &\Leftrightarrow \text{全ての } n \text{ に対して , ある } k = k(\omega)\ (k \geq n) \text{ が存在し } \omega \in A_k \end{aligned}$$

同様に，事象の列 $\{A_n;\ n=1,2,\ldots\}$ に対して $C_n=\cap_{k=n}^{\infty}A_k$ である新しい事象の列 $\{C_n;\ n=1,2,\ldots\}$ を定義する．このとき $\{C_n;\ n=1,2,\ldots\}$ は増加列となる．そこで，その極限を下極限(inferior limit)と呼び，$\liminf_{n\to\infty}A_n$ と書く．

$$\lim_{n\to\infty}C_n=\cup_{n=1}^{\infty}C_n=\cup_{n=1}^{\infty}\cap_{k=n}^{\infty}A_k=\liminf_{n\to\infty}A_n$$

$\liminf_{n\to\infty}A_n$ は，最初の有限個を除いた残り全ての A_n に含まれている要素の全体からなる集合である．すなわち

$$\begin{aligned}\omega\in\liminf_{n\to\infty}A_n &\Leftrightarrow \text{ある } n=n(\omega) \text{ が存在し },\omega\in C_n\\ &\Leftrightarrow \text{ある } n=n(\omega) \text{ が存在し，全ての } k\ (k\geq n) \text{ に対して } \omega\in A_k\end{aligned}$$

上極限と下極限が等しいとき，$A=\limsup_{n\to\infty}A_n=\liminf_{n\to\infty}A_n$ を事象の列 $\{A_n;\ n=1,2,\ldots\}$ の極限(limit)という．

定理 3.1 (確率分布の連続性) 事象の列 $\{A_n;\ n=1,2,\ldots\}$ が極限をもつならば

$$\lim_{n\to\infty}P(A_n)=P\left(\lim_{n\to\infty}A_n\right)$$

が成立する． □

もし，$\{A_n;\ n=1,2,\ldots\}$ が増加列ならば $D_0=\phi$, $D_n=A_n-A_{n-1}$ $(n=1,2,\ldots)$ とすれば $\{D_n;\ n=1,2,\ldots\}$ は排反な事象の列となる．よって，定義 1.2 で与えた σ-加法性より

$$P\left(\bigcup_{n=1}^{\infty}D_n\right)=\sum_{n=1}^{\infty}P(D_n)=\lim_{n\to\infty}\sum_{i=1}^{n}(P(A_i)-P(A_{i-1}))=\lim_{n\to\infty}P(A_n)$$

となり，$\cup_{n=1}^{\infty}D_n=\cup_{n=1}^{\infty}A_n=\lim_{n\to\infty}A_n$ であるので定理を得る．
もし，$\{A_n;\ n=1,2,\ldots\}$ が減少列ならば $\{A_n^c;\ n=1,2,\ldots\}$ は増加列であるので

$$\lim_{n\to\infty}P(A_n^c)=P(\lim_{n\to\infty}A_n^c)=P\left(\bigcup_{n=1}^{\infty}A_n^c\right)=P\left((\cap_{n=1}^{\infty}A_n)^c\right)=1-P\left(\lim_{n\to\infty}A_n\right)$$

となり，$P(A_n^c)=1-P(A_n)$ より定理を得る．
さて，一般の事象の列 $\{A_n;\ n=1,2,\ldots\}$ が極限をもつならば

$$\lim_{n\to\infty}A_n=\liminf_{n\to\infty}A_n=\bigcup_{n=1}^{\infty}\bigcap_{j=n}^{\infty}A_j=\lim_{n\to\infty}\bigcap_{j=n}^{\infty}A_j$$

となる．ここで $\{\cap_{j=n}^{\infty}A_j;\ n=1,2,\}$ は増加列であるので

$$P\left(\lim_{n\to\infty}A_n\right)=\lim_{n\to\infty}P\left(\bigcap_{j=n}^{\infty}A_j\right)\leq\liminf_{n\to\infty}P(A_n)$$

である．同様に $\lim_{n\to\infty}A_n=\limsup_{n\to\infty}A_n$ を考えると，$\{\cup_{j=n}^{\infty}A_j;\ n=1,2,\}$ は減少列であるので

$$P\left(\lim_{n\to\infty}A_n\right)=\lim_{n\to\infty}P\left(\bigcup_{j=n}^{\infty}A_j\right)\geq\limsup_{n\to\infty}P(A_n)$$

となり，定理が得られる．

定理 3.2 (ボレル-カンテリの定理) 事象の列 $\{A_n;\ n=1,2,\ldots\}$ に対して

1. $\sum_{n=1}^{\infty}P(A_n)<\infty$ ならば $P(\limsup_{n\to\infty}A_n)=0$
2. $A_1,A_2,\ldots$ が互いに独立で $\sum_{n=1}^{\infty}P(A_n)=\infty$ ならば $P(\limsup_{n\to\infty}A_n)=1$

が成立する．これらをボレル-カンテリの定理(Borel-Cantelli Theorem)という． □

1. $0\leq P(\limsup_{n\to\infty}A_n)=P(\lim_{n\to\infty}\cup_{k=n}^{\infty}A_k)$ であり $\cup_{k=n}^{\infty}A_k$ は単調な列なので $P(\lim_{n\to\infty}\cup_{k=n}^{\infty}A_k)=\lim_{n\to\infty}P(\cup_{k=n}^{\infty}A_k)$ となる．さらに $\lim_{n\to\infty}P(\cup_{k=n}^{\infty}A_k)\leq\lim_{n\to\infty}\sum_{k=n}^{\infty}P(A_k)=0$．ただし，最後の等号は仮定 $\sum_{n=1}^{\infty}P(A_n)<\infty$ より従う．

2. $A_1, A_2, \ldots$ が互いに独立ならば，$A_1^c, A_2^c, \ldots$ も互いに独立である．よって $1 - x \leq e^{-x}$ $(x \geq 0)$ を用いると，$m \geq n$ なる m, n に対して $P(\cap_{k=n}^m A_k^c) = \prod_{k=n}^m P(A_k^c) = \prod_{k=n}^m (1 - P(A_k)) \leq \exp(-\sum_{k=n}^m P(A_k))$．$m \to \infty$ とすると条件より右辺は 0 へ収束する．よって $0 = \lim_{n\to\infty} P(\cap_{k=n}^\infty A_k^c)$ であり，$\cap_{k=n}^\infty A_k^c$ は単調な列なので $0 = P(\lim_{n\to\infty} \cap_{k=n}^\infty A_k^c) = P(\liminf_{n\to\infty} A_n^c)$ を得る．すなわち，$P((\liminf_{n\to\infty} A_n^c)^c) = P(\limsup_{n\to\infty} A_n) = 1$.

ボレル-カンテリの定理 1. は $\sum_{n=1}^\infty P(A_n) < \infty$ ならば，無限に多くの A_n に含まれている標本 ω 全体からなる事象は確率 0 をもつという意味であり，ボレル-カンテリの定理 2. は $A_1, A_2, \ldots$ が互いに独立で $\displaystyle\sum_{n=1}^\infty P(A_n) = \infty$ ならば，無限に多くの A_n に含まれている標本 ω 全体からなる事象は確率 1 をもつという意味である．

例 **3.1** 例 1.1 を拡張して，コインを無限に投げ続ける実験を考える．A_n $(n = 1, 2, \ldots)$ を $2n-1$ 回目と $2n$ 回目に続けて表が出る事象とする．このとき事象 $A = \limsup_{n\to\infty} A_n$ は「奇数回目と偶数回目に連続して表が出るということが無限回起こる」という事象になる．また $P(A_n) = 1/4$ なので $\sum_{n=1}^\infty P(A_n) = \infty$ であり，事象 $\{A_n;\ n = 1, 2, \ldots\}$ は互いに独立なので，ボレル-カンテリの定理 2. より $P(A) = 1$ である．さらに事象「2 回連続して表が出るということが無限回起こる」という事象を B とすると $A \subset B$ なので $P(A) \leq P(B)$ より $P(B) = 1$ を得る． □

3.2 確率変数列の概収束と確率収束

1.3 節で述べたように，確率変数は標本空間 Ω から実数 $\mathcal{R}$ への関数である．よって確率変数の収束は，数列の収束とは異なり，様々な定義がある．以下では $X_1(\omega), X_2(\omega), \ldots$ を確率空間 $(\Omega, \mathcal{F}, P)$ で定義された確率変数として，代表的な確率変数の収束の定義を与える．

定義 **3.1** (概収束) 事象 Ω_1 が $P(\Omega_1) = 1$ であるとする．このとき，任意の $\omega \in \Omega_1$ に対して

$$\lim_{n\to\infty} X_n(\omega) = X(\omega) \tag{13}$$

となるような確率変数 $X(\omega)$ が存在するとき，確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は確率変数 X へ概収束 (almost sure convergence) する，あるいは確率 1 で収束 (convergence with probability 1) するといい，$X_n \xrightarrow{a.s.} X$ または $\lim_{n\to\infty} X_n = X$ w.p.1 と書く． □

概収束するということは，全ての $\omega \in \Omega_1$ に対して，各標本 ω 毎に収束するということであり，上で与えた定義は $P(\lim_{n\to\infty} X_n = X) = 1$ と等価である．すなわち，式 (13) を満たす ω 全体が形成する事象 $\{\omega;\ \lim_{n\to\infty} X_n(\omega) = X(\omega)\}$ の確率は 1 である (このような事象を Ω_1 とした)．言い換えれば，もし $\lim_{n\to\infty} X_n(\omega) \neq X(\omega)$ であるような ω が存在するならば，そのような ω は $\omega \in \Omega_0 = \Omega - \Omega_1$ (すなわち ω は $P(\Omega_0) = 0$ であるような事象 Ω_0 の要素) である．

系 **3.1** $X_n \xrightarrow{a.s.} X$ であるということは，全ての $\epsilon > 0$ に対して

$$\lim_{n\to\infty} P(\sup_{m\geq n} |X_m - X| \geq \epsilon) = 0$$

となることと等価である． □

$A_n^{(\epsilon)} = \{\omega \mid |X_n(\omega) - X(\omega)| \geq \epsilon\}$, $A^{(\epsilon)} = \limsup_{n\to\infty} A_n^{(\epsilon)} = \cap_{n=1}^\infty \cup_{k=n}^\infty A_k^{(\epsilon)}$ とする．このとき，$A = \{\omega \mid \lim_{n\to\infty} X_n(\omega) \neq X(\omega)\}$ とすると $A = \cup_{\epsilon>0} A^{(\epsilon)} = \cup_{\epsilon>0} [\cap_{n=1}^\infty \cup_{k=n}^\infty A_k^{(\epsilon)}] = \cup_{m=1}^\infty [\cap_{n=1}^\infty \cup_{k=n}^\infty A_k^{(1/m)}] = \cup_{m=1}^\infty A^{(1/m)}$ である．よって $X_n \xrightarrow{a.s.} X \Leftrightarrow P(A) = P(\cup_{m=1}^\infty A^{(1/m)}) = 0 \Leftrightarrow \forall m \geq 1, P(A^{(1/m)}) = 0 \Leftrightarrow \forall \epsilon > 0, P(A^{(\epsilon)}) = 0$.

定義 **3.2** (確率収束) 任意の正数 ϵ に対して

$$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$$

が成立するとき，確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は確率変数 X へ確率収束 (convergence in probability) するといい，$X_n \xrightarrow{P} X$ と書く． □

定義より $X_n \xrightarrow{P} X$ は $\lim_{n\to\infty} P(|X_n - X| \le \epsilon) = 1$ と等価である．確率収束は，概収束のような関数の収束を意味しているのではなく，実数列 $P(|X_n - X| > \epsilon)$ $(n = 1, 2, \ldots)$ の 0 への収束を意味している．

例 3.2 確率空間 $(R, \mathcal{B}(\mathcal{R}), P)$ において，P は区間 $[0,1]$ 上の一様分布とする．この確率空間上で確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$，$\{Y_n;\ n = 1, 2, \ldots\}$ を以下のように定義する．

$$X_n(\omega) = \begin{cases} 1, & \omega \in [0, \frac{1}{n}] \\ 0, & \text{その他} \end{cases}$$

$$Y_n(\omega) = \begin{cases} 1, & \omega \in \left[\dfrac{n \bmod 2^i}{2^i}, \dfrac{n \bmod 2^i + 1}{2^i}\right] \\ 0, & \text{その他} \end{cases} \qquad (\text{ただし } 2^i \le n < 2^{i+1})$$

ここで $n \bmod m$ は n を m で割ったときの余りを表す．確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は任意の選ばれたある $\omega \in [0,1]$ に対して

$$\lim_{n\to\infty} X_n(\omega) = \begin{cases} 1, & \omega = 0, \\ 0, & \text{その他}, \end{cases} \qquad P(|X_n| \ge \epsilon) = \frac{1}{n} \ \ (0 < \epsilon < 1)$$

である．$P(\{0\}) = 0$ であるので確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は

$$X(\omega) = 0, \qquad \omega \in \mathcal{R}$$

へ概収束する．また $\lim_{n\to\infty} P(|X_n| \ge \epsilon) = 0$ なので $\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$ となり，確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は X へ確率収束する．

一方，任意の選ばれたある $\omega \in [0,1]$ に対して確率変数列 $\{Y_n;\ n = 1, 2, \ldots\}$ は

$$\limsup_{n\to\infty} Y_n(\omega) = 1, \qquad \liminf_{n\to\infty} Y_n(\omega) = 0$$

となるので極限は存在せず，概収束しない．しかし，任意の $0 < \epsilon < 1$ に対して，$2^i \le n < 2^{i+1}$ のとき

$$P(|Y_n - X| > \epsilon) = P(Y_n \ne 0) = \frac{1}{2^i}$$

となるので，$n \to \infty$ のとき $2^i \to \infty$ となることに注意すれば，$\lim_{n\to\infty} P(|Y_n - X| > \epsilon) = 0$ となり，確率変数列 $\{Y_n;\ n = 1, 2, \ldots\}$ は X へ確率収束する． □

この例から分かるように概収束は確率収束よりも強い収束である．

定理 3.3 (概収束 $\Rightarrow$ 確率収束) $X_n \xrightarrow{a.s.} X$ ならば $X_n \xrightarrow{P} X$ である． □

$X_n \xrightarrow{a.s.} X$ ならば，$Y_n(\omega) = 1_{\{|X_n - X| > \epsilon\}}(\omega)$ $(n = 1, 2, \ldots)$ としたとき，$Y_n(\omega)$ は $P(\Omega_1) = 1$ なる事象 Ω_1 上で 0 に収束する．よって $Y_n(\omega)1_{\{\Omega_1\}}(\omega)$ は全ての $\omega \in \Omega$ に対して 0 へ収束する．ここで $0 \le Y_n(\omega)1_{\{\Omega_1\}}(\omega) \le 1$ であるので注意 2.1 で与えた有界収束定理を用いると $\lim_{n\to\infty} E(Y_n 1_{\{\Omega_1\}}) = E(\lim_{n\to\infty} Y_n 1_{\{\Omega_1\}}) = 0$ を得る．一方，$P(\Omega - \Omega_1) = 0$ に注意すると $E(Y_n 1_{\{\Omega - \Omega_1\}}) = E(Y_n \mid \Omega - \Omega_1)P(\Omega - \Omega_1) = 0$ なので $E(Y_n 1_{\{\Omega_1\}}) = E(Y_n 1_{\{\Omega_1\}}) + E(Y_n 1_{\{\Omega - \Omega_1\}}) = E(Y_n) = E(1_{\{|X_n - X| > \epsilon\}}) = P(|X_n - X| > \epsilon)$ となり，$\lim_{n\to\infty} P(|X_n - X| > \epsilon) = 0$ を得る．

概収束，確率収束以外の重要な確率変数収束概念として次の法則収束がある．

定義 3.3 (法則収束) 確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ と確率変数 X があるとする．X_n, X の分布関数をそれぞれ $F_n(x)$, $F(x)$ としたとき，$F(x)$ が連続である全ての x で

$$\lim_{n\to\infty} F_n(x) = F(x)$$

が成立するならば，確率変数列 $\{X_n;\ n = 1, 2, \ldots\}$ は確率変数 X へ法則収束 (convergence in law, convergence in distribution) するといい，$X_n \xrightarrow{D} X$ と書く．また $X_n \xrightarrow{D} X$ であるとき，X の分布 $F(x)$ を X_n の漸近分布 (asymptotic distribution) または極限分布 (limiting distribution) という． □

例 **3.3** (例 3.2 の続き) 例 3.2 で与えた非負確率変数 $X_n(\omega)$ の分布関数 $F_n(x)$ は

$$F_n(x) = \begin{cases} 0, & x < 0 \\ \frac{n-1}{n}, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

であるので，極限分布

$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0 \end{cases}$$

をもつ．すなわち $X_n \xrightarrow{D} X$ である．また，例 3.2 で与えた非負確率変数 $Y_n(\omega)$ の分布関数 $G_n(x)$ は $2^i \le n < 2^{i+1}$ なる n に対して

$$G_n(x) = \begin{cases} 0, & x < 0 \\ 1 - \frac{1}{2^i}, & 0 \le x < 1 \\ 1, & x \ge 1 \end{cases}$$

なので，極限分布

$$F(x) = \begin{cases} 0, & x < 0 \\ 1, & x \ge 0 \end{cases}$$

をもつ．すなわち $Y_n \xrightarrow{D} X$ である． □

一方，法則収束するが確率収束しない例は簡単に作ることが出来る．

例 **3.4** 確率空間 $(R, \mathcal{B}(\mathcal{R}), P)$ において，P は区間 $[0,1]$ 上の一様分布とする．この確率空間上で確率変数列 $\{Z_n;\ n = 1, 2, \ldots\}$ を以下のように定義する．

$$n \text{ が奇数のとき } : Z_n(\omega) = \begin{cases} 1 & \omega \in [0, 1/2] \\ 0, & \text{その他} \end{cases} \qquad n \text{ が偶数のとき } : Z_n(\omega) = \begin{cases} 1 & \omega \in (1/2, 1] \\ 0, & \text{その他} \end{cases}$$

定義より $P(Z_n = 0) = P(Z_n = 1) = 1/2$ であるので，全ての n に対して Z_n は同じ分布をもつ．よって $Z_n \xrightarrow{D} Z_1$ である．しかし n が偶数のとき $P(|Z_n - Z_1| > \epsilon) = 1$ となり，確率収束の意味では収束しない． □

このように法則収束は確率収束よりも弱い収束であり，実際，確率収束するならば法則収束することが知られている．

定理 **3.4** (確率収束 $\Rightarrow$ 法則収束) $X_n \xrightarrow{P} X$ ならば $X_n \xrightarrow{D} X$ である． □

x を $F(x)$ の連続点とする．このとき任意の正数 ϵ に対して $F_n(x) - F(x) = P(X_n \le x) - P(X \le x) = P(X_n \le x, X > x) + P(X_n \le x, X \le x) - P(X \le x) \le P(X_n \le x, X > x) \le P(X_n \le x, X > x + \epsilon) + P(x < X \le x + \epsilon) \le P(|X_n - X| > \epsilon) + F(x + \epsilon) - F(x)$ となり，$\epsilon \to 0$ とすると $\lim_{n\to\infty} F_n(x) - F(x) \le 0$ を得る．同様に $F(x) - F_n(x) = P(X \le x) - P(X_n \le x) = P(X \le x, X_n > x) + P(X \le x, X_n \le x) - P(X_n \le x) \le P(X \le x, X_n > x) \le P(X \le x - \epsilon, X_n > x) + P(x - \epsilon < X \le x) \le P(|X_n - X| > \epsilon) + F(x) - F(x - \epsilon)$ となり，$\epsilon \to 0$ とすると $\lim_{n\to\infty} F(x) - F_n(x) \le 0$ を得る．

法則収束に関しては以下の結果が知られている．

定理 **3.5** (連続定理) $\varphi_n(\eta)$ を確率変数 X_n の特性関数とする．もし $\lim_{n\to\infty} \varphi_n(\eta) = \varphi(\eta)$ となるような $\varphi(\eta)$ が存在し，$\varphi(\eta)$ は $\eta = 0$ の近傍で連続であるならば，$\varphi(\eta)$ はある確率変数 X の特性関数であり，$X_n \xrightarrow{D} X$ が成立する．逆に，$X_n \xrightarrow{D} X$ であるとき，$\varphi_n(\eta),\ \varphi(\eta)$ をそれぞれ，確率変数 $X_n,\ X$ の特性関数とすると $\lim_{n\to\infty} \varphi_n(\eta) = \varphi(\eta)$ が成立する．これを連続定理 (continuity theorem) という． □

証明は [2] を参照．連続定理を用いれば確率変数の分布収束を特性関数を通じて議論することができる．

補題 **3.1** ある定数 c に対して

$$X_n \xrightarrow{D} c \ \Leftrightarrow\ X_n \xrightarrow{P} c$$

である． □

定理 3.4 より $X_n \xrightarrow{P} c$ ならば $X_n \xrightarrow{D} c$ である．そこで $X_n \xrightarrow{D} c \Rightarrow X_n \xrightarrow{P} c$ を示す．X_n の分布関数を $F_n(x)$, c の分布関数を $F(x)$ とする．すなわち $F(x) = 0$ $(x < c)$, $F(x) = 1$ $(x \geq c)$ である．もし $X_n \xrightarrow{D} c$ ならば，全ての正数 ϵ に対して $\lim_{n\to\infty} F_n(c-\epsilon) = F(c-\epsilon) = 0$ であり，$\lim_{n\to\infty} F_n(c+\epsilon) = F(c+\epsilon) = 1$ である．よって $\lim_{n\to\infty} P(|X_n - c| \geq \epsilon) = \lim_{n\to\infty} P(X_n \geq c+\epsilon) + \lim_{n\to\infty} P(X_n \leq c-\epsilon) = 1 - \lim_{n\to\infty} P(X_n < c+\epsilon) + \lim_{n\to\infty} F(c-\epsilon) = 0$.

3.3 大数の法則

多くの回数，同じ条件で確率的実験を繰り返したときに得られる数値の平均は，ある値に近付いていくことが知られている．一般にこのような性質は大数の法則 (law of large numbers) と呼ばれている．大数の法則が成り立つ条件は色々と知られているが，ここでは確率変数列 $\{X_n;\ n = 1, 2, \}$ が互いに独立な場合の結果について紹介する．以下では確率変数列 $\{X_n;\ n = 1, 2, \}$ に対して最初の n 個の確率変数の和を

$$S_n = \sum_{k=1}^{n} X_k$$

で表す．

定理 3.6 (大数の弱法則 (1)) $\{X_n;\ n = 1, 2, \}$ を独立な確率変数列とし，$V(X_n) = \sigma_n^2 < \infty$ であるとする．もし，$\lim_{n\to\infty} \sum_{k=1}^{n} \sigma_k^2/n^2 = 0$ ならば，

$$\frac{S_n - E(S_n)}{n} \xrightarrow{P} 0$$

が成立する．これをチェビシェフの大数の弱法則 (weak law of large numbers) という． □

式 (4) で与えたチェビシェフの不等式を $(S_n - E(S_n))/n$ へ適用すると $P(|S_n - E(S_n)| \geq n\epsilon) \leq \frac{1}{n^2\epsilon^2} \sum_{k=1}^{n} \sigma_k^2$ を得る．よって仮定より，$\lim_{n\to\infty} P(|S_n - E(S_n)|/n \geq \epsilon) = 0$.

分散に関する条件 $\lim_{n\to\infty} \sum_{k=1}^{n} \sigma_k^2/n^2 = 0$ が成立する例として、全ての k に対して $\sigma_k^2 \leq K$ となるような定数 K が存在する場合などがある。

大数の弱法則 (1) は分散が有限であるという制約があった．しかし，この制約は外すことができる．

定理 3.7 (大数の弱法則 (2)) $\{X_n;\ n = 1, 2, \}$ を独立で同一な分布に従う (independent and identically distributed, 略して i.i.d.) 確率変数列とする．このとき，$E(X_1) < \infty$ ならば，$S_n = \sum_{k=1}^{n} X_k$ としたとき

$$\frac{S_n}{n} \xrightarrow{P} E(X_1)$$

が成立する． □

この定理の証明は切断 (truncation) と呼ばれる技法を用いて，無限の分散をもつ確率変数を有限な分散をもつ確率変数へ変換することで行われる．詳細は付録 A.1 を参照すること．

上記の二つの大数の弱法則はいずれも確率収束を与えていた．一方，この収束が概収束の意味で成り立つとき，大数の強法則 (strong law of large numbers) という．

定理 3.8 (大数の強法則) $\{X_n;\ n = 1, 2, \}$ を互いに独立で同一な分布に従う確率変数列とする．もし，$E(X_1) < \infty$ ならば，$S_n = \sum_{k=1}^{n} X_k$ としたとき

$$\frac{S_n}{n} \xrightarrow{a.s.} E(X_1)$$

が成立する． □

証明するには多くの準備を必要とするため省略する (例えば [3] を参照) .

大数の弱法則と大数の強法則の違い，すなわち確率収束と概収束の違いをしっかりと理解する必要がある．系 3.1 より大数の強法則は

$$\lim_{n\to\infty} P\left(\sup_{m\geq n}\left|\frac{S_m}{m}-E(X_1)\right|\geq\epsilon\right)=0$$

と書くことが出来る．ここで事象 $\{\sup_{m\geq n}|S_m/m-E(X_1)|>\epsilon\}$ は $|S_m/m-E(X_1)|$ の値が ϵ より大きくなるような m（$m\geq n$）が少なくとも一つあるということである．よって大数の強法則が示していることは，n が十分に大きければ，n 以上の m で $|S_m/m-E(X_1)|>\epsilon\}$ となってしまうことはほとんどないということである．言い換えれば，このことは $S_1/1, S_2/2, \ldots, S_n/n, \ldots$ という列が n が増加するにつれて，徐々に，それより大きな $m>n$ のいずれの項も平均との差が ϵ より大きくならないようになるということである．一方，大数の弱法則がいうことは n が増加するにつれて，徐々に，S_n/n と平均との差が ϵ より大きくならないようになるということである．このように大数の強法則は列 $\{S_n/n;\ n=1,2,\ldots\}$ の最初の n 個を除いた残り全てに関する性質に言及しているのに対し，大数の弱法則は項 S_n/n それ自身に対する性質を述べている．

例 **3.5** 例 1.1 を拡張して，コインを無限に投げ続ける実験を考える．X_n $(n=1,2,\ldots)$ を n 回目に表が出れば 1，裏が出れば 0 の値を取る確率変数とする．このとき S_n/n は確率 1 で $1/2$ へ収束する．一方，S_n/n が $1/2$ でないような標本は簡単に作ることが出来る．例えば全て表が出る，全て裏が出る，全体の $1/3$ だけ表が出る，などである．大数の強法則は，このような S/n が $1/2$ に収束しないような標本全体からなる事象は確率 0 をもつということを主張している．一方，大数の弱法則は n を大きくすれば事象 $\{|S_n/n-1/2|\geq\epsilon\}$ が起こる確率をいくらでも小さくできるということを主張している． □

3.4 分布の弱収束と中心極限定理

確率変数列の法則収束は，対応する分布関数列の収束と考えることが出来る．このような意味で定義された分布関数の収束を弱収束（weak convergence）といい，$F_n(x)\xrightarrow{D}F(x)$ と書く．法則収束の定義から明らかなように，弱収束では $F(x)$ が連続であるような点でのみ収束が要求されており，$F(x)$ の不連続点での収束は要求されていない．

例 **3.6** 以下で定義される分布関数 $F_n(x)$，$F(x)$ を考える．

$$F_n(x)=\begin{cases}0, & x<0\\ nx, & x\in[0,1/n]\\ 1, & x>1/n\end{cases}\qquad F(x)=\begin{cases}0, & x<0\\ 1, & x\geq 0\end{cases}$$

明らかに $F_n(x)$ は $x=0$ を除いて $F(x)$ へ収束するが，$x=0$ では収束していない．しかし，$x=0$ は $F(x)$ の不連続点であるので，定義より $F_n(x)\xrightarrow{D}F(x)$ である． □

分布関数が弱収束することと等価な条件は幾つか知られており，定理 3.5 はその一つである．これ以外には次のようなものがある．

定理 **3.9** $F_n(x)\xrightarrow{D}F(x)$ は，任意の有界で連続な関数 $g(x)$ に対して

$$\lim_{n\to\infty}\int_{-\infty}^{\infty}g(x)dF_n(x)=\int_{-\infty}^{\infty}g(x)dF(x)$$

が成立することと等価である． □

弱収束を用いて表現された非常に重要な定理に中心極限定理がある．

定理 **3.10** (中心極限定理 (1)) $\{X_n;\ n=1,2,\}$ を互いに独立で同一な分布に従う確率変数列とする．もし，$\mu=E(X_1)<\infty$ かつ $\sigma^2=V(X_1)$ が $0<\sigma^2<\infty$ を満たすならば，$S_n=\sum_{k=1}^n X_k$ としたとき

$$\lim_{n\to\infty}P\left(\frac{S_n-n\mu}{\sqrt{n}\sigma}\leq x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du \tag{14}$$

が成立する．これを中心極限定理 (central limit theorem) という． □

定理 3.10 の証明は付録 A.2 に与える．

中心極限定理 (14) の左辺は平均 0，分散 1 をもつ標準正規分布 $N(0,1)$ の分布関数である．よって

$$\lim_{n\to\infty} P\left(a \le \frac{1}{\sqrt{n}\sigma}\sum_{k=1}^{n}(X_k-\mu) \le b\right) = \int_a^b \frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du$$

と等価である．ガウス (Gauss) は実験から得られたデータの期待値からのバラツキに一定の法則があることを見い出した．それゆえ正規分布はガウス分布とも呼ばれる．

定理 3.10 では X_n が同一な分布に従うことを要求しているが，この制約は以下のように外すことが出来ることが知られている．

定理 3.11 (中心極限定理 (2)) $\{X_n;\ n=1,2,\ldots\}$ を互いに独立な確率変数列とする．X_k $(k=1,2,\ldots)$ の分布関数を $F_k(x)$，平均を $\mu_k<\infty$，分散を σ_k^2 $(0<\sigma_k^2<\infty)$ とし，$B_n^2=\sum_{k=1}^n\sigma_k^2$ とする．もし，任意の正数 ϵ に対して

$$\lim_{n\to\infty}\frac{1}{B_n^2}\sum_{k=1}^{n}\int_{|x-\mu_k|\ge\epsilon B_n}(x-\mu_k)^2dF_k(x)=0 \tag{15}$$

となるならば，$S_n=\sum_{k=1}^n X_k$ としたとき

$$\lim_{n\to\infty} P\left(\frac{S_n-E(S_n)}{B_n}\le x\right)=\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du$$

が成立する． □

式 (15) はリンデベルグ条件 (Lindeberg condition) と呼ばれており

$$\lim_{n\to\infty}\frac{\max\limits_{1\le k\le n}\sigma_k^2}{B_n^2}=0$$

と等価であることが知られている．この式から明らかなように，リンデベルグ条件は n が増加するに連れて X_n の分散が (相対的な意味で) 急激には大きくならないということを要求している．また，この式より X_k が同一の分布に従う場合はリンデベルグ条件を満たしていることが分かる．さらに，全ての X_k が有界で $n\to\infty$ のとき $B_n^2\to\infty$ となるならばリンデベルグ条件が満たされることを簡単に示すことが出来る．

標準正規分布の 3 次中心積率は 0 である．よって，もし確率変数 X_1 の 3 次中心積率の絶対値が非常に大きい，あるいは存在しないならば，その収束は非常にゆっくりしたものになることが予想される．実際，有限の n に対して $(S_n-E(S_n))/B_n$ と極限分布 $N(0,1)$ の差の絶対値に関してはベリーエシーンの定理 (Berry-Esseen's theorem) が知られている ([7] 参照)．特に，独立で同一な分布に従う確率変数列 $\{X_n;\ n=1,2,\}$ に対しては，$0<\sigma^2<\infty$ であり，かつ，$E(|X_1|^3)<\infty$ ならば，n とは独立なある定数 c が存在し，

$$\sup_x\left|P\left(\frac{S_n-n\mu}{\sqrt{n}\sigma}\le x\right)-\int_{-\infty}^{x}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{u^2}{2}\right)du\right|\le\frac{c}{\sqrt{n}}\frac{E(|X_1-\mu|^3)}{\sigma^3}$$

である．これは正規分布による近似の誤差の評価を与えているが，残念ながら定数 c は未知であるので，定量的に誤差を評価するためには利用できない．

この節を終えるにあたり，中心極限定理 (1) の意味するところについて述べる．まず，S_n が平均 $n\mu$, 分散 $n\sigma^2$ をもつことに注意する．中心極限定理 (14) の左辺に現れる $Z_n=(S_n-nE(X_1))/(\sqrt{n}\sigma)$ は式 (5) に従って S_n を標準化したものであり，全ての自然数 n に対して平均 0，分散 1 をもつ確率変数である．中心極限定理はこの確率変数 Z_n の従う分布が，確率変数 S_n の分布，すなわち，確率変数 X_1 の分布が如何なるものであろうとも，有限の平均と分散をもつならば，n を増加させると標準正規分布 $N(0,1)$ に弱収束するということを主張している．すなわち，n を増加させると $(S_n-nE(X_1))/\sqrt{n}$ は平均 0，分散 σ^2 をもつ正規分布に弱収束するということである．

中心極限定理は確率変数 S_n が必ずしも (標準正規分布 $N(0,1)$ の密度関数で近似できるような) 密度関数をもつことを主張しているわけではない．実際，X_1 が離散的確率変数ならば全ての n に対して確率変数 S_n は密度関数をもたない．このような場合，大きな n に対して，その分布関数は非常に細かな階段状の変化をする形状をもち，その形状は積分する (すなわち分布関数を考える) ことで平滑化されることになる．

絶対値が非常に大きい負の値 x に対しては，式 (14) の右辺は非常に 0 に近い値となる．よって n がそれほど大きくなければ，右辺と左辺に現れる確率 $P(Z_n \leq y)$ の差の絶対値は小さいが，その比は 1 から懸け離れたものになっていると思われる．特に非負の確率変数の和を考えると，全ての標本に対して $S_n \geq 0$ であるが，正規分布の定義域は $(-\infty, \infty)$ であるので，有限の n に対して $x < -\sqrt{n}\mu/\sigma$ では明らかに近似的評価として採用できない．また同様に，非常に大きい正数 x に対しては右辺と左辺に現れる確率 $P(Z_n \leq y)$ の両方とも 1 に非常に近く，この場合も右辺は左辺に現れる確率の良い近似とはなっていないと思われる．すなわち左辺に現れる確率は適当な大きさの n に対しては，y が 0 に比較的近い場合にのみ良い近似と見なせると考えられ，この定理が極限定理ではなく，中心極限定理と呼ばれている理由もそこにある．

前節で紹介した大数の法則や中心極限定理では独立な確率変数列を仮定していた．次章で示すように，大数の法則は独立でない確率変数列に対しても，一定の条件下で成立する．この意味で，大数の法則の適用範囲は (中心極限定理に比べて) 相当広い．一方，独立でないような確率変数列に対する中心極限定理はほとんど知られていない．すなわち，中心極限定理においては独立性が非常に重要な仮定となっており，中心極限定理を応用で用いる際には，この点に十分注意を払う必要がある．

問 **3.1** 表 ($X_n = 1$)，裏 ($X_n = -1$) がそれぞれ $1/2$ の確率で出るコインを n 回投げる実験を乱数の初期値を変えて 100 回計算機上で行え．そして，この実験から得られたデータを基に，S_n ($n = 4, 10, 50$)，S_n/n ($n = 4, 10, 50, 100, 200$)，$S_n/\sqrt{n}$ ($n = 4, 10, 50, 100, 200$) の累積度数分布関数を計算し，それぞれ 1 つのグラフに表示せよ． □

問 **3.2** 例 1.1 を拡張して，コインを 100 万回投げるとする．表が 51 万回以上出る確率を複数の手法で評価し，評価結果の差を議論せよ．

> (チェビシェフの不等式 (4)，シャーノフの不等式 (8) ならびに中心極限定理 (14) が利用できる．標準正規分布 $N(0,1)$ に従う確率変数 X に対して $P(|X| > 20)$ はおよそ 2.5×10^{-89} .)

□

4 確率過程の基礎

4.1 確率過程とは

最初に以下のような例を考える．

例 **4.1** ある交差点を通過する客を乗せていない，すなわち空車のタクシーの到着を考える．簡単化のため，連続する空車のタクシーの通過時刻の間隔は独立で同一の分布に従うと仮定する． □

ここで $X(t)$ を時刻 t に交差点へやってきた人が次の空車のタクシーを捕まえるまでに待たなければならない時刻とする．$X(t)$ は時間と共に変化するため，時刻 t の関数であるが，時刻 t を固定してもその値は一意には定まらず，確率変数となっている．すなわち $X(t)$ は標本 ω の関数であり，$X(t, \omega)$ と表すことが出来る．

定義 **4.1** (確率過程) 時刻を表すパラメタ t をもつ確率変数列 $\{X(t);\ t \in \mathcal{T}\}$ を確率過程 (stochastic process) という．ここで $\mathcal{T}$ は時刻 t の取り得る値の集合である．さらに標本 $\omega \in \Omega$ を固定したときの時刻 t の関数 $X(t, \omega)$ を標本路 (sample path) あるいは標本関数 (sample function) という． □

例 4.1 において，$\mathcal{T} = [0, \infty)$ とし，$T_n = T_n(\omega)$ ($n = 1, 2, \ldots$) を n 番目に空車のタクシーが交差点を通過する時刻とする．時刻 T_n の標本が与えられる (すなわち，ある ω が固定される) と $X(t)$ の標本路 $X(t, \omega)$ が一意に定まる．$X(t)$ は正である間は傾き -1 で減少し，タクシーの通過時刻 T_n において $X(T_n) = 0$ となり，その次の瞬間，$\lim_{\epsilon \to 0+} X(T_n + \epsilon) = X(T_n+) = T_{n+1} - T_n$ へジャンプする ($T_1 = 4, T_2 = 2, T_3 = 5, T_4 = 3$ の場合に標本路を書いてみよ)．

また，T_n におけるジャンプの大きさを $Y_n = X(T_n+) = T_{n+1} - T_n$ とすれば，確率変数列 $\{Y_n;\ n \in \mathcal{T} = \{1, 2, \ldots\}\}$ も確率過程と見なすことが出来る．$X(t)$ のように時刻を表すパラメタが連続的な場合，連続時間型確率過程 (continuous-time stochastic process) といい，Y_n のように時刻を表すパラメタが離散的な場合，離散時間型確率過程 (discrete-time stochastic process) という．

確率過程に対する考察では，長時間に渡る大まかな特性を把握するため，長い間，標本路を観察したときの平均的特性を考えることが多い．

定義 **4.2** (時間平均と事象平均) 連続時間確率過程 $\{X(t);\ t \geq 0\}$ に対して

$$\overline{X} = \lim_{t \to \infty} \frac{1}{t} \int_0^t X(s) ds$$

が存在するならば，$\overline{X}$ を確率過程 $\{X(t);\ t \geq 0\}$ の時間平均 (time-average) という．また，$N = \{T_1, T_2, \ldots\}$ を単調増加な時刻の列としたとき

$$\overline{X}_N = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X(T_k)$$

が存在すれば，$\overline{X}_N$ を確率過程 $\{X(t);\ t \geq 0\}$ の N に関する事象平均 (event-average) という．なお，一般に T_n は確定値ではなく (単調増加な) 確率変数列である．

同様に，離散時間確率過程 $\{X_n;\ n = 1, 2, \ldots\}$ に対して

$$\overline{X} = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_k$$

が存在するならば，$\overline{X}$ を確率過程 $\{X_n;\ n = 1, 2, \ldots\}$ の時間平均という．また，$N = \{T_1, T_2, \ldots\}$ を単調増加な時刻の列としたとき

$$\overline{X}_N = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X_{T_k}$$

が存在すれば，$\overline{X}_N$ を確率過程 $\{X_n;\ n = 1, 2, \ldots\}$ の N に関する事象平均という． □

確率過程を考えるとき，時間平均と事象平均をしっかりと区別することは重要である．時間平均は $X(t)$ の標本路を長時間観察したときの平均に対応している．一方，事象平均は，$X(t)$ の標本路をある特定の事象が起こった時点でのみ観察したときの平均である．このように一つの標本路に沿って観測した結果から得られる平均を一般に標本平均 (sample mean) という．これに対して，時間を表すパラメタを固定したときの期待値 $E(X(t))$ や $E(X_n)$ を空間平均 (ensemble mean) という．

4.2 定常過程

確率過程が全ての時刻において同じ分布に従うとき，確率過程は定常状態 (steady state) にあるという．確率過程の多くは時間が経過するに従って，定常状態に近付くことが知られている．ここでは，そのような定常状態にある確率過程について幾つかの結果を離散時間型と連続時間型に分けて紹介する．

定義 **4.3** (離散時間型定常過程) 確率空間 $(\Omega, \mathcal{F}, P)$ 上で定義された離散時間型確率過程 $\{X_n; n = 1, 2, \ldots\}$ において，任意の自然数 n，m に対して $(X_1, X_2, \ldots, X_n)$ と $(X_{m+1}, X_{m+2}, \ldots, X_{m+n})$ が同じ結合分布をもつ場合，確率過程 $\{X_n;\ n = 1, 2, \ldots\}$ は定常過程 (stationary process) と呼ばれる． □

定義において $n=1$ とすると，離散時間型定常過程は全ての自然数 m に対して X_1 と X_m は同じ分布に従うことが分かる．よって，離散時間型確率過程 $\{X_n;\ n=1,2,\ldots\}$ が定常ならば，

$$E\left(\sum_{k=m}^{n} X_k\right) = (n-m+1)E(X_1)$$

である．しかし，X_n は独立であるとは限らないことに注意する．

例 **4.2** $\Omega=\{\omega_1,\omega_2\}$, $P(\{\omega_1\})=P(\{\omega_2\})=1/2$ であるような離散型確率空間を考える．ここで任意の整数 n に対して

$$X_n(\omega) = \begin{cases} 1, & \omega=\omega_1 \\ -1, & \omega=\omega_2 \end{cases}$$

によって離散時間型確率過程 $\{X_n;\ n=1,2,\ldots\}$ を定義する．この確率過程は任意の自然数 $n,\ m$ に対して $P(X_{m+1}=1, X_{m+2}=1,\ldots,X_{m+n}=1) = P(X_{m+1}=-1, X_{m+2}=-1,\ldots,X_{m+n}=-1) = 1/2$ となるので定常過程である．しかし，$P(X_1=1,X_2=1,\ldots,X_n=1)=1/2 \neq (1/2)^n = P(X_1=1)P(X_2=1)\cdots\cdots P(X_n=1)$ であるので $X_1,X_2,\ldots,X_n$ は互いに独立ではない． □

$E(X_1)<\infty$ であるような定常過程 $\{X_n;\ n=1,2,\ldots\}$ に対して，標本平均

$$Y_n = \frac{X_1+X_2+\cdots+X_n}{n}$$

を考える．Y_n の期待値 $E(Y_n)$ は全ての n に対して $E(X_1)$ に等しい．一方，$\lim_{n\to\infty} Y_n$ は概収束することが知られているが，その極限は必ずしも $E(X_1)$ とはならない．

例 **4.3** (例 4.2 の続き) $E(X_1)=0$ であるが，任意の自然数 n に対して $Y_n(\omega)$ は $\omega=\omega_1$ ならば 1，$\omega=\omega_2$ ならば -1 となり，$E(X_1)=0$ へは収束しない． □

確率 1 で $\lim_{n\to\infty} Y_n = E(X_1)$ となるとき，定常過程 $\{X_n;\ n$ は整数 $\}$ はエルゴード的 (ergodic) と呼ばれる．すなわち，定常過程がエルゴード的であるということは，時間平均が標本路に依存しないということである．特に独立で同一な確率変数列はエルゴード的であることに注意する．

最初に述べたように，多くの確率過程は時間が十分に経過すると定常状態に近付く．よって，そのような確率過程が定常状態においてエルゴード的であるということは，

(i) 時間平均の極限が確率 1 で存在し，かつ，確率 1 で (標本路に依存しない) 定数となり，
(ii) 空間平均の極限 $\lim_{n\to\infty} E(X_n)$ が確率 1 で存在し，
(iii) 時間平均と空間平均の極限が一致する，

という三つの性質を満たしているいうことである．

さらに，定常でエルゴード的な離散時間型確率過程 $\{X_n;\ n=1,2,\ldots\}$ と実数から実数への可測関数 $f(x)$ で定義された確率過程 $\{f(X_n);\ n=1,2,\ldots\}$ は定常でエルゴード的であることが知られている．

例 4.2 では定常性は最初から仮定されていたが，性質 (i) が満たされていなかった．エルゴード的でない確率過程の中で，性質 (i), (ii) が満たされているが (iii) が満たされていないような例には次のようなものがある．

例 **4.4** 任意の自然数 n に対して $P(X_n=0)=P(X_n=2)=1/2$ であるような，独立で同一な分布に従う確率変数列 $\{X_n;\ n=1,2,\ldots\}$ を用いて，離散時間型確率過程 $\{M_n;\ n=1,2,\ldots\}$ を $M_n=X_1X_2\cdots X_n$ で定義する．定義より，任意の自然数 n に対して M_n は $X_1,\ldots,X_n$ 全てが値 2 を取ったときのみ正となり，それ以外の場合は値 0 を取る．すなわち，$P(M_n=2^n)=(1/2)^n$，$P(M_n=0)=1-(1/2)^n$ である．また，確率 1 で $\lim_{n\to\infty} M_n=0$ であり，ある n に対して $M_n=0$ とならば全ての $m>n$ に対して $M_m=0$ となる．よって，時間平均 Y_n の極限は存在し，確率 1 で

$$\lim_{n\to\infty} Y_n = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} M_k = 0$$

となる．一方，空間平均 $E(M_n)$ は $E(M_n) = 2^n \times (1/2)^n = 1$ であるので，空間平均の極限 $\lim_{n\to\infty} E(M_n)$ も存在し，1 である．このように離散時間型確率過程 $\{M_n;\ n = 1, 2, \ldots\}$ では時間平均の極限が存在し，確率 1 で標本路に依存しない定数となり，空間平均の極限も存在するが，両者は一致しない． □

連続時間型確率過程 $\{X(t);\ t \geq 0\}$ に対しても同様に定常過程を定義することが出来る．

定義 4.4 (連続時間型定常過程) 連続時間型確率過程 $\{X(t);\ t \geq 0\}$ が任意の正数 s と任意の自然数ならびに任意の単調増加列 $0 \leq t_1 < t_2 < \cdots < t_m$ に対して，$(X(t_1), \ldots X(t_m))$ と $(X(s+t_1), \ldots X(s+t_m))$ が同じ結合分布をもつとき，定常過程と呼ばれる． □

連続時間型確率過程 $\{X(t);\ t \geq 0\}$ が定常ならば

$$E\left(\int_s^t X(\tau)d\tau\right) = (t-s)E(X(0))$$

が成立する．さらに，連続時間型定常過程 $\{X(t);\ t \geq 0\}$ において

$$\lim_{t\to\infty} \frac{1}{t}\int_0^\infty X(t)dt = E(X(0)) \qquad \text{w.p.1}$$

となるならば，$\{X(t);\ t \geq 0\}$ はエルゴード的と呼ばれる．

また，定常でエルゴード的な連続時間型確率過程 $\{X(t);\ t \geq 0\}$ と実数から実数への可測関数 $f(x)$ で定義された確率過程 $\{f(X(t));\ t \geq 0\}$ は定常でエルゴード的であることが知られている．

エルゴード的な定常過程は応用上，非常に重要である．例えば，確率的な挙動をするシステムを観察あるいは測定する場合，そこからは有限の時間区間で観察された (その極限は時間平均となる) 標本平均しか手に入れることが出来ない．よって，観測結果から空間平均に関する性質を議論するためには，観察している確率過程がエルゴード的な定常過程である必要がある．一方，時間平均に比べて，空間平均は格段に解析的な扱いが容易である．よって，空間平均を解析的に求めることで，時間平均を得ることが可能となる．

例 4.5 パケット長が独立で同一な分布に従い，平均パケット長が有限である十分に多くのパケットを，伝送速度が時間と共に変化する回線を用いて，連続して送信する状況を考える．n 番目のパケットの長さを L_n (bit) とする．また，時刻 t での伝送速度を $\lambda(t)$ (bit/sec) とする．すなわち，時刻 s から時刻 $t > s$ までに送信できる総ビット数は $\int_s^t \lambda(\tau)d\tau$ (bit) で与えられる．D_n を n 番目に送信されたパケットの伝送時間とする．$\lambda(t)$ がエルゴード的な定常過程であり，平均伝送率 $\overline{\lambda} = E(\lambda(0))$ が $0 < \overline{\lambda} < \infty$ であるとき，$\{D_n;\ n = 1, 2, \ldots\}$ は定常でエルゴード的な離散時間型確率過程となる．

以下では，例 4.5 において 1 つのパケットを伝送するのに必要な平均送信時間 $\overline{D}$ を求める．$\overline{D}$ は本来，離散時間型確率過程 $\{D_n;\ n = 1, 2, \ldots\}$ の時間平均として定義されるものであるが，エルゴード性により空間平均と等しくなるので，まず，空間平均を考える．T_n を n 番目のパケットの伝送終了時間とする．n 番目のパケットは時刻 T_{n-1} に送信が開始されるので

$$L_n = \int_{T_{n-1}}^{T_n} \lambda(\tau)d\tau$$

が成立する．さらに $T_n = T_{n-1} + D_n$ であり，$\lambda(t)$ は定常過程であるので

$$P(L_n \leq x) = P\left(\int_{T_{n-1}}^{T_{n-1}+D_n} \lambda(\tau)d\tau \leq x\right) = P\left(\int_0^{D_n} \lambda(\tau)d\tau \leq x\right)$$

を得る．よって

$$E(L_n) = E\left(\int_0^{D_n} \lambda(\tau)d\tau\right) = E\left(E\left(\int_0^{D_n} \lambda(\tau)d\tau \;\middle|\; D_n\right)\right) = E\left(D_n E(\lambda(0))\right) = E(D_n)\overline{\lambda}$$

となるので，$E(L_n) = E(L_1)$ に注意すると，$E(D_n) = E(D_1) = E(L_1)/\overline{\lambda}$ を得る．さらに，D_n はエルゴード的な定常過程であるので，

$$\overline{D} = \lim_{n\to\infty} \frac{1}{n}\sum_{k=1}^{n} D_n = \frac{E(L_1)}{\overline{\lambda}} \qquad \text{w.p.1}$$

となる．□

4.3 計数過程とポワソン過程

特定の事象が時間間隔 $(0,t]$ の間に起こった回数を表す確率過程を考える．ただし，有限時間内には高々有限回しか注目する事象は起こらないと仮定する．

定義 **4.5** (計数過程) $0 \le t < \infty$ で定義された $N(t)$ が

1. $N(0) = 0$
2. $N(t)$ は非負整数値を取り，t に対して単調非減少
3. $N(t)$ は右連続，すなわち，全ての t $(0 \le t < \infty)$ に対して $\lim_{\epsilon \to 0+} N(t+\epsilon) = N(t)$

であるとき，連続時間型確率過程 $\{N(t);\ t \ge 0\}$ を計数過程 (counting process) という．□

特に事象の起こる時間間隔が独立で同一な分布に従う場合，計数過程は再生過程 (renewal process) と呼ばれる．定義より，再生過程においては $N(t)$ の不連続点間の長さは独立で同一な確率変数となっている．

例 4.1 において，$N(t)$ を時刻 t までに通過した空車のタクシーの数とすると $N(t)$ は再生過程となる．$N(t)$ の n 番目の不連続点を T_n とする．例 4.1 では n 番目の空車のタクシーが通過した時刻に対応する．$T_0 = 0$ とし，$Y_n = T_{n+1} - T_n$ で不連続点間の長さ, すなわち注目する事象が生起する時間間隔を表す．$E(Y_1) = m_1 > 0$ とすると，大数の強法則より

$$\lim_{n\to\infty} \frac{T_n(\omega)}{n} = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} Y_k(\omega) = m_1 \qquad \text{w.p.1}$$

が成立する．また

$$\lim_{t\to\infty} N(t,\omega) = \infty \qquad \text{w.p.1}$$

であるので，これらが同時に起こる標本 ω からなる事象 B は確率 1 をもつ．事象 B に含まれる任意の標本 ω に対して上の二つの式が成立するので

$$\lim_{t\to\infty} \frac{T_{N(t)}}{N(t)} = m_1 \qquad \text{w.p.1} \tag{16}$$

を得る．式 (16) は大数の強法則に現れる n を確率変数 $N(t)$ で置き換えたものになっている．

また，

$$\lim_{t\to\infty} \frac{T_{N(t)}}{T_{N(t)+1}} = \lim_{t\to\infty} \frac{T_{N(t)}}{N(t)} \cdot \frac{N(t)+1}{T_{N(t)+1}} \cdot \frac{N(t)}{N(t)+1}$$

なので，

$$\lim_{t\to\infty} \frac{T_{N(t)}}{T_{N(t)+1}} = 1 \qquad \text{w.p.1}$$

を得る．さらに $T_{N(t)} \le t \le T_{N(t)+1}$ なので，$T_{N(t)}/T_{N(t)+1} \le T_{N(t)}/t \le 1$ となり，

$$\lim_{t\to\infty} \frac{T_{N(t)}}{t} = 1 \qquad \text{w.p.1} \tag{17}$$

を得る．よって $N(t)/t = (T_{N(t)}/t)(N(t)/T_{N(t)})$ とし，式 (16), (17) を用いると次の定理を得る．

定理 **4.1** (初等再生定理) 再生過程 $\{N(t); t \ge 1\}$ において $E(Y_1) = m_1 > 0$ ならば

$$\lim_{t\to\infty} \frac{N(t)}{t} = \frac{1}{m_1} \qquad \text{w.p.1}$$

が成立する．これを初等再生定理 (elementary renewal theorem) という．□

さて，例 4.1 において，タクシーの通過間隔の平均を $E(Y_1) = m_1 > 0$，2 次積率を $E(Y_1^2) = m_2$ として，次の空車のタクシーが通過するまでの時間を表す確率過程 $X(t)$ の時間平均と $X(T_n+) = Y_n$ の事象平均を考える．$T_0 = 0$ に注意すると

$$\begin{aligned}
\overline{X} &= \lim_{t\to\infty} \frac{1}{t} \int_0^t X(s)ds \\
&= \lim_{t\to\infty} \frac{1}{t} \left(\sum_{k=1}^{N(t)} \int_{T_{k-1}}^{T_k} X(s)ds + \int_{T_{N(t)}}^{t} X(s)ds \right) \\
&= \lim_{t\to\infty} \frac{N(t)}{t} \cdot \frac{1}{N(t)} \sum_{k=1}^{N(t)} \int_0^{Y_{k-1}} (Y_{k-1} - s)ds + \lim_{t\to\infty} \frac{1}{t} \int_0^{t-T_{N(t)}} (Y_{N(t)} - s)ds \\
&= \lim_{t\to\infty} \frac{N(t)}{t} \cdot \lim_{t\to\infty} \frac{1}{N(t)} \sum_{k=1}^{N(t)} \frac{Y_{k-1}^2}{2} + \lim_{t\to\infty} \frac{1}{t} \int_0^{t-T_{N(t)}} (Y_{N(t)} - s)ds
\end{aligned}$$

となるので，

$$\overline{X} = \frac{m_2}{2m_1} + \lim_{t\to\infty} \frac{1}{t} \int_0^{t-T_{N(t)}} (Y_{N(t)} - s)ds \qquad \text{w.p.1}$$

を得る．ここで $Y_{N(t)}$ は有限なので，式 (17) を用いると

$$\lim_{t\to\infty} \frac{1}{t} \int_0^{t-T_{N(t)}} (Y_{N(t)} - s)ds \le \lim_{t\to\infty} \frac{t - T_{N(t)}}{t} Y_{N(t)} = \lim_{t\to\infty} \left(1 - \frac{T_{N(t)}}{t}\right) Y_{N(t)} = 0 \qquad \text{w.p.1}$$

となるので，最終的に

$$\overline{X} = \frac{m_2}{2m_1} \qquad \text{w.p.1}$$

を得る．

一方，$N = \{T_1+, T_2+, \ldots\}$ に関する事象平均 $\overline{X}_N$ は

$$\overline{X}_N = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{\infty} X(T_k+) = \lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{\infty} Y_k = m_1 \qquad \text{w.p.1}$$

となる．

$\overline{X} = m_2/(2m_1) = m_1/2 + V(Y_1)/(2m_1)$ はランダムに交差点へやってきた人が空車のタクシーを捕まえるまで待つ平均時間と見ることが出来る．一方，$\overline{X}_N = m_1$ は空車のタクシーが通過した直後に交差点へやってきた人が空車のタクシーを捕まえるまで待つ平均時間を表している．$V(Y_1) \ge 0$ なので，もし $V(Y_1) = 0$, すなわち空車のタクシーが等間隔でやってくるならば $\overline{X} = m_1/2$ であるが，通過間隔にバラツキがあると $m_1/2$ よりも大きくなる．もし空車のタクシーの通過間隔が指数分布に従っているならば $V(Y_1) = m_1^2$ なので，$\overline{X} = m_1$ となり，ランダムに交差点へやってきた人の平均待ち時間と空車のタクシーが通過した直後に交差点へやってきた人の平均待ち時間は等しくなる．一方，もし $V(Y_1) > m_1$ ならば，ランダムに交差点へやってきた人の平均待ち時間は空車のタクシーが通過した直後に交差点へやってきた人の平均待ち時間よりも長くなる．このようなことが起こる理由は，ランダムに交差点へやってきた人の到着時刻は長い通過間隔によりはまりやすいからである．

計数過程の重要な例にポワソン過程がある．

定義 4.6 (ポワソン過程) 注目する事象が生起する時間間隔が互いに独立で，同じパラメタ λ の指数分布に従う再生過程を率 λ をもつポワソン過程 (Poisson process) という． □

ここで，事象の生起する時間間隔 X がパラメタ λ の指数分布に従っているとき $X > s$ であるという条件の下での条件付き確率 $P(X \le t + s \mid X > s)$ を考える．定義より

$$P(X \le t + s \mid X > s) = \frac{P(s < X \le t + s)}{P(X > s)} = \frac{P(X \le t + s) - P(X \le s)}{1 - P(X \le s)}$$

となるので $P(X \leq x) = 1 - \exp(-\lambda x)$ を代入すると

$$P(X \leq t+s \mid X > s) = \frac{\exp(-\lambda s) - \exp(-\lambda(t+s))}{\exp(-\lambda s)} = 1 - \exp(-\lambda t)$$

となり，$P(X \leq t+s \mid X > s) = P(X \leq t)$ を得る．これは指数分布の無記憶性 (memoryless property) と呼ばれている．この無記憶性を利用すると，ポワソン過程は以下の性質をもつことを示すことが出来る．

1. 任意の自然数 $n \geq 2$ に対して，互いに交わらない任意の区間 $(u_{k-1}, u_k]$ $(k = 1, \ldots, n,\ 0 \leq u_0 < u_1 < \cdots < u_n)$ における $N(t)$ の増分 $N(u_1) - N(u_0), \ldots, N(u_n) - N(u_{n-1})$ は互いに独立な確率変数となる
2. 任意の $0 \leq s < t$ に対して $N(t) - N(s)$ は平均 $\lambda(t-s)$ をもつポワソン分布に従う

これらの性質より，任意の n と $0 \leq u_0 < u_1 < \cdots < u_n$ に対して，$N(t)$ の増分 $N(s+u_1) - N(s+u_0), \ldots, N(s+u_n) - N(s+u_{n-1})$ の結合分布は全ての $s \geq 0$ に対して同一であることがわかる．この性質をもつ計数過程は一般に定常であるといわれる．すなわち，ポワソン過程は定常である．ポワソン過程の重要性は以下の定理による．

定理 **4.2** (PASTA) $\{X(t);\ t \geq 0\}$ を右連続で左極限をもつ有界な連続時間型確率過程とし，$\{N(t);\ t \geq 0\}$ を有限の率をもつポワソン過程とする．T_n $(n = 1, 2, \ldots)$ を $N(t)$ の n 番目の不連続点とする．さらに $N = \{T_1-, T_2-, \ldots\}$ を不連続点の直前の時点からなる集合とする．もし，この N に関する事象平均

$$\overline{X}_N = \lim_{n \to \infty} \frac{1}{n} \sum_{k=1}^{n} X(T_k-)$$

と時間平均

$$\overline{X} = \lim_{t \to \infty} \frac{1}{t} \int_0^t X(s) ds$$

が存在し，かつ，任意の $t > 0$ に対して $\{X(s) \mid 0 \leq s \leq t\}$ が $\{N(t+u) - N(t) \mid u > 0\}$ と独立であるならば

$$\overline{X}_N = \overline{X} \qquad \text{w.p.1}$$

が成立する．それゆえ $E(\overline{X}_N) = E(\overline{X})$ が成り立つ．この性質を **PASTA** (Poisson Arrivals See Time Averages) という． □

ここで $X(t)$ はポワソン過程 $N(t)$ によって駆動されていても良いことに注意する．特に $X(t)$ が定常であるならば，$\overline{X}_N$ は定常状態分布の平均 $E(X(0))$ と等しい．

例 **4.6** ある病院の待合室に客がポワソン過程に従って到着しているとする．時刻 t において待合室にいる客の数を $X(t)$ とする．このとき $X(t)$ の時間平均は客が到着する直前に関する事象平均，すなわち，到着する客が見る平均客数に等しい． □

例 **4.7** ネットワークの輻輳度を調べるため，あるルータへ向けて試験パケットを送出することを考える．今，試験パケットがこのルータに到着したとき，バッファが一杯でパケット損となった割合を p とすると，p はパケットがルータに到着する直前に関する事象平均と見なすことが出来る．もし，試験パケットのルータへの到着時刻がポワソン過程であるならば，すなわち，到着間隔が独立で同一の指数分布に従うならば，p はルータのバッファが一杯であるとき $X(t) = 1$，そうでないとき $X(t) = 0$ であるような確率過程 $X(t)$ の時間平均に等しい． □

しかし，一般には，事象平均は時間平均と異なることに注意する．

例 **4.8** ある受け付けに客が到着する状況を考える．客はちょうど 2 分毎に一人到着し，ちょうど 1 分だけサービスを受けて受け付けから去るとする．このとき受け付けにいる客数の時間平均は $1/2$ であるが，客が到着した直前に関する事象平均は 0 である． □

ポワソン過程は時間平均と事象平均が一致する特殊な確率過程であることを喚起しておく．

問 **4.1** $N(t)$ を率 λ のポワソン過程とする．$N(t)$ の変化が起こる時刻を $\tau_1, \tau_2, \ldots$ と書く．このとき，以下のような確率過程 $\{X(t);\ t \geq 0\}$ を考える．$X(t) \in \{0, 1, 2\}$ であり，$X(t)$ は 時刻 τ_n $(n = 1, 2, \ldots)$ において状態を

一つ増加させようとする．すなわち，時刻 τ_n において，もし，$X(\tau-) \in \{0,1\}$ ならば $X(\tau_n) = X(\tau_n-)+1$ となる．また，$X(\tau_n-) = 2$ ならば状態に変化は起こらず $X(\tau_n) = X(\tau_n-) = 2$ である．一方，$X(t)$ は時間が 1 だけ経過する毎に状態を一つ減少させようとする．すなわち，$n = 1, 2, \ldots$ に対して，もし $X(n-) \in \{1,2\}$ ならば $X(n) = X(n-)-1$ であり，$X(n-) = 0$ ならば状態に変化は起こらず $X(n) = X(n-) = 0$ である．このような確率過程 $\{X(t);\ t \geq 0\}$ を $\lambda = 0.5, 1.0.2.0$ の場合について，計算機上でシミュレーションせよ．さらに時間平均，ポワソン過程で変化があった直前の時点 $N_1 = \{\tau_1-, \tau_2-, \ldots\}$ に関する事象平均，状態が減少しようとする直前 $N_2 = \{1-, 2-, \ldots\}$ に関する事象平均，ならびに $N_3 = \{0.5, 1.5, 2.5, \ldots\}$ に関する事象平均を求め，これらをを比較せよ．[17] □

5 離散時間マルコフ連鎖

離散時間型確率過程 $\{X_n;\ n = 0, 1, \ldots\}$ を考える．以下では，確率変数 X_n の取り得る値は非負の整数値とし，その集合を $\mathcal{S}$ で表す．このとき，離散時間型確率過程 $\{X_n;\ n = 0, 1, \ldots\}$ を $\mathcal{S}$ 上の確率過程という．

定義 5.1 (離散時間マルコフ連鎖) 離散時間型確率過程 $\{X_n;\ n = 0, 1, \ldots\}$ が

$$P(X_{n+1} = j_{n+1} \mid X_k = j_k, k = 0, \ldots, n) = P(X_{n+1} = j_{n+1} \mid X_n = j_n)$$

を満たすとき，離散時間マルコフ連鎖 (discrete-time Markov chain) という． □

離散時間マルコフ連鎖では将来の挙動 X_m $(m = n+1, n+2, \ldots)$ が現在の状態 X_n のみで定まり，それ以前の状態 X_k $(k = 0, \ldots, n-1)$ とは独立である．この性質を マルコフ性 (Markovian property) をいう．

5.1 離散時間マルコフ連鎖の遷移確率

離散時間マルコフ連鎖における確率 $P(X_{n+1} = j \mid X_n = i)$ は状態 i から状態 j への遷移確率 (transition probability) と呼ばれる．以下では，遷移確率 $P(X_{n+1} = j \mid X_n = i)$ が n に依存しないと仮定し，(i, j) 要素が

$$p_{i,j} = P(X_{n+1} = j \mid X_n = i), \qquad i, j \in \mathcal{S}$$

で与えられる行列 $\boldsymbol{P} = (p_{i,j})$ を考える．$\boldsymbol{P}$ は遷移確率行列 (transition probability matrix) と呼ばれる．行列 $\boldsymbol{P}$ の各行は，X_n の値が与えられたという条件の下での X_{n+1} の確率分布を表している．よって 行列 $\boldsymbol{P}$ は各要素が非負であり，かつ，各行の総和が 1 となる正方行列である．さらに (i, j) 要素が

$$p_{i,j}^{[n]} = P(X_n = j \mid X_0 = i)$$

で与えられる正方行列 $\boldsymbol{P}^{[n]} = (p_{i,j}^{[n]})$ を n ステップ遷移確率行列と呼ぶ．特に $\boldsymbol{P}^{[1]} = \boldsymbol{P}$ であることに注意する．マルコフ性より，$n, m = 1, 2, \ldots$ に対して

$$\begin{aligned}
p_{i,j}^{[n+m]} &= P(X_{n+m} = j \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} P(X_{n+m} = j, X_n = k \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} P(X_{n+m} = j \mid X_n = k, X_0 = i) P(X_n = k \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} P(X_{n+m} = j \mid X_n = k) P(X_n = k \mid X_0 = i) \\
&= \sum_{k \in \mathcal{S}} p_{i,k}^{[n]} p_{k,j}^{[m]}
\end{aligned}$$

[17] $\lambda = 0.5$ の場合の理論値は以下の通りである．時間平均 $= 0.37095$, N_2 に関する事象平均 $= 0.60146$, N_3 に関する事象平均 $= 0.37373$.

となる．これを行列を用いて表現すると

$$\boldsymbol{P}^{[n+m]} = \boldsymbol{P}^{[n]}\boldsymbol{P}^{[m]}$$

である．特に $m = n = 1$ とすれば $\boldsymbol{P}^{[2]} = \boldsymbol{P}^2$ となるので，$m = 1$ として帰納法を用いると n ステップ遷移確率行列 $\boldsymbol{P}^{[n]}$ は

$$\boldsymbol{P}^{[n]} = \boldsymbol{P}^n$$

で与えられることが分かる．

ここで時点 n において状態が i である確率を $\pi_i^{[n]} = P(X_n = i)$ $(i \in \mathcal{S})$ とし，X_n の確率分布 $\boldsymbol{\pi}^{[n]}$ を次式で定義する．

$$\boldsymbol{\pi}^{[n]} = (\pi_0^{[n]}, \pi_1^{[n]}, \ldots), \qquad n = 0, 1, \ldots$$

このとき

$$\begin{aligned} P(X_n = j) &= \sum_{i=0}^{\infty} P(X_n = j, X_0 = i) \\ &= \sum_{i=0}^{\infty} P(X_n = j \mid X_0 = i) P(X_0 = i) \end{aligned}$$

であるので $\boldsymbol{\pi}^{[n]} = \boldsymbol{\pi}^{(0)}\boldsymbol{P}^{[n]}$ となり，次の定理を得る．

定理 **5.1** (時点 n での確率分布) 離散時間マルコフ連鎖の時点 n における確率分布 $\boldsymbol{\pi}^{[n]}$ は，初期状態確率分布 $\boldsymbol{\pi}^{(0)}$ ならびに遷移確率行列 $\boldsymbol{P}$ によって，以下のように一意に定まる．

$$\boldsymbol{\pi}^{[n]} = \boldsymbol{\pi}^{(0)}\boldsymbol{P}^n$$

□

5.2 再帰時間と状態の分類

ある時点で離散時間マルコフ連鎖が状態 i にあり，これ以降初めて状態 j に到達するまでの遷移回数 $T_{i,j}$ を状態 i から状態 j への初到達時間 (first passage time) という．特に $i = j$ のとき，$F_{j,j}$ は状態 j の再帰時間 (recurrence time) と呼ばれる．状態 i から状態 j への初到達時間が n である確率 $P(F_{i,j} = n)$ を $f_{i,j}^{[n]}$ とする．すなわち

$$f_{i,j}^{[n]} = P(X_n = j, X_{n-1} \neq j, \ldots, X_1 \neq j \mid X_0 = i), \;\; n = 1, 2, \ldots$$

である．さらに，$f_{i,j}^{[n]}$ の総和を $f_{i,j}$ とする．

$$f_{i,j} = \sum_{n=1}^{\infty} f_{i,j}^{[n]}$$

$f_{i,j}$ は状態 i から出発し，いずれは状態 j に到達する確率である．もし $f_{j,j} = 1$ であるならば状態 j は再帰的 (recurrent) と呼ばれ，$f_{j,j} < 1$ ならば状態 j は過渡的 (transient) と呼ばれる．

まず，状態 i が再帰的であるか過渡的であるかを判定する条件を $p_{j,j}^{[n]}$ を用いて表すことを考える．ここで状態 i から出発するという条件の下で n ステップ目に状態 j にある確率を，最初に状態 j に訪れた時点を m として排反な事象に分けると

$$p_{i,j}^{[n]} = \sum_{m=1}^{n} f_{i,j}^{[m]} p_{j,j}^{[n-m]} \tag{18}$$

を得る．ただし $p_{j,j}^{[0]} = 1$ である．そこで，式 (18) において $i = j$ とし，両辺を n について和をとると

$$\begin{aligned} \sum_{n=1}^{\infty} p_{j,j}^{[n]} &= \sum_{n=1}^{\infty}\sum_{m=1}^{n} f_{j,j}^{[m]} p_{j,j}^{[n-m]} = \sum_{m=1}^{\infty}\sum_{n=m}^{\infty} f_{j,j}^{[m]} p_{j,j}^{[n-m]} = f_{j,j} \sum_{n=0}^{\infty} p_{j,j}^{[n]} \\ &= f_{j,j} + f_{j,j} \sum_{n=1}^{\infty} p_{j,j}^{[n]} \end{aligned}$$

となるので，形式的に

$$\sum_{n=1}^{\infty} p_{j,j}^{[n]} = \frac{f_{j,j}}{1-f_{j,j}}$$

を得る．これより，状態 j は $\sum_{n=1}^{\infty} p_{j,j}^{[n]}$ が発散すれば再帰的であり，有限であれば過渡的であることがわかる．

次に再帰的な状態 j に対して平均再帰時間 $\nu_j = E[F_{j,j}]$ を考える．明らかに

$$\nu_j = \sum_{n=1}^{\infty} n f_{j,j}^{[n]}$$

である．もし，$f_{j,j}=1$ かつ $\nu_j < \infty$ ならば状態 j は正再帰的 (positive recurrent) と呼ばれ，$f_{j,j}=1$ かつ $\nu_j = \infty$ ならば零再帰的 (null recurrent) と呼ばれる．

次に連結 (communicate) という考え方を導入する．まず，ある n ($n=1,2,\dots$) に対して $p_{i,j}^{[n]} > 0$ ならば，状態 j は状態 i から到達可能という．さらに状態 i と j が互いに到達可能である時，状態 i と j は連結しているという．特に，状態 i はそれ自身と連結であるとする．このとき，マルコフ連鎖の全状態の集合 $\mathcal{S}$ の中から一つの状態を選び，それと連結している全ての状態を集めて集合 $\mathcal{S}$ の部分集合 $\mathcal{C}_1$ を作り (これを連結クラスという)，さらに $\mathcal{C}_1$ に含まれていない状態を一つ選び，それと連結している全ての状態を集めて連結クラス $\mathcal{C}_2$ を作る，という手続きを繰り返すと，状態集合 $\mathcal{S}$ は

$$\mathcal{S} = \mathcal{C}_1 \cup \mathcal{C}_2 \cup \cdots$$

のように互いに素な複数個の連結クラスの和集合で書くことができる．状態数が無限の場合は，連結クラスも無限個必要となる場合がある．

定義より，任意に選ばれた連結クラス $\mathcal{C}$ に含まれる二つの状態 i と j に対して，$p_{j,i}^{[n]} > 0$，$p_{i,j}^{[m]} > 0$ なる n，m が存在する．さらに

$$p_{j,i}^{[n]} p_{i,i}^{(l)} p_{i,j}^{[m]} \le p_{j,j}^{[n+l+m]} \tag{19}$$

が成立するので，両辺を l について和をとれば

$$p_{j,i}^{[n]} \left(\sum_{l=1}^{\infty} p_{i,i}^{[l]} \right) p_{i,j}^{[m]} \le \sum_{l=1}^{\infty} p_{j,j}^{[n+l+m]} \le \sum_{l=1}^{\infty} p_{j,j}^{[l]}$$

を得る．これより次のことが分かる．もし，状態 i が再帰的ならば (すなわち $\sum_{l=1}^{\infty} p_{i,i}^{[l]} = \infty$)，同じ連結クラスに属する他の状態 j もまた再帰的である (すなわち $\sum_{l=1}^{\infty} p_{j,j}^{[l]} = \infty$)．逆に状態 j が過渡的ならば (すなわち $\sum_{l=1}^{\infty} p_{j,j}^{[l]} < \infty$)，同じ連結クラスに属する他の状態 i も過渡的である．

さらに，ある連結クラスに属する状態 i が正再帰的 (零再帰的) であるならば，同じ連結クラスに属する他の状態 j もまた正再帰的 (零再帰的) であることを示すことができる．特に再帰的な連結クラスに含まれる状態数が有限の場合，これらの状態は全て正再帰的である．

再帰的な連結クラス $\mathcal{C}$ の任意に選ばれた状態 i に対して，$p_{i,i}^{[l]} > 0$ となる l の最大公約数 d_i を考える．d_i は状態 i の周期と呼ばれる．同じ連結クラスに属する状態は同じ周期を持つことが知られており，$d=1$ の時，連結クラスは非周期 (aperiodic) であると呼ばれ，$d \ge 2$ のとき周期的 (periodic) であると呼ばれる．これは以下のように示すことが出来る．

まず，i とは異なる状態 $j \in \mathcal{C}$ に対して，$p_{j,i}^{[n]} > 0$，$p_{i,j}^{[m]} > 0$ なる n，m を選ぶ．定義より $n+m$ は d_j の倍数である．さて，$p_{i,i}^{[l]} > 0$，すなわち l が d_i の倍数ならば，式 (19) の右辺は正となる．よって，$p_{j,j}^{[n+l+m]} > 0$ であり，$n+l+m$ は d_j の倍数であり，$n+m$ は d_j の倍数なので l も d_j の倍数でなければならない．以上の議論より，$p_{j,i}^{[n]} > 0$，$p_{i,j}^{[m]} > 0$ なる n，m に対して，l が d_i の倍数であることは，$p_{j,j}^{[n+l+m]} > 0$ の十分条件となっていることが分かる．一方，l が d_j の倍数であることは $p_{j,j}^{[n+l+m]} > 0$ の必要条件である．よって，$d_j \le d_i$ を得る．ここで状態 i，j の選び方は任意であるので，この議論を i と j を逆にしても成り立つ．すなわち $d_i \le d_j$ も同様に成立する．これより，$d_i = d$ ($i \in \mathcal{C}$) が成立する．以上をまとめて次の定理を得る．

定理 **5.2** 連結クラス内の各状態は全て正再帰的か，全て零再帰的か，あるいは全て過渡的かのいずれかである．また，ある状態が周期的であるならば，その状態と同じ連結クラスに属する全ての状態は同じ周期をもつ． □

5.3 既約な離散時間マルコフ連鎖の極限確率と定常状態分布

全ての状態 $i \in \mathcal{S}$ が一つの連結クラスに含まれるようなマルコフ連鎖は既約 (irreducible) と呼ばれる．定義より，既約なマルコフ連鎖の任意に選ばれた状態 $i, j \in \mathcal{S}$ に対して，$p_{i,j}^{[n(i,j)]} > 0$ となるような自然数 $n(i,j)$ が存在する．

定理 **5.3** (定常状態分布) 既約で正再帰的な離散時間マルコフ連鎖において

$$\pi_j = \sum_{i \in \mathcal{S}} \pi_i p_{i,j}, \qquad \sum_{j \in \mathcal{S}} \pi_j = 1 \tag{20}$$

を満たす正数 π_j ($j = 0, 1, \ldots$) が唯一定まり，$\pi_j = 1/\nu_j$ で与えられる．π_j を状態 j の定常状態確率 (steady state probability) といい，$\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ を定常状態分布 (steady state distribution) という． □

$\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ としたとき，式 (20) の最初の式は $\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ と等価であることに注意する．定義より，もし $\boldsymbol{\pi}^{[0]} = \boldsymbol{\pi}$ ならば $\boldsymbol{\pi}^{[1]} = \boldsymbol{\pi}\boldsymbol{P} = \boldsymbol{\pi}$ であるので，帰納法により

$$\boldsymbol{\pi}^{[n]} = \boldsymbol{\pi}^{[n-1]}\boldsymbol{P} = \boldsymbol{\pi}, \qquad n = 1, 2, \ldots$$

を得る．すなわち初期状態分布 $\boldsymbol{\pi}^{[0]}$ が定常状態分布 $\boldsymbol{\pi}$ に等しければ，任意に選ばれた時点 n での状態分布 $\boldsymbol{\pi}^{[n]}$ は定常状態分布 $\boldsymbol{\pi}$ に等しい．

定理 5.3 は以下のようにして証明することができる．今，$X_0 = i_0 \in \mathcal{S}$ とし，この後再び状態 i_0 を訪れる時刻を順に $T_{i_0}(n)$ $(n = 1, 2, \ldots)$ とする．さらに，$T_{i_0}(0) = 0$ とすると，マルコフ性から全ての自然数 n に対して $T_{i_0}(n) - T_{i_0}(n-1)$ は独立で同一な分布に従っていることに注意する．すなわち，$N_{i_0}(n)$ を時刻 1 から時刻 n までに状態 i_0 を訪問した回数を表す確率過程とすると，$N_{i_0}(n)$ は再生過程であり，$E(T_{i_0}(n) - T_{i_0}(n-1)) = E(T_{i_0}(1) - T_{i_0}(0)) = E(T_{i_0}(1)) = \nu_{i_0}$ である．

さて，$v_{j|i_0}$ を次式で定義する．

$$v_{j|i_0} = E\left(\sum_{n=1}^{T_{i_0}(1)} 1_{\{X_n = j\}} \mid X_0 = i_0 \right)$$

定義より $v_{j|i_0}$ は時刻 0 で状態 i_0 にあるマルコフ連鎖が時刻 1 以降に再び状態 i_0 を訪れるまでに，状態 j を訪れる平均回数である．連鎖は規約で再帰的であるので $v_{i_0|i_0} = 1$ である，ここで

$$\sum_{j \in \mathcal{S}} \sum_{n=1}^{T_{i_0}(1)} 1_{\{X_n = j\}} = \sum_{n=1}^{T_{i_0}(1)} \sum_{j \in \mathcal{S}} 1_{\{X_n = j\}} = \sum_{n=1}^{T_{i_0}(1)} 1 = T_{i_0}(1)$$

であるので，

$$\sum_{j \in \mathcal{S}} v_{j|i_0} = E(T_{i_0}(1)) = \nu_{i_0}$$

となることに注意する．

上で導入した $v_{j|i_0}$ は

$$\begin{aligned} v_{j|i_0} &= E\left(\sum_{n=1}^{T_{i_0}(1)} 1_{\{X_n = j\}} \mid X_0 = i_0 \right) = E\left(\sum_{n=1}^{\infty} 1_{\{X_n = j, n \le T_{i_0}(1)\}} \mid X_0 = i_0 \right) \\ &= \sum_{n=1}^{\infty} E\left(1_{\{X_n = j, n \le T_{i_0}(1)\}} \mid X_0 = i_0 \right) = \sum_{n=1}^{\infty} P(X_n = j, n \le T_{i_0}(1) \mid X_0 = i_0) \end{aligned}$$

と書き換えることが出来る．ここで，$P(X_1 = j, 1 \leq T_{i_0}(1) \mid X_0 = i_0) = P(X_1 = j \mid X_0 = i_0) = p_{i_0,j}$ であり，$n = 2, 3, \ldots$ に対して

$$P(X_n = j, n \leq T_{i_0}(1) \mid X_0 = i_0) = \sum_{k \neq i_0} P(X_{n-1} = k, n-1 \leq T_{i_0}(1) \mid X_0 = i_0) p_{k,j}$$

が成立するので，これらの式を両辺足し合わせると

$$v_{j|i_0} = p_{i_0,j} + \sum_{k \neq i_0} v_{k|i_0} p_{k.j}$$

を得る．さらに $v_{i_0|i_0} = 1$ であるので

$$v_{j|i_0} = \sum_{k \in \mathcal{S}} v_{k|i_0} p_{k.j} \tag{21}$$

となる．

ここで $\boldsymbol{v} = (v_{0|i_0}, v_{1|i_0}, \ldots)$ とおくと上式は $\boldsymbol{v} = \boldsymbol{v}\boldsymbol{P}$ であるので，任意の自然数 n に対して $\boldsymbol{v} = \boldsymbol{v}\boldsymbol{P}^n = \boldsymbol{v}\boldsymbol{P}^{[n]}$ が成立する．すなわち，$v_{i_0|i_0} = 1$ に注意すると

$$v_{j|i_0} = \sum_{k \in \mathcal{S}} v_{k|i_0} p^{[n]}_{k.j} = p^{[n]}_{i_0,j} + \sum_{k \neq i_0} v_{k|i_0} p^{[n]}_{k.j}$$

が成立する．もし，$v_{j|i_0} = 0$ ならば全ての自然数 n に対して $p^{[n]}_{i_0,j} = 0$ となり，既約であるという仮定に反する．よって $v_{j|i_0} > 0$ である．一方，正再帰であるという仮定から $\sum_{i \in \mathcal{S}} v_{i|i_0} = \nu_{i_0} < \infty$ であるので，全ての $i \in \mathcal{S}$ に対して $v_{i|i_0} < \infty$ である．

さらに，式 (21) の両辺を ν_{i_0} で割ると

$$\frac{v_{j|i_0}}{\nu_{i_0}} = \sum_{k \in \mathcal{S}} \frac{v_{k|i_0}}{\nu_{i_0}} p_{k.j}$$

を得る．よって $\pi_j = v_{j|i_0}/\nu_{i_0}$ とおくと π_j は式 (20) を満たす．

次に式 (20) を満たす $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ が唯一であることを示す．今 $\pi_{i_0} = 1/\nu_{i_0}$ であるので，$j \neq i_0$ なる全ての $j \in \mathcal{S}$ に対して

$$\frac{1}{\nu_j} = \frac{v_{j|i_0}}{\nu_{i_0}}$$

を示せばよい．そこで $N_j(n \mid i_0) = \sum_{k=1}^{n} 1_{\{X_k = j | X_0 = i_0\}}$ を考える．最初に $X_k = j$ となるまでの時間は $F_{i_0,j}$ であり，これ以降の再び $X_k = j$ となるまでの時間は $F_{j,j}$ であり，平均 $1/\nu_j$ をもつ独立で同一な分布に従っている．よって

$$\lim_{n \to \infty} \frac{N_j(n \mid i_0)}{n} = \lim_{n \to \infty} \frac{F_{i_0,j}}{n} + \lim_{n \to \infty} \frac{n - F_{i_0,j}}{n} \frac{N_j(n - F_{i_0.j} \mid j)}{n - F_{i_0,j}}$$

を得る．規約性より $F_{i_0,j}$ は確率 1 で有限である．よって大数の強法則より

$$\lim_{n \to \infty} \frac{N_j(n \mid i_0)}{n} = \lim_{n \to \infty} \frac{N_j(n \mid j)}{n} = \frac{1}{\nu_j} \qquad \text{w.p.1} \tag{22}$$

を得る．

一方，

$$A_{i_0}(n) = \sum_{k=1}^{n} 1_{\{X_k = i_0\}} = N_{i_0}(n \mid i_0)$$

とすると，$T_{i_0}(A_{i_0}(n)) \leq n < T_{i_0}(A_{i_0}(n+1))$ であるので

$$\lim_{n \to \infty} \frac{N_j(T_{i_0}(A_{i_0}(n)) \mid i_0)}{n} \leq \lim_{n \to \infty} \frac{N_j(n \mid i_0)}{n} \leq \lim_{n \to \infty} \frac{N_j(T_{i_0}(A_{i_0}(n+1)) \mid i_0)}{n} \tag{23}$$

である．ここで

$$\lim_{n \to \infty} \frac{N_j(T_{i_0}(A_{i_0}(n)) \mid i_0)}{n} = \lim_{n \to \infty} \frac{A_{i_0}(n)}{n} \cdot \frac{N_j(T_{i_0}(A_{i_0}(n))}{A_{i_0}(n)}$$

となり，定理 4.1 より確率 1 で $\lim_{n\to\infty} A_{i_0}(n)/n = 1/\nu_{i_0}$ となるので

$$\lim_{n\to\infty} \frac{N_j(T_{i_0}(A_{i_0}(n)) \mid i_0)}{n} = \frac{1}{\nu_{i_0}} \cdot \lim_{n\to\infty} \frac{N_j(T_{i_0}(A_{i_0}(n))}{A_{i_0}(n)} \qquad \text{w.p.1}$$

を得る．さらに $Y_n = N_j(T_{i_0}(n)) - N_j(T_{i_0}(n-1))$ とするとマルコフ性より $\{Y_n;\ n = 1, 2, \ldots\}$ は独立で同一な分布に従い，平均 $E(Y_1) = v_{j|i_0}$ をもつ．よって，大数の強法則より

$$\lim_{n\to\infty} \frac{N_j(T_{i_0}(A_{i_0}(n))}{A_{i_0}(n)} = \lim_{n\to\infty} \frac{1}{T_{i_0}(n)} \sum_{k=1}^{T_{i_0}(n)} Y_k = v_{j|i_0} \qquad \text{w.p.1}$$

となるので，

$$\lim_{n\to\infty} \frac{N_j(T_{i_0}(A_{i_0}(n)) \mid i_0)}{n} = \frac{v_{j|i_0}}{\nu_{i_0}} \qquad \text{w.p.1}$$

を得る．$\lim_{n\to\infty} N_j(T_{i_0}(A_{i_0}(n+1)) \mid i_0)/n$ についても同じ議論により

$$\lim_{n\to\infty} \frac{N_j(T_{i_0}(A_{i_0}(n+1)) \mid i_0)}{n} = \frac{v_{j|i_0}}{\nu_{i_0}} \qquad \text{w.p.1}$$

を示すことが出来る．よって式 (23) より

$$\lim_{n\to\infty} \frac{N_j(n \mid i_0)}{n} = \frac{v_{j|i_0}}{\nu_{i_0}} \qquad \text{w.p.1} \tag{24}$$

となる．最後に式 (22) と (24) を組合わせると $1/\nu_j = v_{j|i_0}/\nu_{i_0}$ を得る．

さらに証明の中で示した式 (22) より，次の定理を得る．

定理 5.4 (マルコフ連鎖の定常状態確率と時間平均) 既約で正再帰的なマルコフ連鎖において各状態を訪れる回数の時間平均は定常状態分布に等しく，初期状態分布とは独立である． □

すなわち，既約で正再帰的なマルコフ連鎖はエルゴード的であり，時間平均と定常状態における空間平均が等しいという性質をもち，以下の定理が成立する．

定理 5.5 $\mathcal{S}$ 上の既約で正再帰的なマルコフ連鎖 $\{X_n;\ n = 0, 1, \ldots\}$ の定常状態分布を $\boldsymbol{\pi} = (\pi_0, \pi_1, \ldots)$ としたとき，$\mathcal{S}$ から $\mathcal{R}$ への関数 $g(j)$ が $\sum_{j\in\mathcal{S}} |g(j)|\pi_i < \infty$ を満たすならば

$$\lim_{n\to\infty} \frac{1}{n} \sum_{k=1}^{n} g(X_k) = \sum_{j\in\mathcal{S}} g(j)\pi_j \qquad \text{w.p.1}$$

が成立する． □

一方，既約で非周期的なマルコフ連鎖に対して次式が成立する．

$$\lim_{n\to\infty} p_{i,j}^{[n]} = \lim_{n\to\infty} \pi_j^{[n]} = \pi_j^{[\infty]}$$

上式は，$n \to \infty$ の極限においてマルコフ連鎖がある状態 j にある確率が初期状態分布とは独立であることを示している．確率 $\pi_j^{[\infty]}$ は極限確率 (limiting probability) と呼ばれる．極限確率に関しては次の定理が知られている．

定理 5.6 (マルコフ連鎖の極限確率) 既約なマルコフ連鎖が過渡的あるいは零再帰的である場合，全ての状態 $j \in \mathcal{S}$ に対して $\pi_j^{[\infty]} = 0$ である．また，正再帰的かつ非周期的である場合，極限確率 $\pi_j^{[\infty]}$ は定常状態確率 π_j に等しい．
□

5.4 応用：$M/G/1$ 待ち行列の系内客数分布

離散時間マルコフ連鎖の応用として，待ち行列モデル (queueing model) を取り上げる．待ち行列モデルとは，多くの利用者によって共有されている資源の利用状況を表す確率モデルである．待ち行列モデルでは共有資源へ

客が到着し，共有資源の利用を要求する．客の到着時に共有資源が他の利用者に利用されていなければ，到着した客は直ちに共有資源の利用を開始する．一方，到着時に共有資源が他の客に利用されている場合は，待合室で共有資源が利用できるようになるまで待つ．待ち行列モデルでは，共有資源をサーバと呼んでいる．よって，客をパケット，サーバを回線と見なすと，これは十分大きなバッファをもつルータと見ることが出来る[18]．

ここでは，一つのサーバと容量に制限のない待合室から成るシステムへ，パラメタ λ のポワソン過程に従い客が到着し，それぞれの客がサーバを占有する時間，すなわち，サービス時間は互いに独立で同一な分布関数 $H(x)$ をもつ分布に従うと仮定する．このような待ち行列モデルは $M/G/1$ と呼ばれている[19]．単一サーバ待ち行列では，平均到着時間間隔を $1/\lambda$，平均サービス時間 b としたとき，$\rho = \lambda b$ を利用率という．利用率 ρ はサーバが客をサービス中であるという事象の時間平均に等しいことが知られている．以下では $\rho < 1$ であると仮定する．

定常状態における系内客数分布を求めるため，客の離脱直後に注目する．n 番目の客の離脱直後の系内客数を X_n とし，A_n を n 番目の客のサービス時間の間に新たに到着する客数とする．

図 2 に系内客数の変化を示す．$X_n = 0$ ならば，この後，暫くの間サーバは休止しており，次の客が到着するとサービスを開始する．そして，この客のサービスが終了し離脱が起こった直後の系内客数は，この客のサービス中に新たに到着した客数に等しい．すなわち $X_n = 0$ ならば $X_{n+1} = A_{n+1}$ となる．

一方，$X_n = i$ $(i \geq 1)$ ならば，客の離脱の後，直ちに次のサービスが開始される．このサービスの終了直後における系内客数は，サービス開始時点で既に系内にいた $i-1$ 人の客とサービス中に新たに到着した客数の和で与えられる．すなわち $X_n \geq 1$ ならば $X_{n+1} = X_n - 1 + A_{n+1}$ である．よって

$$X_{n+1} = \max(X_n - 1, 0) + A_{n+1} \tag{25}$$

が成立する．

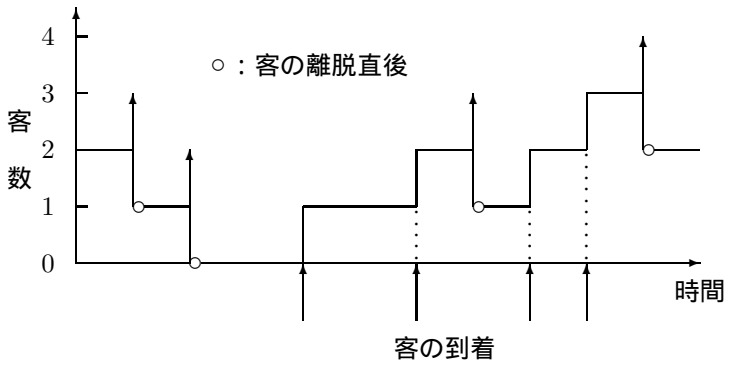

図 2: $M/G/1$ の系内客数の変化

さらに，一人の客のサービス時間の間に k 人の客が到着する確率を $a_k = P(A_n = k)$ とすると，ポワソン到着の仮定より a_k は

$$a_k = \int_0^\infty e^{-\lambda x} \frac{(\lambda x)^k}{k!} dH(x), \qquad k = 0, 1, \ldots \tag{26}$$

で与えられ，n とは独立である．ここで、全ての $k = 0, 1, \ldots$ に対して $a_k > 0$ であることに注意する．

以上の議論から，$n+1$ 番目の客の離脱直後の系内客数 X_{n+1} の確率分布は n 番目の客の離脱直後の系内客数 X_n のみに依存しており，それ以前の客の離脱時点での系内客数とは独立となる．すなわち，X_n はマルコフ性をもつことがわかる．このように系内客数が注目する時点でマルコフ性をもつとき，これらの時点を隠れマルコフ点 (imbedded Markov point) といい，そのような時点のみに注目して構成されたマルコフ連鎖 X_n を隠れマルコフ連鎖 (imbedded Markov chain) と言う．

[18]待ち行列モデルには待つことができる客数に制限があるものや，複数のサーバがあるものなど，様々な種類がある．例えば [11] を参照．
[19]待ち行列モデルは一般に $A/B/c$ の形で表現される．A は客の到着間隔の分布，B はサービス時間の分布，c はサーバ数を表しており，$M/G/1$ における M は指数分布，G は一般の分布を表している

定常状態における離脱直後の系内客数分布を求めるため，隠れマルコフ連鎖 X_n の遷移確率 $p_{i,j} = P(X_{n+1} = j \mid X_n = i)$ を考える．式 (25) より $X_n = 0$ ならば $A_{n+1} = X_{n+1}$ なので

$$p_{0,j} = a_j, \qquad j = 0, 1, \ldots$$

を得る．また，$X_n = i \geq 1$ ならば $A_{n+1} = X_{n+1} - X_n + 1$ なので

$$p_{i,j} = \begin{cases} a_{j-i+1}, & j = i-1, i, \ldots \\ 0, & j = 0, \ldots, i-2 \end{cases}$$

を得る．以上より，客の離脱直後の系内客数の遷移確率行列 $\boldsymbol{P}$ は

$$\boldsymbol{P} = \begin{pmatrix} a_0 & a_1 & a_2 & a_3 & a_4 & \ldots \\ a_0 & a_1 & a_2 & a_3 & a_4 & \ldots \\ 0 & a_0 & a_1 & a_2 & a_3 & \ldots \\ 0 & 0 & a_0 & a_1 & a_2 & \ldots \\ 0 & 0 & 0 & a_0 & a_1 & \ldots \\ \vdots & \vdots & \vdots & \vdots & \vdots & \ddots \end{pmatrix}$$

で与えられる．状態 0 からは 1 ステップで任意の状態 j に到達可能であり，かつ，任意の状態 j $(j = 1, 2, \ldots)$ からは j ステップで状態 0 に到達可能なので (そのような確率は a_0^j)，マルコフ連鎖 X_n は既約であることが分かる．よって以下では X_n の定常状態確率が存在すると仮定する．

客の離脱直後における系内客数が k 人である定常状態確率を π_k とする．このとき定常状態分布 $\boldsymbol{\pi} = (\pi_j)$ は

$$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}, \qquad \sum_{j=0}^{\infty} \pi_j = 1$$

を満たす．$\boldsymbol{\pi} = \boldsymbol{\pi}\boldsymbol{P}$ を要素毎に書き下せば

$$\pi_j = \pi_0 a_j + \sum_{i=1}^{j+1} \pi_i a_{j+1-i}, \qquad j = 0, 1, \ldots \tag{27}$$

である．

ここで式 (27) の両辺を $j = 0$ から $k-1$ まで加えると

$$\begin{aligned} \sum_{j=0}^{k-1} \pi_j &= \pi_0 \sum_{j=0}^{k-1} a_j + \sum_{j=0}^{k-1} \sum_{i=1}^{j+1} \pi_i a_{j+1-i} \\ &= \pi_0 \sum_{j=0}^{k-1} a_j + \sum_{i=1}^{k} \sum_{j=i-1}^{k-1} \pi_i a_{j+1-i} \\ &= \pi_0 \sum_{j=0}^{k-1} a_j + \sum_{i=1}^{k} \pi_i \sum_{j=0}^{k-i} a_j \end{aligned}$$

を得る．これを π_k について解くと

$$\pi_k a_0 = \pi_0 \left(1 - \sum_{j=0}^{k-1} a_j \right) + \sum_{i=1}^{k-1} \pi_i \left(1 - \sum_{j=0}^{k-i} a_j \right) \tag{28}$$

となる．一回の遷移で高々一つしか系内客数が減らないことに注意すると，式 (28) の左辺は系内客数が k 以上から $k-1$ 以下に減少する確率フローの量と見なすことができ，一方，右辺は系内客数が $k-1$ 以下から k 以上になる確率フローの量と見なすことが出来る．一般に，定常状態では系内客数が k と $k-1$ の間を横切るフローの量は等しくなる．

定常状態において任意に選ばれた時点で系内客数が k 人である確率を p_k とする．π_k は客の離脱直後における系内客数分布であり，一般には p_k とは異なるが，$M/G/1$ においては以下の理由により両者が一致する．

まず，系内客数を時間の関数として眺めると一回の変化で高々 1 しか動かない階段関数である（図 2 参照）．そこで，$A_k(t)$ を時間区間 $(0,t]$ の間に系内客数が k から $k+1$ に増加した時点の数とし，$D_k(t)$ を時間区間 $(0,t]$ の間に系内客数が $k+1$ から k に減少した時点の数とする．このとき $A(t)=\sum_{k=0}^{\infty}A_k(t)$ は時間区間 $(0,t]$ の間に到着する総客数を表し，$D(t)=\sum_{k=0}^{\infty}D_k(t)$ は時間区間 $(0,t]$ の間に離脱した総客数を表すことに注意する．定義より

$$\pi_k=\lim_{t\to\infty}\frac{D_k(t)}{D(t)}\qquad \text{w.p.1}$$

である．ここで $L(t)$ を時刻 t における系内客数とすると $L(t)=L(0)+A(t)-D(t)$ が成立する．よって

$$\lim_{t\to\infty}\frac{D_k(t)}{D(t)}=\lim_{t\to\infty}\frac{A_k(t)}{A(t)}\cdot\frac{A(t)}{A(t)+L(0)-L(t)}+\lim_{t\to\infty}\frac{D_k(t)-A_k(t)}{D(t)}$$

を得る．さらに $L(0)<\infty$，かつ，$L(t)$ が極限分布をもつと仮定すると，$t\to\infty$ の極限では確率 1 で $A(t)/[A(t)+L(0)-L(t)]\to 1$ となり，一方，$|A_k(t)-D_k(t)|\le 1$ であるので確率 1 で $[D_k(t)-A_k(t)]/D(t)\to 0$ となる．すなわち

$$\lim_{t\to\infty}\frac{D_k(t)}{D(t)}=\lim_{t\to\infty}\frac{A_k(t)}{A(t)}\qquad \text{w.p.1}\tag{29}$$

が成立する．式 (29) は客の離脱直後の系内客数分布が客の到着直前の系内客数分布と等しいということを示している．また式 (29) を

$$\lim_{t\to\infty}\frac{D_k(t)}{t}\frac{t}{D(t)}=\lim_{t\to\infty}\frac{A_k(t)}{t}\frac{t}{A(t)}\qquad \text{w.p.1}$$

のように変形し，確率 1 で $\lim_{t\to\infty}A(t)/t=\lim_{t\to\infty}D(t)/t$ であることに注意すると

$$\lim_{t\to\infty}\frac{D_k(t)}{t}=\lim_{t\to\infty}\frac{A_k(t)}{t}\qquad \text{w.p.1}\tag{30}$$

を得る．式 (30) は単位時間当たりに状態 k を見る平均到着客数が離脱直後に状態 k を見る平均離脱客数に等しいことを示している．以上をまとめて次の補題を得る．

補題 5.1 客の到着ならびに離脱が順序性をもつ定常状態にある待ち行列モデルにおいて，客の到着直前の系内客数分布と客の離脱直後の系内客数分布は確率 1 で等しい．さらに，このとき，系内客数が k から $k+1$ へ変化する回数の時間平均は系内客数が $k+1$ から k へ変化する回数の時間平均に確率 1 で等しい． □

一方，客の到着がポワソン過程に従うならば，PASTA の性質より，定常状態において客の到着直前の系内客数分布は定常状態分布に等しい．それゆえ補題 5.1 より，離脱直後の系内客数分布は定常状態分布に等しくなる．特に $\pi_0=p_0$ であり，p_0 はサーバが稼働していない確率なので，$\pi_0=p_0=1-\rho$ を得る．

このように，式 (28) より π_k が順次求まり，離脱直後の系内客数分布 $\{\pi_k;k=0,1,\ldots\}$ は定常状態における系内客数分布 $\{p_k;k=0,1,\ldots\}$ に等しい．以上をまとめて次の定理を得る．

定理 5.7 $\rho<1$ のとき，$M/G/1$ の定常状態における系内客数分布 $\{p_k;k=0,1,\ldots\}$ は次式により計算される．

$$\begin{aligned}p_0&=1-\rho\\ p_k&=\frac{1}{a_0}\left[p_0\left(1-\sum_{j=0}^{k-1}a_j\right)+\sum_{i=1}^{k-1}p_i\left(1-\sum_{j=0}^{k-i}a_j\right)\right],\qquad k=1,2,\ldots\end{aligned}$$

□

図 3 はサービス時間が 1 である $M/D/1$ の系内客数 L の裾野分布（tail distribution）$P(L>k)$ を対数軸で示している．図より，大きな k の値に対しては $P(L>k)$ は一定の率で減少することがわかる．すなわち，十分大きな k に対して系内客数の裾野分布は $d>0$，$0<r<1$ なる定数 d，r を用いて $P(L>k)\sim dr^k$ となっている．

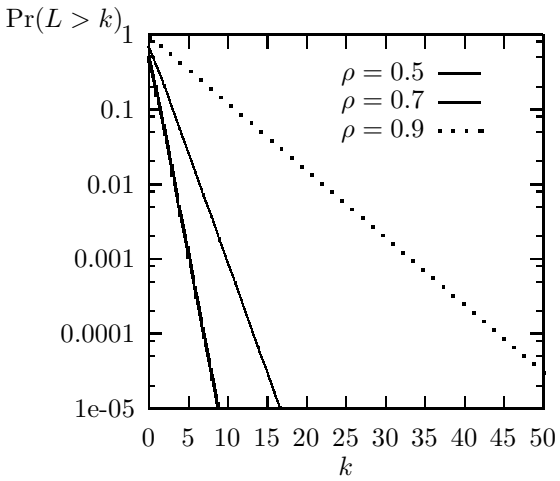

図 3: $M/D/1$ の系内客数の裾野分布 ($b = 1$)

次に系内客数分布の積率を計算するため．以下では $M/G/1$ の客の離脱時点における系内客数分布の確率母関数

$$L^*(z) = \sum_{j=0}^{\infty} \pi_j z^j$$

を導く．$L^*(z)$ は定常状態における系内客数分布の確率母関数に等しいことに注意する．式 (27) の両辺に z^j を掛け，$j = 0$ から無限大まで和を取ると

$$\begin{aligned} L^*(z) &= \pi_0 \sum_{j=0}^{\infty} a_j z^j + \sum_{j=0}^{\infty} \sum_{i=1}^{j+1} \pi_i a_{j+1-i} z^j \\ &= \pi_0 A^*(z) + \frac{1}{z} \sum_{i=1}^{\infty} \pi_i z^i \sum_{j=i-1}^{\infty} a_{j+1-i} z^{j+1-i} \\ &= \pi_0 A^*(z) + \frac{1}{z} \left(L^*(z) - \pi_0 \right) A^*(z) \end{aligned} \tag{31}$$

を得る．ただし $A^*(z)$ は任意に選ばれた客のサービス時間の間に到着する客数の確率分布 $\{a_k; k = 0, 1, \ldots\}$ の確率母関数であり

$$\begin{aligned} A^*(z) &= \sum_{k=0}^{\infty} a_k z^k = \sum_{k=0}^{\infty} \int_0^{\infty} e^{-\lambda x} \frac{(\lambda x)^k}{k!} dH(x) z^k \\ &= \int_0^{\infty} e^{-\lambda x} \sum_{k=0}^{\infty} \frac{(\lambda z x)^k}{k!} dH(x) = \int_0^{\infty} e^{-\lambda(1-z)x} dH(x) \end{aligned}$$

で与えられる．ここで，サービス時間分布のラプラス・スティルチェス変換

$$H^*(s) = \int_0^{\infty} e^{-sx} dH(x)$$

を用いれば

$$A^*(z) = H^*(\lambda - \lambda z)$$

となる．式 (31) を $L^*(z)$ について解くと

$$L^*(z) = \frac{\pi_0 (z-1) H^*(\lambda - \lambda z)}{z - H^*(\lambda - \lambda z)}$$

を得る．右辺に現れる未知数 π_0 は正規化条件 $L(1) = \sum_{j=0}^{\infty} \pi_j = 1$ を用いて定めることができる．すなわち

$$1 = \lim_{z \to 1-} L^*(z) = \frac{\pi_0}{1 - \rho} \tag{32}$$

となるので，$\pi_0 = 1 - \rho$ が得られ $L^*(z)$ が完全に決定される．

定理 5.8 ($M/G/1$ の系内客数分布の母関数) $\rho < 1$ のとき，$M/G/1$ の系内客数分布 $\{p_k; k = 0, 1, \ldots\}$ の確率母関数 $L^*(z)$ は

$$L^*(z) = \frac{(1-\rho)(z-1)H^*(\lambda - \lambda z)}{z - H^*(\lambda - \lambda z)} \tag{33}$$

で与えられる．□

確率母関数の性質より，$L^*(z)$ を微分することで系内客数分布の階乗積率を求めることができる．定常状態における系内客数分布の n 次の階乗積率を $L^{(n)}$ とし，$b^{(n)}$ をサービス時間分布の n 次積率とする．式 (33) の両辺を n 回微分し，$z \to 1$ とすれば，系内客数分布の階乗積率は次の再帰式を満たすことを示すことができる．

$$L^{(n)} = \sum_{k=0}^{n-1} \binom{n+1}{k} \frac{\lambda^{n+1-k} b^{(n+1-k)} L^{(k)}}{(n+1)(1-\rho)} + \lambda^n b^{(n)}, \qquad n = 1, 2, \ldots$$

ただし $L^{(0)} = 1$ とした．特に平均系内客数 $E[L] = L^{(1)}$ は

$$E[L] = \frac{\lambda^2 b^{(2)}}{2(1-\rho)} + \rho$$

で与えられる．図 4 は利用率 ρ に対して平均系内客数を示したものである．平均系内客数 は利用率 ρ のみならず，サービスの変動要因 $b^{(2)}$ にも依存することに注意する．特に ρ が比較的大きい場合，$b^{(2)}$ の影響が顕著になる．また，いずれの場合も ρ が 1 に近付くと平均系内客数が急激に増大することが分かる．

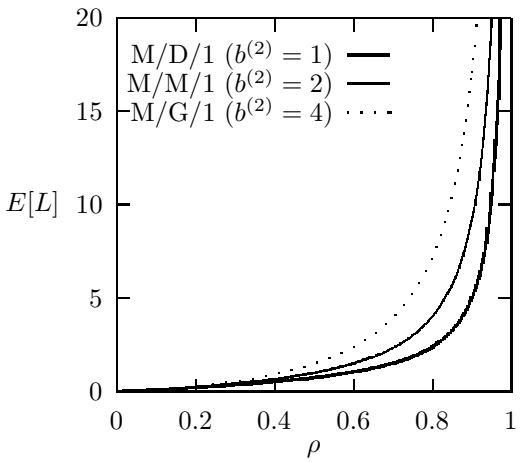

図 4: $M/G/1$ の平均系内客数 ($b = 1$)

6 確率的シミュレーションと統計理論

シミュレーション (simulation) とは実際のシステムの挙動を計算機上で模擬的に再現するということである．確率的要素を含むシステムのシミュレーションを特に確率的シミュレーションということもあるが，ここでは単にシミュレーションという．シミュレーションから得られるデータは，非常に複雑な確率的実験から得られた標本であり，確定値ではない．すなわち，シミュレーションとは計算機を使った複雑な確率過程の標本抽出実験であり，同じシミュレーションモデルあるいはプログラムであっても，異なる乱数列に対する出力結果は一見，全く異なる値をもった列となる．よって，シミュレーションから得られた信じるに足る結果とは様々な理論的背景から得られる手法を正しく適用したものである．言い替えれば，シミュレーションをする際にはどのような手法を適用すべきか，という議論を行う必要がある．以下では，シミュレーションを行う際に必要となる統計理論と実際にシミュレーションを行う際に有益であると思われる幾つかのシミュレーション技法ついて解説する．

6.1 統計的推測と推定量

$\boldsymbol{X} = (X_1, \ldots, X_n)$ をある確率モデルから得られた標本とし，その標本空間を Ω，その結合分布はパラメタ化されたある分布の族 $\mathcal{P} = \{P_\theta;\ \theta \in \Theta\}$ に属しているとする．この標本の実現値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ に基づいて $\boldsymbol{X}$ が従う分布のパラメタ θ の値を推測することを点推定 (point estimate) という．パラメタ θ は例えば X_i の平均や予め与えられた x に対する $P(X > x)$ の値などである．

標本 $\boldsymbol{X} = (X_1, \ldots, X_n)$ の (可測) 関数 $T(\boldsymbol{X})$ を統計量 (statisitic) という．統計量は標本関数と呼ばれることもある．特に，θ の値を推定する点推定において用いられる統計量を θ の推定量 (estimator) といい，標本 $\boldsymbol{X}$ の実現値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ から計算した統計量 $T(\boldsymbol{x})$ の値を θ の推定値 (estimate) という．推定量は確率変数であり，推定値はその実現値であることに注意する．

一般には，パラメタ θ を推定する際，複数の統計量が θ の推定量になり得る場合がある．そのようなとき，どの推定量を用いれば良いか，逆にいうと良い推定量とはどのようなものであるかということが問題になる．推定量 $T(\boldsymbol{X})$ は確率変数であるので，推定量と求めたいパラメタの値 θ との近さを測るのに，$|T(\boldsymbol{X}) - \theta|$ のような距離を直接定義することは出来ない．そこで，通常，全ての θ, t に対して $l(\theta, t) \geq 0$ であり，かつ，全ての θ に対して $t = \theta$ のとき $l(\theta, t) = 0$ となるような関数 $l(\theta, t)$ を考える．関数 $l(\theta, t)$ は推定値 t と θ の距離を表す関数と見ることができ，このような関数を損失関数 (loss function) という．すなわち $t = \theta$ ならば損失は 0 であり，$t \neq \theta$ ならば損失は正となる．さらに推定量 $T(\boldsymbol{X})$ とパラメタの値の差を測るものとして

$$R(\theta, T(\boldsymbol{X})) = E_\theta(l(\theta, T(\boldsymbol{X})))$$

を考える．ここで E_θ はパラメタ θ の下での期待値を表す．この関数 $R(\theta, T(\boldsymbol{X}))$ は危険関数 (risk function) と呼ばれており，$T(\boldsymbol{X})$ を推定量として採用したときの平均損失を表している．そして，この危険関数の値が小さい推定量を良い推定量と考える．

良く使われる危険関数に平均二乗誤差がある．

定義 6.1 (平均二乗誤差) パラメタ θ の推定量 $T(\boldsymbol{X})$ に対して次の危険関数 $\mathrm{MSE}(T(\boldsymbol{X}), \theta)$

$$\mathrm{MSE}(T(\boldsymbol{X}), \theta) = E_\theta((T(\boldsymbol{X}) - \theta)^2)$$

は平均二乗誤差 (mean squared error, MSE) と呼ばれる． □

簡単な計算より，

$$\begin{aligned} E_\theta((T(\boldsymbol{X}) - \theta)^2) &= E_\theta([\{T(\boldsymbol{X}) - E_\theta(T(\boldsymbol{X}))\} + \{E_\theta(T(\boldsymbol{X})) - \theta\}]^2) \\ &= E_\theta(\{T(\boldsymbol{X}) - E_\theta(T(\boldsymbol{X}))\}^2) + \{E_\theta(T(\boldsymbol{X})) - \theta\}^2 \\ &= V_\theta(T(\boldsymbol{X})) + \{E_\theta(T(\boldsymbol{X})) - \theta\}^2 \end{aligned}$$

を得る．ただし V_θ はパラメタ θ の下での分散を表す．すなわち MSE は推定量の平均と分散で定まる危険関数である．ここで現れた $E_\theta(T(\boldsymbol{X})) - \theta$ を偏り (bias) という．

複数の推定量を危険関数を用いて比較するとき，ある推定量 $T(\boldsymbol{X})$ が他の如何なる推定量 $S(\boldsymbol{X})$ よりも良い推定量であるということは，全ての $\theta \in \Theta$ に対して $R(\theta, T(\boldsymbol{X})) \leq R(\theta, S(\boldsymbol{X}))$ であり，少なくとも一つの θ に対して不等号が成立することである．しかし，このような推定量は存在しない．なぜならば，$S(\boldsymbol{X}) = \theta_0 \in \Theta$ とすれば $R(\theta_0, S(\boldsymbol{X})) = 0$ となるので，$R(\theta_0, T(\boldsymbol{X})) = 0$ でなければならない．他の θ についても同様に考えると，最も良い推定量 $T(\boldsymbol{X})$ は全ての $\theta \in \Theta$ で $R(\theta, T(\boldsymbol{X})) = 0$ とならなければならないが，これは一般に不可能である．

そこで，全ての推定量の中で最良のものを考える代わりに，推定量として相応しい適当な性質をもった推定量に限定し，そのような性質を満たす推定量の中で最良のものを探すことがある．相応しい性質の中で，特に良く用いられるのは，推定量の偏り $E_\theta(T(\boldsymbol{X})) - \theta$ が 0 であるという性質である．すなわち，全ての $\theta \in \Theta$ で

$$E_\theta(T(\boldsymbol{X})) = \theta$$

となる推定量に限定して考えることが多い．このような性質をもった推定量 $T(\boldsymbol{X})$ を不偏推定量 (unbiased estimator) という．よって，平均二乗誤差を危険関数として採用した場合，不変推定量の中で最良の推定量を見い出すということは，推定量の分散 $V_\theta(T(\boldsymbol{X}))$ が最小となる不変推定量を求めることと等価である．

ここまでの議論は，推定量となり得る統計量が明らかであることを暗黙の内に仮定していた．しかし，状況が非常に複雑であるため，推定したいパラメタと観察できる標本との対応関係が自明でない場合もある．そのような場合に適用できる推定量に最尤推定量がある．

定義 6.2 (尤度関数と最尤推定量) 標本 $\boldsymbol{X} = (X_1, \ldots, X_n)$ に対して，標本の実現値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ が与えられたとき，結合確率密度 (あるいは結合確率関数) $f_{\boldsymbol{X}}(\boldsymbol{x}, \theta)$ を θ の関数と見なしたものを尤度関数 (likehood function) と呼び，$L(\theta;\ \boldsymbol{x})$ と書く．尤度関数を最大にするパラメタの値 $\hat{\theta} = \hat{\theta}(\boldsymbol{x}) \in \Theta$ を最尤推定値 (maximum likehood estimate) といい，$\hat{\theta} = \hat{\theta}(\boldsymbol{X})$ を θ の最尤推定量 (maximum likehood estimator, MLE) という． □

独立同一な分布に従う標本の場合の尤度関数は $L(\theta;\ \boldsymbol{x}) = \prod_{k=1}^{n} f(x_k, \theta)$ となり，密度関数の積で与えられることに注意する．また，尤度関数の最大化はパラメタ空間 Θ 内で行わなければならない．よって最尤推定量は存在しないこともあるし，複数存在することもある．

尤度関数の対数をとった $\log L(\theta;\ \boldsymbol{x})$ を対数尤度関数 (log likehood function) という．対数関数 $\log x$ は x の増加関数なので，尤度関数を最大にする $\hat{\theta}$ は対数尤度関数も最大にする．特に $\theta = (\theta_1, \ldots, \theta_m)$ が開集合で，それぞれの θ_i に対して尤度関数の 1 階偏導関数が存在するならば，最尤推定値 $\hat{\theta}$ は全ての $i = 1, \ldots, m$ に対して

$$\frac{\partial}{\partial \theta_i} \log L(\theta;\ \boldsymbol{x}) = 0$$

を満たす．これを尤度方程式 (likehood equations) という．

6.2 標本平均と標本分散

前節の予備知識の下で，シミュレーションから得られた注目する指標の出力データ X_i $(i = 1, 2, \ldots)$ からこの指標の平均 b_X を推定することを考える．まず，予め定められた自然数 n に対して標本関数 $T(\boldsymbol{X}) = \overline{X}(n)$ を

$$\overline{X}(n) = \frac{1}{n} \sum_{i=1}^{n} X_i \tag{34}$$

とする．この $\overline{X}(n)$ を標本平均 (sample mean) という．$\overline{X}(n)$ は確定的な値ではなく，確率変数であることに注意する．

例 6.1 サイコロを 10 回振るという試行を行う．X_i $(i = 1, \ldots, 10)$ を i 回目に出た目の数としたとき，標本平均 $\overline{X}(10)$ の値は確定的ではなく，試行毎に異なる． □

また X_i は例 6.1 におけるサイコロの目のようにシミュレーションから直接得られたデータである必要はないことに注意する．

例 6.2 ルータにおけるパケットの待ち時間特性をシミュレーションによって評価することを考える．W_i をシミュレーションにおける i 番目のパケットの待ち時間とする．もし平均待ち時間に興味があれば $X_i = W_i$ である．また，待ち時間が x より大きい確率に興味があれば，定義関数を用いて $X_i = 1_{\{W_i > x\}}$ とすれば良い．一般には，求めたい指標に対応する (可測) 関数 $u(x)$ を用いて $X_i = u(W_i)$ という形で与えられる． □

シミュレーションを行えば X_i の列に対する標本値が得られ，それを用いて標本平均の値を計算することができる．しかし，標本平均は確率変数なので，1 回の実験から得られた標本平均の値が求めたい指標の真の値と一致するわけではない．

最初に X_n が独立同一分布に従う場合を考える．X_i $(i = 1, 2, \ldots)$ を独立同一分布に従う確率変数列とし，X_i は有限の平均 b_X と有限の分散 σ_X^2 をもつとする．ここでの主な目的は，X_i の標本値から b_X の良い推定値を見い出すことである．標本平均 $\overline{X}(n)$ は確率変数なので，最初に $\overline{X}(n)$ の期待値を考える．平均の性質より

$$E\left(\overline{X}(n)\right) = \frac{1}{n}\sum_{i=1}^{n} E(X_i) = b_X$$

となり，$\overline{X}(n)$ の期待値は元の確率変数 X_i の期待値 b_X に等しい．すなわち，$\overline{X}(n)$ は平均 b_X の不偏推定量となっている．

$\overline{X}(n)$ の一つの実現値，すなわち，一回のシミュレーションから得られた結果があるとする．これ以外に何の情報もなければ，この値から真の平均 b_X の値を推定することは本質的に無理がある．なぜなら $\overline{X}(n)$ は分散 $V(\overline{X}(n))$ を持った確率変数であり，$\overline{X}(n)$ の一つの実現値が真の平均 b_X とどの程度近いか，あるいは離れているか判断できないからである．

そこで標本平均の分散 $V(\overline{X}(n))$ を計算すると

$$V(\overline{X}(n)) = V\left(\sum_{i=1}^{n}\frac{X_i}{n}\right) = \frac{1}{n^2}V\left(\sum_{i=1}^{n} X_i\right) = \frac{1}{n^2}\sum_{i=1}^{n} V\left(X_i\right) = \frac{\sigma_X^2}{n} \tag{35}$$

となる．ここで，$\sum$ と V の交換は X_i が独立である時のみ成立することに注意する．式 (35) より，n の値を大きくとれば，分散が小さくなるため $\overline{X}(n)$ の実現値が b_X に近い確率は大きくなると考えられる．しかし，真の分散 σ_X^2 は未知であり，依然として標本平均がどの程度の分散を持っているかは明らかでない．

式 (35) で与えられる標本平均の分散を推定するということは，X_i の分散 σ_X^2 を推定することと等価である．そこで，標本分散 (sample variance) $S_X^2(n)$ を次式で定義する．

$$S_X^2(n) = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \overline{X}(n))^2 \tag{36}$$

$S_X^2(n)$ もまた確率変数であることに注意する．標本分散の期待値は

$$\begin{aligned} E\left(S_X^2(n)\right) &= \frac{1}{n-1}\left(\sum_{i=1}^{n} E(X_i^2) - 2\sum_{i=1}^{n} E(X_i\overline{X}(n)) + nE(\overline{X}^2(n))\right) \\ &= \frac{1}{n-1}\left(nE(X^2) - 2\left\{E(X^2) + (n-1)E(X)^2\right\} + \left\{E(X^2) + (n-1)E(X)^2\right\}\right) \\ &= E(X^2) - E(X)^2 \end{aligned}$$

となり，$E(S_X^2(n)) = \sigma_X^2$ である．よって，$S_X^2(n)$ は σ_X^2 の不偏推定量であることが分かる．この結果，$S_X^2(n)/n$ は標本平均の分散 $V(\overline{X}(n))$ の不変推定量となる[20]．

X_i をシミュレーションの出力とすると，一般には上記の議論をそのまま適用することは出来ない．なぜならば，一般にシミュレーションの出力には相関 (correlation) があるからである．

例 **6.3** FCFS でパケットを転送するルータにおけるパケットの待ち時間の列には正の相関がある．なぜならば，あるパケットの待ち時間が非常に長かったとすると，その次に到着したパケットの待ち時間もまた長くなる傾向にあるからである． □

図 5 は $\rho = 0.8$ の $M/M/1$[21] における n 番目の客の待ち時間 W_n と $n+1$ 番目の客の待ち時間 W_{n+1} の散布図である．図より待ち時間の列 $\{W_n;\, n = 1, 2, \ldots\}$ は強い相関をもつことが分かる．以下に示すように，このような相関のある確率変数列に対しては，前節の独立同一分布に従う場合の結果をそのまま適用することはできない．

[20] 式 (36) で与えられる σ_X^2 の不偏推定量の左辺が $n-1$ で割られたものになっていることに注意する．もし 真の平均 b_X が既知であるならば $\sum_{i=1}^{n}(X_i - b_X)^2/n$ は σ_X^2 の不偏推定量となる．

[21] $M/M/1$ とは，客の到着がポワソン過程に従い，客のサービス時間が指数分布に従う単一サーバ待ち行列である．

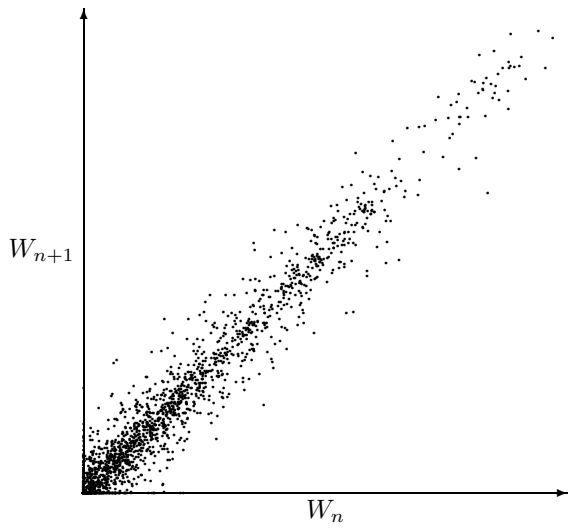

図 5: $M/M/1$ における待ち時間の散布図

相関のある確率変数列 X_i $(i = 1, 2, \ldots)$ を特徴付ける最も簡便な指標は確率変数 X_i と X_j の共分散 $Cov(X_i, X_j)$ である (式 (6) 参照) .

$$Cov(X_i, X_j) = E((X_i - b_i)(X_j - b_j)) = E(X_i X_j) - b_i b_j$$

ここで $b_i = E(X_i)$ である . もし X_i と X_j が独立ならば共分散は 0 となる . また , 両者の間に正の相関がある , すなわち , X_i が平均より大きな値を取るとき , X_j も平均より大きな値を取る傾向がある場合は共分散は正の値を取る . 逆に両者の間に負の相関があれば共分散は負の値を取る .

以下では確率変数列 X_i ($i = 1, 2, \ldots$) は定常で , かつ , 有限の分散をもつと仮定する . このとき , 確率変数列 X_i は以下で定義される共分散定常な確率変数列となる .

定義 6.3 以下の性質が成り立つとき , 確率変数列 X_i ($i = 1, 2, \ldots$) は共分散定常 (covariance stationary) な確率変数列と呼ばれる .

1. $E(X_i) = b_X < \infty$ が全ての i に対して成り立つ .
2. $V(X_i) = \sigma_X^2 < \infty$ が全ての i に対して成り立つ .
3. $Cov(X_i, X_{i+j}) = C_j$ が全ての i , j ($i, j = 1, 2, \ldots$) に対して成り立つ .

□

例えば , 定常な $G/G/1$[22] における待ち時間の列は共分散定常である[23] . これをシミュレーションに置き換えれば , 十分時間が経過した後 , 確率モデルが定常状態に近付けば . それ以降のデータ系列は近似的に共分散定常とみなすことが可能ということである .

さて , 式 (34) で与えられる標本平均 $\overline{X}(n)$ は X_i が共分散定常の場合にも平均値 b_X の不偏推定量である . しかし , 式 (36) で与えられる標本分散 $S_X^2(n)$ は X_i の分散 σ_X^2 の不偏推定量ではない . 実際に計算してみると ,

$$(n-1)E(S_X^2(n)) = (n-1)\sigma_X^2 - 2\sigma_X^2 \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \phi_j \tag{37}$$

となる . ただし $\phi_j = C_j/\sigma_X^2$ は X_i と X_{i+j} の相関関数である (式 (7) 参照) . よって , 標本分散の期待値 $E(S_X^2(n))$ は

$$E(S_X^2(n)) = \sigma_X^2 \left(1 - \frac{2\Phi(n)}{n-1}\right) \tag{38}$$

[22] 一般の単一サーバ待ち行列を $G/G/1$ という .
[23] 正確には , 待ち時間の分散が有限であるという条件が必要である .

となる．ただし $\Phi(n)$ は次式で与えられる．

$$\Phi(n) = \sum_{j=1}^{n-1} \left(1 - \frac{j}{n}\right) \phi_j \leq \frac{n-1}{2} \tag{39}$$

独立同一分布に従う場合は $\phi_j = 0$ なので標本分散の平均 $E[S_X^2(n)]$ は真の分散 σ_X^2 に等しいが，共分散定常の場合，$\phi_j > 0$ ならば $E[S_X^2(n)] < \sigma_X^2$ となる．多くのシミュレーションパッケージでは，$S_X^2(n)$ を用いて出力データの分散を推定しているため，強い相関のあるデータ系列，例えば重負荷時における待ち時間などに対しては，パッケージから得られる分散を用いた結果には，大きな誤差を含んでいることに留意しなければならない．

表 1 は $\rho = 0.99$，平均サービス時間が 1 の $M/M/1$ において 100,000 人分の待ち時間を求める独立なシミュレーションを 10 回行った際の結果である．どのシミュレーションにおいても到着間隔，サービス時間の標本平均は理論値に近いが，個々のシミュレーションから得られる待ち時間の標本平均ならびに標本分散の値には大きなばらつきが見られる．また，標本分散 $S_X^2(n)$ は真の分散 σ_X^2 より小さくなる傾向にある．

表 1: 重負荷時における $M/M/1$ の待ち時間 ($n = 100,000$)

	平均到着間隔	平均サービス時間	待ち時間	
試行 #	$\overline{X}(n)$	$\overline{X}(n)$	$\overline{X}(n)$	$S^2(n)$
1	1.0184	0.9986	79.6614	5553.93
2	1.0035	0.9991	114.8397	5902.72
3	1.0106	0.9997	73.3527	3930.30
4	1.0159	0.9993	54.8584	2243.79
5	1.0125	0.9962	53.4275	2394.71
6	1.0098	0.9976	139.1561	12325.07
7	1.0121	1.0025	72.0533	3179.14
8	1.0081	1.0024	98.0347	4236.82
9	1.0138	0.9946	41.5243	1145.61
10	1.0086	1.0062	119.7758	4454.63
平均	1.0114	0.9997	84.6685	4536.68
理論値	1.0101	1.0000	99.0000	9999.00

共分散定常である確率変数列 X_i に対する標本平均 $\overline{X}(n)$ の分散は，同様の計算により次式で与えられる．

$$V(\overline{X}(n)) = \frac{\sigma_X^2}{n}(1 + 2\Phi(n))$$

よって

$$E\left(\frac{S_X^2(n)}{n}\right) = \frac{\frac{n}{1+\Phi(n)} - 1}{n-1} V(\overline{X}(n))$$

を得る．もし，確率変数列 X_i が正の相関をもつ ($\phi_j > 0$) ならば $\Phi(n) > 0$ となり，$E(S_X^2(n)/n) < V(\overline{X}(n))$ である．すなわち独立同一分布に従う場合の公式を用いて標本平均の分散を推定すると真の値よりも小さな値となる．以上をまとめて次の定理を得る．

定理 6.1 X_i $(i = 1, \ldots, n)$ が同一な分布に従うならば，式 (34) で与えられる標本平均 $\overline{X}(n)$ は真の平均 b_X の不変推定量である．また，X_i $(i = 1, \ldots, n)$ が独立で同一な分布に従うならば，式 (36) で与えられる標本分散は $S_X^2(n)$ は真の分散 σ_X^2 の不変推定量であり，$S_X^2(n)/n$ は標本平均 $\overline{X}(n)$ の分散の不変推定量である．しかし，X_i $(i = 1, \ldots, n)$ が独立でないならば標本分散 $S_X^2(n)$ は真の分散 σ_X^2 の不変推定量ではない． □

6.3 仮説検定と信頼区間

標本平均は確率変数であるので，標本平均の実現値のみを用いて何らかの結論を下すのは心許ない．もし，標本平均が真の平均とどの程度近いかが統計的に評価できれば，それから得られる結論がどの程度信頼できるものかがわかる．この節では，最初に仮説検定の概要について説明し，その後，標本平均の信頼区間について解説する．

確率空間 $(\Omega, \mathcal{F}, P)$ において確率測度 P が未知のパラメタ $\theta \in \Theta$ を含むとする．このとき，パラメタ空間 Θ の部分集合 Θ_0 が与えられたとき，n 個の標本 $\boldsymbol{X} = (X_1, \ldots, X_n)$ の実現値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ に基づいて，$\theta \in \Theta_0$ と $\theta \in \Theta_1 = \Theta \cap \Theta_0^c$ のいずれが真であるかを決定することを統計的仮説検定 (statistic hypothesis testing) という．ここで，パラメタ θ がパラメタ空間の部分集合 Θ_0 に含まれるという仮説を帰無仮説 (null hypothesis) といい，$H_0 :\ \theta \in \Theta_0$ と書く．また，仮説 $H_1 :\ \theta \in \Theta_1$ を対立仮説 (alternative hypothesis) という．特に，部分集合 Θ_0 あるいは Θ_1 が一つの要素からなる場合，それらに対する仮説は単純仮説 (simple hypothesis) と呼ばれる．

例 **6.4** 未知の平均 θ をもつ分布から得られた n 個の標本値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ に基づいた標本平均の実現値 $\theta_0 = (x_1 + \cdots + x_n)/n$ が与えられたとき，ある正数 $\epsilon > 0$ に対して $\Theta_0 = [\theta_0 - \epsilon, \theta_0 + \epsilon]$ とおくと，仮説検定では，与えられたデータ $x_1, \ldots, x_n$ に基づいて H_0 を採択し，真の平均 θ が Θ_0 に含まれるという結論か，あるいは H_1 を採択 (すなわち H_0 を棄却) し，真の平均 θ が $\Theta_1 = (-\infty, \theta_0 - \epsilon) \cup (\theta_0 + \epsilon, \infty)$ に含まれるという結論を出す． □

標本 $\boldsymbol{X} = (X_1, \ldots, X_n)$ がどのような値を取るとき帰無仮説 $H_0 :\ \theta \in \Theta_0$ を採択し，どのような値を取るとき棄却するかという規則を一般に検定 (test) という．特に，パラメタ空間 Θ が実数であり，$H_0 :\ \theta = \theta_0$, $H_1 :\ \theta \neq \theta_0$，すなわち未知のパラメタ θ の値が θ_0 であるという帰無仮説と θ_0 ではないという対立仮説をもつ検定を両側検定 (two-side test) という．検定を行う際には $\boldsymbol{X}$ の標本空間 Ω の部分集合 C を定め，$\boldsymbol{X}$ の実現値 $\boldsymbol{x}$ が C に含まれているならば帰無仮説を棄却し，そうでないならば採択することになる．このような部分集合 C を棄却域 (critical region) という．

検定を行った場合，帰無仮説が真であって，それを採択する，あるいは対立仮説が真 (すなわち帰無仮説が偽) であって，帰無仮説を棄却するというのは正しい決定である．一方，帰無仮説が真であるが，それを棄却する，あるいは対立仮説が真であるが，帰無仮説を採択するという誤りを犯す可能性がある．前者を第 1 種の誤りあるいはタイプ 1 の誤りといい，後者を第 2 種の誤りあるいはタイプ 2 の誤りという．

	$\theta \in \Theta_0$	$\theta \in \Theta_1$
$H_0 :\ \theta \in \Theta_0$ を採択	正しい決定	第 2 種の誤り
$H_1 :\ \theta \in \Theta_1$ を採択	第 1 種の誤り	正しい決定

定義より，第 1 種の誤りを犯す確率は $P_\theta(X \in C \mid \theta \in \Theta_0) = P_\theta(X \in C \mid H_0)$ であり，第 2 種の誤りを犯す確率は $P_\theta(X \notin C \mid \theta \in \Theta_1) = P_\theta(X \notin C \mid H_1)$ である．これらの誤りを犯す確率が共に小さな検定が望ましいが，与えられた標本に対して，同時にこれら両方の確率を小さくすることは出来ない．

そこで，第 1 種の誤りを犯す確率が，ある決められた値 α $(0 < \alpha < 1)$ 以下になるような検定の中で，第 2 種の誤りを犯す確率が最も小さな検定を選ぶことを考える．これを式で書くと

$$\begin{aligned} &\text{目的関数} : P_\theta(X \notin C \mid \theta \in \Theta_1) \longrightarrow \text{最小} \\ &\text{制約条件} : P_\theta(X \in C \mid \theta \in \Theta_0) \leq \alpha \end{aligned}$$

となる．これを達成するような棄却域 C を見い出すことが出来れば，そのような棄却域を用いた検定は良い検定ということができる．

上記のように予め決められた値 α を検定の有意水準 (level of significance) という．よって，有意水準 α の検定は第 1 種の誤りを犯す確率が α 以下であるような検定となり，帰無仮説 H_0 が棄却されたとき，帰無仮説が有意水準 α で棄却されたという．上の議論から明らかなように，第 1 種の誤りを犯す確率は予め設定できるが，第 2 種の誤りを犯す確率を制御することは出来ない．よって，どのような誤りがより重大であるかを判断した上，帰

無仮説と対立仮説を設定する必要がある．

帰無仮説を棄却する確率 $P_\theta(X \in C)$ を $\theta \in \Theta$ の関数

$$\beta(\theta) = P_\theta(X \in C)$$

として捉えたとき，$\beta(\theta)$ ををを検定の検出力関数 (power function) という．さらに，$\theta_1 \in \Theta_1$ であるような θ_1 に対する検出力関数の値 $\beta(\theta_1)$ を θ_1 に対する検定の検出力 (power against θ_1) という．言い換えると，検出力とは帰無仮説が正しくないとき，それを棄却する確率である．一方，$\theta \in \Theta_0$ のとき $\beta(\theta)$ は帰無仮説が正しいのにも関わらず，帰無仮説を棄却する，すなわち，第 1 種の誤りを犯す確率となる．検出力関数 $\beta(\theta)$ の $\theta \in \Theta_0$ における上界

$$\sup_{\theta \in \Theta_0} P_\theta(X \in C) = \sup_{\theta \in \Theta_0} \beta(\theta)$$

は検定の大きさ (size of test) と呼ばれる．

与えられた標本値に対して，帰無仮説が棄却できる最小の有意水準を有意確率 (significance probability) という．ある標本値の有意確率が p^* であるということは，有意水準が p^* より大きい仮説検定では帰無仮説は必ず棄却され，有意水準が p^* より小さな仮説検定では帰無仮説は必ず採択されるということである．

さて，上記とは異なり，α $(0 < \alpha < 1)$ と二つの統計量 $T_1 = T_1(\boldsymbol{X})$，$T_2 = T_2(\boldsymbol{X})$ が全ての $\theta \in \Theta$ に対して

$$P_\theta(T_1 \leq \theta \leq T_2) \geq 1 - \alpha$$

であるとき，区間 $[T_1, T_2]$ を θ の $100(1-\alpha)\%$ 信頼区間(confidence interval)といい，$1-\alpha$ を信頼水準(confidence level) という．θ の $100(1-\alpha)\%$ 信頼区間が $[T_1, T_2]$ であるということは，θ が確率 $1-\alpha$ で区間 $[T_1, T_2]$ に含まれるということではないことに注意する．上の定義において，T_1，T_2 は確率変数である．よって，独立な試行を数多く繰り返し，それぞれの試行に対して信頼区間を求めたとすると，パラメタ θ の値はそれらの信頼区間の内の $100(1-\alpha)\%$ の中には含まれており，$100\alpha\%$ の中には含まれないということである．パラメタ θ は未知の定数であり，実際に分布しているのは区間 $[T_1, T_2]$ であることに注意する．よって，ある標本値 $\boldsymbol{x} = (x_1, \ldots, x_n)$ に基づいて得られた区間 $[T_1(\boldsymbol{x}), T_2(\boldsymbol{x})]$ は θ の値を含んでいるか，含んでいないかのいずれかである．この区間 $[T_1(\boldsymbol{X}), T_2(\boldsymbol{X})]$ は θ の区間推定量 (interval estimator) と呼ばれることもある．

信頼区間と仮説検定は密接な関係があり，実際，仮説検定を基に信頼区間を構築することが可能である．$S(\boldsymbol{X}) = [T_1(\boldsymbol{X}), T_2(\boldsymbol{X})]$ とすると，$\theta = \theta_0$ であるとき，定義関数 $1_{\{\theta \notin S(\boldsymbol{X})\}}$ に対して

$$\beta(\theta_0) = E_{\theta_0}(1_{\{\theta_0 \notin S(\boldsymbol{X})\}}) = 1 - P_{\theta_0}(\theta_0 \in S(\boldsymbol{X})) \leq 1 - (1 - \alpha) = \alpha$$

となり，$\beta(\theta)$ は有意水準 α の検定 $H_0 :\ \theta = \theta_0$ の検定力である．

逆に，$C(\theta_0)$ を帰無仮説 $H_0 :\ \theta = \theta_0$ についての有意水準 α の棄却域とし，$\overline{S}_0(\boldsymbol{x})$ を $\overline{S}_0(\boldsymbol{x}) = \{\theta_0 \mid \boldsymbol{x} \in C(\theta_0)\}$ とすると

$$P_{\theta_0}(\theta_0 \notin \overline{S}_0(\boldsymbol{x})) = P_{\theta_0}(\boldsymbol{X} \in C(\theta_0)) \leq \alpha$$

となる．よって，X の取り得る値全体の集合から $C(\theta_0)$ を除いた残りの領域は θ の信頼水準 $1-\alpha$ の信頼領域 (区間) となる．すなわち，$[T_1, T_2]$ を信頼水準 $1-\alpha$ の信頼区間とするということは，もし $\theta_0 \notin [T_1, T_2]$ ならば帰無仮説 $H_0 :\ \theta = \theta_0$ を水準 α で棄却するということと等価である．

以下では，平均 b，分散 σ^2 をもつ独立同一分布に従う確率変数列 Y_i ($i = 1, \ldots, n$) を考える．以下の議論をシミュレーションの出力に適用するときは，Y_i は i 番目の独立なシミュレーションから得られた標本平均 $\overline{X}(n)$ に対応することに注意する．例えば，表 1 の場合，$n = 10$ であり，Y_i は i 番目のシミュレーションにおける標本平均 $\overline{X}(100,000)$ に対応する．

まず，Y_i の標本平均 $\overline{Y}(n)$ に対して，新しい確率変数 V_n を定義する．

$$V_n = (\overline{Y}(n) - b)/\sqrt{\sigma^2/n} \tag{40}$$

このとき定理 3.10 の中心極限定理より次式が成立する．

$$\lim_{n\to\infty} P(V_n \le z) = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{z} e^{-y^2/2} dy \tag{41}$$

これは，Y_i が独立同一分布に従うならば，十分に大きな n に対して V_n が平均 0，分散 1 をもつ標準正規分布に従うと近似的にみなすことができるということを示している．また，この結果より，十分に大きな n に対して $\overline{Y}(n)$ は平均 b，分散 σ^2/n の正規分布に従うと近似的にみなすことができる．

一般に，分散 σ^2 は未知であるので，この結果を実際に用いるのは容易ではない．しかし，十分大きな n に対して標本分散 $S^2(n)$ は σ^2 に収束するため，式 (41) の σ^2 を $S^2(n)$ で置き換えることができる．よって十分に大きな n に対して

$$\begin{aligned} P\left(\left|\frac{\overline{Y}(n)-b}{\sqrt{S^2(n)/n}}\right| \le z_{1-\alpha/2}\right) &= P\left(\overline{Y}(n) - z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}} \le b \le \overline{Y}(n) + z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}\right) \\ &\approx 1-\alpha \end{aligned} \tag{42}$$

を得る．ここで $z_{1-\alpha/2}$ は標準正規分布の上側 $1-\alpha/2$ の点である．よって，十分大きな n に対して近似的な $100(1-\alpha)\%$ 信頼区間は

$$\overline{Y}(n) \pm z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}} \tag{43}$$

で与えられる．

上でも述べたように，十分に大きい n に対して式 (43) で与えられた信頼区間が意味することは，独立な試行を数多く繰り返し，それぞれの試行に対して信頼区間を求めたとすると，真の平均値 b はそれらの信頼区間の内の $100(1-\alpha)\%$ の中には含まれており，$100\alpha\%$ の中には含まれないということである．

さて，確率変数列 Y_i が正規分布に従うと仮定する．このとき

$$\frac{\overline{Y}(n)-b}{\sqrt{S^2(n)/n}}$$

は自由度 $n-1$ の t 分布に従うことが知られている．さらに，b に対する厳密な $100(1-\alpha)\%$ 信頼区間は

$$\overline{Y}(n) \pm t_{n-1,1-\alpha/2}\sqrt{\frac{S^2(n)}{n}} \tag{44}$$

で与えられる．ただし $t_{n-1,1-\alpha/2}$ は自由度 $n-1$ の t 分布の上側 $1-\alpha/2$ の点である．式 (44) より，標本数 n が 4 倍になると，信頼区間の幅はおよそ半分になることがわかる．

実際には Y_i が正規分布に従うことは極めて稀であり，それゆえ式 (44) も近似的なものとなる．しかし，$t_{n-1,1-\alpha/2} > z_{1-\alpha/2}$ であるので，式 (44) の方が式 (43) よりも真の平均値 b を含みやすい．例えば $t_{24,0.975} \approx 2.064$，$t_{49,0.975} \approx 2.010$ に対して $z_{0.975} \approx 1.960$ である．よって t 分布を用いる方がより安全側の評価となる．

表 2 は $\rho = 0.95$，$\mu = 1.0$ の $M/M/1$ において 50,000 人分の待ち時間を求める独立なシミュレーションを複数回行い，式 (44) を用いて 95% 信頼区間を求めた結果である．試行回数を増やすことで信頼区間の幅が小さくなっていることが観察される．

これらの結果を利用するためには，どの程度の標本数があれば近似的に正規分布とみなせるかということが問題となるが，これには明確な答はない．一般に，分布の対称性を示す尺度である歪度 (skewness)

$$\nu = E((X-b)^3)/\sigma^3$$

の絶対値が大きいほど，収束に多くの標本数を要するといわれている[24]．

[24] p.31 のベリーエシーンの定理を参照．

表 2: 標本数と信頼区間の関係

試行 #	推定値		信頼区間
16	19.4517	±	1.5773
31	20.5451	±	1.2421
51	20.1113	±	1.1829
76	19.7898	±	0.9593
理論値	19.0000		—

6.4 シミュレーションによる定常状態での特性評価

この節ではシミュレーションによってシステムの定常状態における性能指標を評価する方法について解説する．これまで議論してきたように，注目する性能指標に対する標本平均 $\overline{X}(n)$ が一つだけ得られたとしても，そのようなことが起こり得るという以上の情報は得られない．よって，シミュレーション結果から意味ある結論を引き出すためには，独立同一な分布に従う標本平均をなるべく多く集める必要がある．

標本平均を多く集める方法の一つに，システムの挙動が独立なサイクルの繰り返しと見なせるような時点を見い出し，サイクル毎に標本平均を集める再生型シミュレーション (regenerative simulation) がある．一方，このようなサイクルが存在しない場合や，サイクルの長さが非常に長く，結果として多くの標本平均を得ることができない場合には，乱数の初期値を変えて同じシミュレーションプログラムを複数回走らせることで標本平均を集めるか，あるいは非常に長いシミュレーションを行い，そこから近似的に独立同一な分布に従うと見なせる標本平均を複数集めることになる．これらは一般に非停止シミュレーション (non-terminating simulation) と呼ばれている．

以下では，再生型シミュレーションと非停止シミュレーションのそれぞれにおいて，定常状態における性能指標を評価する手法とその際に注意しなければならない点について述べる．

再生型シミュレーション

シミュレーションの出力に対応する離散時間確率過程 $\{X_i; i = 1, 2, \ldots\}$ と関数 $u(x)$ によって定まる性能指標 $Y_i = u(X_i)$ の平均 b_Y を評価することを考える．ただし b_Y は次式を満たすものとする．

$$b_Y = \lim_{N\to\infty} \frac{1}{N} \sum_{i=1}^{N} Y_i \qquad \text{w.p.1}$$

例 6.2 に関数 $u(x)$ の具体例が与えられている．

定義 6.4 以下の条件を満たすとき，離散時間確率過程 $\{X_i;\ i = 1, 2, \ldots\}$ は再生型確率過程(regenerative process)と呼ばれる[25]．

(i) $\{N_{k+1} - N_k;\ k = 1, 2, \ldots\}$ が独立同一分布に従うような増加時点列 $\{N_k;\ k = 1, 2, \ldots\}$ が存在する．

(ii) あらゆる時点列 $0 \le n_1 < n_2 < \cdots < n_m$ ($m \ge 1$) および $l, k \ge 1$ に対して確率ベクトル $\{X_{N_1+n_1}, \ldots, X_{N_1+n_m}, N_2 - N_1, \ldots, N_{l+1} - N_l\}$ と $\{X_{N_{k+1}+n_1}, \ldots, X_{N_{k+1}+n_m}, N_{k+2} - N_{k+1}, \ldots, N_{k+1+l} - N_{k+l}\}$ が同じ分布に従う．

(iii) $\{X_j\,(j = 1, \ldots, N_l - 1), N_{k+1} - N_k\,(1 \le k \le l-1)\}$ と $\{X_{N_l+j}\,(j = 0, 1, \ldots),\ N_{k+1} - N_k\,(k = l, l+1, \ldots)\}$ は独立である．

□

[25]連続時間確率過程 $\{X(t);\ t \ge 0\}$ に対しても同様に再生型確率過程を定義することができるが，ここでは省略する．

概していえば，再生型確率過程 $\{X_i;\, i=1,2,\ldots\}$ とは，(i) N_k で区切られる時点があり，それらの間隔は独立同一分布に従っており，(ii) N_k 以降の X_i の振舞いは N_1 以降の振舞いと確率的に同じであり，かつ，(iii) N_k 以前の X_i の振舞いと N_k 以降の振舞いは独立であるようなものである．このような時点 N_k は再生点と呼ばれ，再生点間は再生サイクルと呼ばれる．

例 **6.5** ルータの振舞いが $GI/G/1$[26] でモデル化され，i 番目に到着したパケットの待ち時間 W_i が X_i に対応するとする．このとき最初の客の待ち時間が 0 ならば，その後，待ち時間が 0 になる客の番号を順次再生点とすれば，待ち時間過程 $\{W_i;\, i=1,2,\ldots\}$ は再生型確率過程となる． □

以下では，X_i の再生点が明確に定義でき，かつ許容範囲内の時間で十分な数の再生サイクルが出現する場合を考える．このとき i 番目の再生サイクルにおいて，注目する性能指標 $Y_k = u(X_k)$ $(k = N_i, \ldots, N_{i+1}-1)$ の累積量 $C_i(Y)$ を次式で定義する．

$$C_i(Y) = \sum_{k=N_i}^{N_{i+1}-1} Y_k, \qquad i = 1,2,\ldots$$

また，i 番目の再生サイクル長 を $\tau_i = N_{i+1} - N_i$ とする．このとき，再生型確率過程の性質より確率変数列 $\{(C_i(Y), \tau_i);\, i=1,2,\ldots\}$ は独立で同一な分布に従う[27]．さらに，緩い仮定の下で $Y_k = u(X_k)$ の時間平均 b_Y は

$$b_Y = \frac{E(C_1(Y))}{E(\tau_1)} \tag{45}$$

で与えられることが知られている．

ここで，二つの確率変数 $C_1(Y)$ と τ_1 には相関があることに注意する．

例 **6.6** (例 6.5 の続き) パケットの伝送が連続的に行われている時間帯を全稼働期間 (busy period) という．$Y_i = u(W_i)$ とすると，全稼働期間の間に伝送されるパケット数が τ_i に対応し，これらのパケットの待ち時間 W_i で定まる Y_i の総和が $C_i(Y)$ に対応する．全稼働期間の間に伝送されるパケットが比較的少ない場合，それらのパケットの待ち時間も比較的短いと予想される．一方，全稼働期間の間に伝送されるパケットが比較的多い場合，輻輳が予想され，これらのパケットの待ち時間も比較的長くなる．すなわち，全稼働期間の間に伝送されるパケット数とそれらのパケットの待ち時間の間には正の相関がある． □

すなわち，再生型シミュレーションでは，式 (45) で与えられる求めたい指標 b_Y が，相関をもつの二つの確率変数 $C_1(Y)$ と τ_1 の期待値の比で与えられる．b_Y の推定量を得るために，予め定められた自然数 n に対して，以下の標本平均 $\overline{C}(n)$，$\overline{\tau}(n)$ を定義する．

$$\overline{C}(n) = \frac{1}{n}\sum_{i=1}^{n} C_i(Y), \qquad \overline{\tau}(n) = \frac{1}{n}\sum_{i=1}^{n} \tau_i$$

$\{(C_i(Y), \tau_i);\ i = 1,2,\ldots\}$ は独立で同一な分布に従う確率変数列であり，標本平均 $\overline{C}(n)$，$\overline{\tau}(n)$ はそれぞれ $E(C_1(Y))$，$E(\tau_1)$ の不変推定量となっている．しかし，$C_i(Y)$ と τ_i の間には相関があるため，b_Y に対する推定量

$$\hat{Y}(n) = \frac{\overline{C}(n)}{\overline{\tau}(n)} \tag{46}$$

には偏りがある．すなわち $E(\hat{Y}(n)) \neq b_Y$ であり，推定量 $\hat{Y}(n)$ は b_Y の不変推定量ではない．その結果，6.3 節で示した結果をそのまま用いることはできない．以下では，この偏りの取り扱いについて議論する[28]．

有限の n に対しては $E(\hat{Y}(n)) \neq b_Y$ であるが，大数の強法則より，推定量 $\hat{Y}(n)$ は確率 1 で $\lim_{n\to\infty} \hat{Y}(n) = b_Y$ を満たす．このように極限において期待値が確率 1 で真の値と一致する推定量を強一致推定量(strongly consistent estimator) という．今の場合，$\hat{Y}(n)$ は b_Y の強一致推定量である．

[26] $GI/G/1$ とは客の到着時間間隔ならびにサービス時間がそれぞれ独立で同一な分布に従う単一サーバ待ち行列を指す．
[27] すなわち，$C_i(Y)$ と τ_i の結合分布 $P(C_i(Y) \le x, \tau_i \le y)$ は i とは独立であり，同じ結合分布に従う．
[28] 以下の議論は，再生シミュレーションのみならず，求めたい指標が相関のある確率変数の期待値の比で与えられる場合に適用できる．

推定量 $\hat{Y}(n)$ の信頼区間を得るため，確率変数列 $\{Z_i;\ i=1,2,\ldots\}$ を

$$Z_i = C_i(Y) - b_Y \tau_i$$

で定義する．Z_i は独立同一分布に従う確率変数列であり，式 (45) より $E(Z_i)=0$ である．また，Z_i の分散 $\sigma_Z^2 = V(Z_i)$ は

$$\sigma_Z^2 = V(C_i(Y)) - 2b_Y Cov(C_i(Y),\tau_i) + b_Y^2 V(\tau_i) \tag{47}$$

で与えられる．また，Z_i の分散 σ_Z^2 の強一致推定量 $S^2(n)$ は

$$S^2(n) = S_{11}(n) - 2\hat{Y}(n)S_{12}(n) + \hat{Y}^2(n)S_{22}(n)$$

で与えられる．ただし，$S_{11}(n)$，$S_{22}(n)$，$S_{12}(n)$ はそれぞれ $V(C_1(Y))$，$V(\tau_1)$，$Cov(C_1(Y),\tau_1)$ に対する不変推定量であり，次式で与えられる．

$$S_{11}(n) = \frac{1}{n-1}\sum_{i=1}^{n}(C_i(Y)-\overline{C}(n))^2, \qquad S_{22}(n) = \frac{1}{n-1}\sum_{i=1}^{n}(\tau_i-\overline{\tau}(n))^2$$

$$S_{12}(n) = \frac{1}{n-1}\sum_{i=1}^{n}(C_i(Y)-\overline{C}(n))(\tau_i-\overline{\tau}(n))$$

さて，中心極限定理より十分大きな n に対して

$$\frac{1}{\sqrt{\sigma_Z^2/n}}\sum_{i=1}^{n}\frac{Z_i}{n}$$

は近似的に標準正規分布に従うとみなせる．さらに

$$\frac{1}{\sqrt{\sigma_Z^2/n}}\sum_{i=1}^{n}\frac{Z_i}{n} = \frac{\sqrt{n}}{\sqrt{\sigma_Z^2}}(\overline{C}(n)-b_Y\overline{\tau}(n)) = \frac{\sqrt{n}}{\sqrt{\sigma_Z^2}}\left(\frac{\overline{C}(n)}{\overline{\tau}(n)}-b_Y\right)\overline{\tau}(n) = \frac{\sqrt{n}\{\hat{Y}(n)-b_Y\}}{\sqrt{\sigma_Z^2/E(\tau_1)}}\cdot\frac{\overline{\tau}(n)}{E(\tau_1)}$$

であるので，

$$\lim_{n\to\infty}\frac{\sqrt{n}\{\hat{Y}(n)-b_Y\}}{\sqrt{\sigma_Z^2/E(\tau_1)}} = \frac{1}{\sqrt{\sigma_Z^2/n}}\sum_{i=1}^{n}\frac{Z_i}{n}\cdot\frac{E(\tau_1)}{\overline{\tau}(n)}$$

より

$$\frac{\sqrt{n}\{\hat{Y}(n)-b_Y\}}{\sqrt{\sigma_Z^2}/E(\tau_1)} \tag{48}$$

は十分大きな n に対して近似的に標準正規分布に従う．

この結果に基づいて信頼区間を求める際には，式 (48) における σ_Z^2，$E(\tau_1)$ に対して，それぞれ強一致推定量 $S^2(n)$ と不変推定量 $\overline{\tau}(n)$ を用いれば良い．最終的に，式 (43) と同様に，十分大きな n に対して近似的な (あるいは漸近的な) $100(1-\alpha)\%$ 信頼区間は

$$\hat{Y}(n) \pm z_{1-\alpha/2}\sqrt{\frac{S^2(n)}{n}}\Big/\overline{\tau}(n) \tag{49}$$

で与えられる．

ここで式 (49) は十分に大きな n についての結果であることに注意する．実際の再生型シミュレーションでは高々有限個のサイクルしか手に入れることはできず，また，状況によってはそれほど多くの再生サイクルを生み出すことが出来ない場合も多い．既に述べたように，有限の n に対しては $E(\hat{Y}(n)) \neq b_Y$ である．実際，$\hat{Y}(n)$ と b_Y の差を考え，$1/n$ について展開することにより次式を得る．

$$E\left(\hat{Y}(n)\right) = b_Y - \frac{1}{n}\frac{E(Z_1\tau_1)}{\{E(\tau_1)\}^2} + \frac{1}{n^2}\frac{E(Z_1\{\tau_1-E(\tau_1)\}^2)}{\{E(\tau_1)\}^3} + O(n^{-3})$$

この偏りは再生サイクルが少ししか手に入れることができなかった場合，特に問題になる．そこで，小さな n に対してこの偏りをより小さくすべく幾つかの提案が行われており，代表的なものにジャックナイフ推定量 $\hat{Y}^*(n)$ がある．

定義 6.5 次式で与えられる $\hat{Y}^*(n)$ をジャックナイフ推定量 (jackknife estimator) という .

$$\hat{Y}^*(n) = n\hat{Y}(n) - \frac{n-1}{n}\sum_{i=1}^{n}\frac{\sum_{k\neq i} C_k(Y)}{\sum_{k\neq i}\tau_k} \tag{50}$$

□

ジャックナイフ推定量 $\hat{Y}^*(n)$ は以下の性質を持つことが知られている .

1. $\hat{Y}^*(n)$ は

$$E(\hat{Y}^*(n)) = b_Y + O(n^{-2})$$

を満たし , b_Y の強一致推定量である .

2. $\hat{S}^2(n)$ を

$$\hat{S}^2(n) = S_{11}(n) - 2\hat{Y}^*(n)S_{12}(n) + \left(\hat{Y}^*(n)\right)^2 S_{22}(n)$$

としたとき , 十分大きな n に対して

$$\frac{\sqrt{n}\left(\hat{Y}^*(n) - b_Y\right)}{\sqrt{\hat{S}^2(n)/n}}$$

は漸近的に標準正規分布に従い , $\hat{Y}^*(n)$ の近似的な (あるいは漸近的な) $100(1-\alpha)\%$ 信頼区間は

$$\hat{Y}^*(n) \pm z_{1-\alpha/2}\sqrt{\frac{\hat{S}^2(n)}{n}}$$

で与えられる .

表 3 は $\rho = 0.99$, 平均サービス時間が 1 の $M/M/1$ において , 10 個の全稼働期間から推定した平均待ち時間である .

表 3: ジャックナイフ推定量

理論値	Jackknife	式 (46)
99.0000	98.5419	67.3741

この例では長い全稼働期間を含む 10 個の全稼働期間の間のデータしか用いていないが , 到着間隔ならびにサービス時間の標本平均はそれぞれ 0.9942 , 1.004 となっており , ほぼ理論値と一致している . ジャックナイフ推定量は決して万能ではないが , 再生サイクルが十分に手に入らなかった場合には効果が期待できる .

非停止シミュレーション

再生点が見つからない場合や再生点の出現する割合が極めて小さい場合は , 適当な初期状態から長いシミュレーションを走らせることになる . このようなシミュレーションを非停止シミュレーションという . 以下では非停止シミュレーションを用いて定常状態における性能指標を得る場合について考える .

$\{Y_i; i = 1, 2, \ldots\}$ を一回のシミュレーションから得られた性能指標の出力結果を表す確率変数列とする . 十分に長い時間が経った後にはシステムは定常状態に達すると考えられることから , ある時点 $i = l$ 以降の確率変数 Y_i ($i = l+1, l+2, \ldots$) は近似的に未知の定常状態分布に従うとみなすことができる . あるいは , 少なくとも

$$\overline{Y}(m, l) = \sum_{i=l+1}^{m} Y_i/(m-l) \tag{51}$$

は $\overline{Y}(m) = \sum_{i=1}^{m} Y_i/m$ より偏りの少ない $b_Y = E(\lim_{i\to\infty} Y_i)$ の推定量であると考えられる．なぜなら，確率変数列の最初の部分ほど，より強く初期状態の影響を受けていると考えられるからである．

このように，非停止シミュレーションによって定常状態における性能指標を評価する際には，再生型シミュレーションと異なり，まず，過渡状態を同定すること，すなわち，適当な l を見い出すことが必要となる．さらに注意すべきことは確率変数列 $\{Y_i; i = l+1, \ldots, m\}$ が相関を持つことである．以下では，確率変数列 $\{Y_i; i = 1, 2, \ldots\}$ から過渡的な振舞いをしている部分を取り除く方法について議論したのち，信頼区間を得るために独立な標本平均を集める手法について述べる．

過渡状態の同定

一般に，単一の確率変数列 $\{Y_i; i = 1, 2, \ldots\}$ の実現値から過渡的振舞いをする部分系列 $\{Y_i; i = 1, \ldots, l\}$ を同定することは容易でない．そこで，同じ初期状態から開始された複数の独立なシミュレーションを行い，それらの結果から過渡状態を示す部分系列を同定することを考える．以下にその手順を要約する．

1. 同じ初期状態から開始された n' (≥ 10) 個の複数の独立なシミュレーションを行う．個々のシミュレーションにおける標本数 m' は時間が許す範囲で十分大きくとる．$Y_{j,i}$ を j 番目のシミュレーションにおける i 番目の出力結果とする．
2. i 番目の出力結果に対する標本平均 $\overline{Y}_i = \sum_{j=1}^{n'} Y_{j,i}/n'$ を計算する．定義より $\overline{Y}_i$ は $E(Y_i)$ の不偏推定量である．
3. 大きさ (window) w ($< m'/2$) の移動平均 $\overline{Y}_i(w)$ を次式に従い計算する．

$$\overline{Y}_i(w) = \begin{cases} \displaystyle\sum_{k=-(i-1)}^{i-1} \overline{Y}_{i+k}/(2i-1) & i = 1, \ldots, w \\ \displaystyle\sum_{k=-w}^{w} \overline{Y}_{i+k}/(2w-1) & i = w+1, \ldots, m'-w \end{cases} \tag{52}$$

   ここで w の値は $w = 1$ から順に大きくしていき，複数の値について $\overline{Y}_i(w)$ を計算する．
4. $\overline{Y}_i(w)$ のグラフを書き，滑らかなグラフになっているものの内，最小の w を持つものに注目する．このグラフにおいて，$i = l$ 以降の $\overline{Y}_i(w)$ が収束しているとみなすことのできる l の値を見い出す．もし，適当なグラフが得られなかった場合は，さらに 10 個程度の独立なシミュレーションを行い，これらの結果を加えて上記の手順を繰り返す．

図 6 は $\rho = 0.8$，平均サービス時間が 1 の $M/M/1$ に対して，システムが空，すなわち最初の客の待ち時間が 0 という初期状態から始め，$w = 10$ に対して $n' = 10, 30, 100$ の場合の結果を示したものである．$n' = 10$ では移動平均 $\overline{Y}_i(w)$ の値は大きく変動しているが，n' の値を大きくすることで変動は押えられている．また，$n' = 100$ の図より，初期状態が空であるという条件の影響は $i = 100$ 付近でほぼ消えていると考えられる．

ここでの目的は，確率変数列 $\{Y_i; i = 1, 2, \ldots\}$ の実現値から過渡的な振舞いをする部分系列 $\{Y_i; i = 1, \ldots, l\}$ を同定することであった．換言すれば $F_i(y) = P(Y_i \leq y)$ とすると，十分大きな i に対しては $F_i(y) \approx F(y) = \lim_{i\to\infty} F_i(y)$ であり，この近似式を満たす i の範囲を同定することであった．

上記の手法は，その手段として $E(Y_i)$ の収束を調べた．すなわち $E(Y_i) \approx b$ とみなせる i を求める手法である．確かに $F_i(y) \approx F(y)$ ならば $E(Y_i) \approx b$ であるが，あくまで $E(Y_i) \approx b$ は必要条件であり，その逆は成立しないことに注意する．よって，仮に $i > l$ に対して $E(Y_i) \approx b$ であっても，確率変数列 $\{Y_i; i = l+1, l+2, \ldots\}$ が定常状態にあるとは結論できない．この意味で，グラフより l の値を定めるときには，より安全側の選択が要求される．

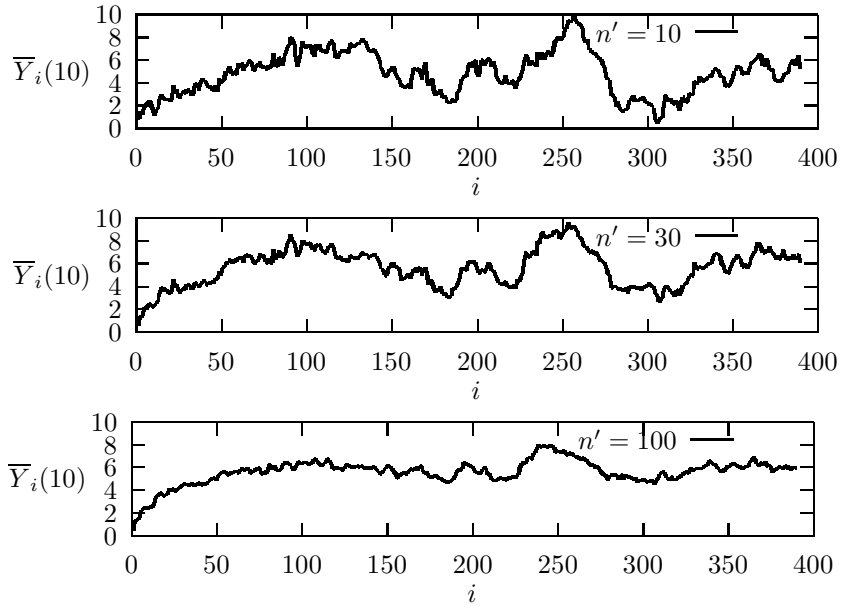

図 6: 過渡状態の同定

複製 / 消去法と集団平均法

以下では，確率変数列 $\{Y_i;\, i = 1, 2, \ldots\}$ から得られる部分系列 $\{Y_i;\, i = l+1, l+2, \ldots\}$ が近似的に定常であるとみなせる場合について考える．

非停止シミュレーションを用いて複数の独立な標本平均を得る最も一般的な方法は，n ($\geq n'$) 個の独立な長さ m ($\geq m'$) のシミュレーションを行い，個々のシミュレーションの標本平均を

$$\overline{Y}_j(m, l) = \sum_{i=l+1}^{m} Y_{j,i}/(m-l) \tag{53}$$

より計算するというものである．ただし $Y_{j,i}$ は j 番目のシミュレーションにおける i 番目の出力データから得られた性能指標を表す．$\overline{Y}_j(m, l)$ は独立同一分布に従う確率変数なので，6.3 節の結果を用いて信頼区間を得ることができる．これは複製 / 消去法 (Replication/Deletion Method) と呼ばれている．この手法を適用する際に注意すべき点は予め定められた l が適当かどうかである．もし，l が小さ過ぎると，推定量 $\overline{Y}(n) = \sum_{j=1}^{n} \overline{Y}_j(m, l)$ に偏りが生じる．また，推定量 $\overline{Y}(n)$ の分散を小さくする，すなわち狭い信頼区間を得るためには，n の値をなるべく大きく取る必要がある．表 4 は $\rho = 0.8$，平均サービス時間が 1 の $M/M/1$ にの平均待ち時間とその 95%信頼区間を合計 $n \times (m-l) = 150,000$ 個の標本に基づき求めたものである．いずれのシミュレーションにおいても $l = 500$ とした．n の値が小さいときは大きな信頼区間しか得られていないことが分かる．

表 4: 標本数の影響

n	$m-l$	平均待ち時間		
5	30,000	4.270	$\pm$	0.446
10	15,000	4.255	$\pm$	0.356
15	10,000	4.237	$\pm$	0.299
20	7,500	4.078	$\pm$	0.281
25	6,000	4.076	$\pm$	0.254
理論値		4.000		

もう一つの良く知られた方法は集団平均法 (Batch Means Method) と呼ばれる 1 回のシミュレーションの出力から信頼区間を得る手法である．確率変数列 $\{Y_i;\, i = 1, 2, \ldots\}$ は十分時間が経てば定常状態に達すると考えら

れる．定常な系列は共分散定常であるので，以下では確率変数列 $\{Y_i; i = l+1, l+2, \ldots, m\}$ が共分散定常であると仮定する．さらに，相関関数 ϕ_j (式 (7) 参照) が，$j \geq k'$ に対して

$$\phi_j \approx 0 \qquad \text{for } j \geq k' \tag{54}$$

であると仮定する．

上記の仮定の下で，$m-l$ 個の確率変数 Y_i ($i = l+1, \ldots, m$) を大きさ k ($\gg k'$) の n 個の部分系列 $\{Y_{l+i}; i = (j-1)k+1, \ldots, jk\}$ ($j = 1, \ldots, n$) に分ける (以下では，簡単化のため $nk = m-l$ を仮定する) . このとき，各部分系列の標本平均

$$\overline{Y}_j(k) = \frac{1}{k}\sum_{i=1}^{k} Y_{l+(j-1)k+i}, \qquad j = 1, \ldots, n \tag{55}$$

は近似的に独立同一分布に従う．よって 6.3 節の結果を用いて信頼区間を得ることができる．この集団平均法の困難さは集団の大きさ k をどのように選ぶかという点にある．小さな集団サイズを選ぶと，6.2 節 で示したように標本分散 $S^2(n)$ に大きな偏りが生じ，結果として誤った，真の値に対して非常に狭い信頼区間が得られてしまう．

表 5 は $\rho = 0.95$，平均サービス時間が 1 の $M/M/1$ の平均待ち時間とその 95%信頼区間を独立な 25 個の標本平均の値を用いて求めたものである．サイズは一つの標本平均に含まれる待ち時間の標本数を示している．表より，サイズが小さい場合，集団平均法は相関の影響を強く受けているため，信頼区間が真の幅よりも狭くなり，理論値を含んでいないことが観察される．

表 5: 複製 / 消去法と集団平均法

サイズ	複製 / 消去法			集団平均法		
500	16.528	±	3.268	37.279	±	13.101
1000	16.394	±	2.458	31.134	±	10.044
1500	19.093	±	2.263	25.722	±	7.867
2000	19.228	±	3.180	23.000	±	7.043
6000	20.108	±	2.206	18.907	±	3.733
10000	21.899	±	2.641	18.298	±	3.060
30000	20.523	±	1.353	18.483	±	1.531
50000	19.420	±	0.841	18.478	±	0.955
理論値	19.000			19.000		

6.5 シミュレーションの加速技法：分散減少法

シミュレーションは一般に時間消費型の数値実験である．それゆえ，より短い時間で所望の精度を持った解を得るにはどうすれば良いかといった問題が自然に生じてくる．シミュレーションに要する時間を短縮する一つの方法は，汎用のシミュレーションパッケージを用いることなく，自らの手で全てをコーディングし，そのプログラムをできる限り高速化することである．同じシミュレーション実験を行う場合でもコーディングの善し悪しで，同数の標本を得るのに要する時間は相当異なる．特に，ポインタの切替ならびにループ内で参照される配列を最小限に押えることは重要である．

以下では，上記のようなプログラミング技法ではなく，統計的な側面からシミュレーションの加速について議論する．すなわち，より少ない標本数で同程度の精度の推定値を得るにはどうすれば良いかという問題を扱う．この問題を正しく理解するために，平均 b，分散 σ^2 をもつ独立同一分布に従う確率変数列 $\{X_i; i = 1, 2, \ldots\}$ を考える．標本平均 $\overline{X}(n)$ の絶対誤差が β 以下である確率が $1-\alpha$ であるとすると，6.3 節の結果より，

$$\begin{aligned} 1-\alpha &\approx P\left(|\overline{X}(n) - b| \leq t_{n-1,1-\alpha/2}\sqrt{\sigma^2/n}\right) \\ &\leq P(|\overline{X}(n) - b| \leq \beta) \end{aligned}$$

を得る．よって，絶対誤差が β 以下であるために必要な標本数 $n^*(\beta)$ は近似的に

$$n^*(\beta) = \min\left\{n; t_{n-1,1-\alpha/2}\sqrt{\sigma^2/n} \leq \beta\right\} \tag{56}$$

で与えられる．式 (56) から標本平均 $\overline{X}(n)$ の分散 σ^2/n が小さければ，より少ない標本数で所望の絶対誤差 β が得られることが分かる．

このように，シミュレーションの加速とは標本平均 $\overline{X}(n)$ の分散を減少させることと等価である．以下では，一般に分散減少法 (variance reduction technique) と呼ばれる標本平均の分散を減少させる手法の幾つかについて解説する．

共通乱数法

二つのシステム 1 と 2 の性能を比較するため，これらのシステムにおける性能指標の差を求める場合を考える．例えば，以下のような状況を想定する．

例 6.7 回線容量 C_1 をもつルータにおけるパケットの平均待ち時間を改善するため，回線容量を C_2 $(C_2 > C_1)$ へ変更すると，どれだけパケットの平均待ち時間が減少するかを知りたい． □

以下では，複製 / 消去法によりシミュレーションを行い，性能指標の差を求めるものとする．それぞれのシステムに対する j 番目の複製シミュレーションから得られた注目する性能指標に対する標本平均を Y_{1j} , Y_{2j} とする．ここで評価したい量は $\eta = E(Y_{1j}) - E(Y_{2j})$ である．そこで，$Z_j = Y_{1j} - Y_{2j}$ ($j = 1, \ldots, n$) と置けば，$E(Z_j) = \eta$ であるので

$$\overline{Z}(n) = \sum_{j=1}^{n} Z_j/n \tag{57}$$

は η の不偏推定量である．さらに Z_j は独立同一分布に従うので

$$V(\overline{Z}(n)) = \frac{V(Z_j)}{n} = \frac{V(Y_{1j}) + V(Y_{2j}) - 2Cov(Y_{1j}, Y_{2j})}{n} \tag{58}$$

を得る．

もし，二つのシステムに対するシミュレーションが独立に行われるならば $Cov(Y_{1j}, Y_{2j}) = 0$ である．一方，何らかの方法で二つのシミュレーションから得られる標本平均 Y_{1j} と Y_{2j} の間に正の相関を持たすことができれば，$Cov(Y_{1j}, Y_{2j}) > 0$ となるので，結果として $V(\overline{Z}(n))$ の分散を減少させることができる．

標本平均 Y_{1j} と Y_{2j} の間に正の相関を持たせるために，二つのシステムに対するシミュレーションにおいて共通の乱数を用いる共通乱数法 (Common Random Numbers: CRN) と呼ばれる手法がある．どのように正の相関を持たせるかは，対象となっているシステムに依存するため一般的な議論はできない．そこで，以下の 2 つの例を用いてその考え方を示す．

例 6.8 $M/M/1$ と $M/M/2$[29] の平均待ち時間の差を求めることを考える．$M/M/c$ のシステムを支配する確率過程はポワソン到着を記述する乱数列と個々のサービス時間 (指数分布) を記述する乱数列によって表現される．そこで，二つのシステムにおける $k-1$ 番目と k 番目の客の到着間隔 $A_{i,k}$ ($i = 1, 2$) と k 番目の客のサービス時間 $B_{i,k}$ ($i = 1, 2$) に対して同じ標本値を用いる．このように，二つのシステムの間で入力となる確率過程に共通の標本値を用いてシミュレーションを行えば，両者の性能指標の間に正の相関 (一方の待ち時間が長ければ，他方も長い傾向にある) が生まれる． □

例 6.9 平均サービス時間が $M/M/1$ の平均待ち時間に与える影響を調べる場合を考える．まず，到着時間間隔に関しては同じ標本値を用いる．さらにサービス時間に関しては，同じ一様乱数列を用い，かつ，一様乱数に対して単調増加あるいは単調減少の関数による写像を利用した変換法，例えば，逆変換法を用いる．こうすることに

[29] 一般に $M/M/c$ とは客の到着がポワソン過程に従い，サービス時間が指数分布に従う c サーバ待ち行列を指す．

よって，一方のシステムで平均に比べて大きなサービス時間を持つ客があれば，他方のシステムにおける対応する客も平均に比べて大きなサービス時間を持つことになり，結果として両者の間に正の相関が生まれる． □

例 6.9 において，もし，非単調な関数による写像を利用した変換法が用いられているとすると，一方のシステムで得られたサービス時間が平均より大きいとき，他方のシステムで得られたサービス時間が平均より小さくなってしまう可能性がある．その結果，両者の性能指標の間には負の相関が生まれ，独立に行った場合よりも分散が増加してしまうことに注意する．それゆえ，例 6.9 のような形で CRN を用いる場合は，確率変数列を生成する際，どのような変換法が用いられているかということに十分注意を払う必要がある．特にシミュレーションパッケージでは逆変換法よりも高速な変換法が用いられていることが多く，これらの手法は一様乱数列に対して非単調の関数による写像を利用している場合が多いため注意を要する．

表 6 は到着時間間隔が 1 の $M/M/1$ において $\rho = 0.95$ の場合の平均待ち時間と $\rho = 0.9$ の場合の平均待ち時間の差とその 95%信頼区間を 25 回のシミュレーション用いて求めたものである．表におけるサイズは一つのシミュレーションに含まれる待ち時間の標本数を示している．表より，共通乱数法を用いることでより信頼度の高い結果が得られていることが分かる．

表 6: 共通乱数法 (CRN) による $M/M/1$ の平均待ち時間の差

サイズ	CRN	独立な試行
10000	9.462 ± 1.898	9.317 ± 2.495
20000	10.263 ± 1.199	10.440 ± 1.670
30000	10.104 ± 1.205	10.446 ± 1.537
40000	10.054 ± 1.106	10.259 ± 1.332
理論値	9.950	9.950

対照変量法

一つのシステムを複製 / 消去法によりシミュレーションを行い評価することを考える．ここで，複製 / 消去法によって $2n$ 個の標本平均を求め，それらを二つのグループ $\{Y_j^{(1)}; j = 1, \ldots, n\}$，$\{Y_j^{(2)}; j = 1, \ldots, n\}$ に分け，$Y_j = (Y_j^{(1)} + Y_j^{(2)})/2$ ($j = 1, \ldots, n$) を定義する．Y_j の標本平均 $\overline{Y}(n)$ は求めたい性能指標の不偏推定量であり，その分散は

$$V(\overline{Y}(n)) = \frac{V(Y_j)}{n} = \frac{V(Y_j^{(1)}) + V(Y_j^{(2)}) + 2Cov(Y_j^{(1)}, Y_j^{(2)})}{4n} \tag{59}$$

で与えられる．よって，$Y_j^{(1)}$ と $Y_j^{(2)}$ の間に負の相関を持たすことが出来れば結果として標本平均の分散を減少させることができる．このような負の相関を導入する手法を対照変量法 (Antithetic Variates: AV) といい，具体的には以下のように行われる．

まず，CRN と同様に，システムの入力となる各確率過程に対して一つの一様乱数列を割り当てる．さらに，(0,1) の間の値を取る一様乱数 U_i に対して，逆変換法のような単調な関数による写像を利用した変換法によって確率変数の実現値が得られるものとする．このような仮定の下で，1 番目の複製 / 消去法で用いられた各乱数列 U_i に対して，2 番目の複製 / 消去法では $U_i' = 1 - U_i$ なる乱数を用いる．定義より U_i' もまた 0 から 1 の間の値を取る一様乱数であるが，変換法の単調性より U_i で発生する確率変数の実現値と U_i' で発生する実現値の間には負の相関が生じる．例えば，ある U_i が大きなサービス時間を生成したとすると，写像の単調性より U_i' は比較的小さなサービス時間を生成する．これより，$(Y_j^{(1)}, Y_j^{(2)})$ の間に負の相関を導入することができる．

表 7 は平均到着間隔が $1/0.9$ の $M/D/1$[30] において $\rho = 0.95$ の場合の平均待ち時間とその 95%信頼区間を 24 回のシミュレーションを用いて求めたものである．サイズは一つのシミュレーションに含まれる待ち時間の標本数を示している．表より，AV を用いることでより良い推定量が得られていることが分かる．

[30] $M/D/1$ とは客の到着がポワソン過程に従い，サービス時間が一定である単一サーバ待ち行列を指す．

表 7: 対照変量法 (AV) による $M/D/1$ の平均待ち時間

サイズ	AV	独立な試行
5000	8.685 ± 0.827	8.498 ± 1.005
10000	9.269 ± 0.883	8.925 ± 0.665
15000	9.485 ± 0.768	8.733 ± 0.551
20000	9.428 ± 0.665	9.073 ± 0.557
理論値	9.500	9.500

CRN，AV に共通して言えることは，異なる二つの複製 / 消去シミュレーションにおいて，目的に応じた相関を導入し，結果として標本平均の分散を減少させることにある．これらは，(i) 到着時間間隔が大きいとき，待ち時間は減少するといった，性能指標 Y_j と入力確率過程との相関が明らかであり，(ii) 乱数を各確率過程に個別に割り当て，(iii) 一様乱数から確率変数の実現値への写像に単調性をもたせる，といった条件が揃ったときにうまく機能する．

例 **6.10** $M/M/1$ の平均待ち時間を求めるシミュレーションにおいて到着間隔とサービス時間の両方に AV を適用したとき，標本平均の分散は減少するであろうか．到着間隔が長ければ待ち時間は減少する傾向にあり，サービス時間が長ければ待ち時間は増加する傾向にある．よって短い到着間隔と長いサービス時間のペア，あるいは長い到着間隔と短いサービス時間のペアに対しては AV を用いることで待ち時間に負の相関を導入できるが，到着時間とサービス時間が共に長い，あるいは共に短い場合，AV によって負の相関が導入できるかどうかは定かでなく，十分な効果は期待できない． □

例 **6.11** 二つのシステムの比較を行いたいときに CRN のみならず AV を併用すると更に分散を減少させることができるであろうか．すなわち，CRN を用いて，システム i($i = 1, 2$)に対して $\{(Y_{i,j}^{(1)}, Y_{i,j}^{(2)}); i = 1, 2, j = 1, \ldots, n\}$ を生成しようとする場合である．一見，個々の標本の分散が AV により減少するため，より小さな分散が得られるように思えるかも知れない．しかし，このとき，図 7 に示すように $Y_{1,j}^{(1)}$ と $Y_{2,j}^{(2)}$，および $Y_{1,j}^{(2)}$ と $Y_{2,j}^{(1)}$ の間には負の相関が生じるため，正の相関を利用した分散減少法である CRN はうまく機能しない． □

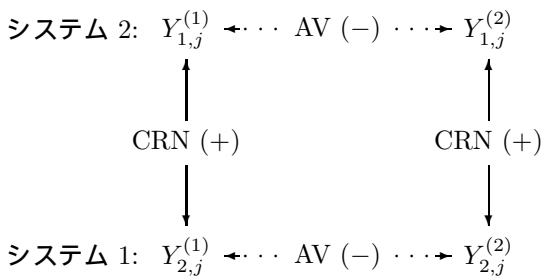

図 7: CRN と AV を併用したときの変数間の相関

このように，複数の手法を組合わせる場合は，それぞれの手法の元になっている考え方を十分理解した上で利用しなければならない．

制御変量法

シミュレーションを用いて，ある性能指標 X の期待値を評価することを考える．このとき，X と相関のある確率変数 Y があり，この平均 $E(Y) = \nu$ が既知であるとする．例えば $G/G/1$ において X が待ち時間であり，Y はサービス時間とすれば正の相関がある．

ここでパラメタ a を導入し，以下の制御された確率変数 X_C を考える．

$$X_C = X - a(Y - \nu) \tag{60}$$

このとき，$E(X_C) = E(X)$ であり，制御された確率変数 X_C の分散は

$$V(X_C) = V(X) + a^2 V(Y) - 2aCov(X, Y) \tag{61}$$

で与えられる．

さて $V(X_C)$ をパラメタ a の関数と見なすと

$$\frac{dV(X_C)}{da} = 2aV(Y) - 2Cov(X, Y), \qquad \frac{d^2V(X_C)}{da^2} = 2V(Y) \geq 0 \tag{62}$$

である．これより，$V(Y) = 0$ でなければ，すなわち Y が定数でなければ，$V(X_C)$ は a の下に凸な関数であり，次式で与えられる $a = a^*$ で最小となる．

$$a^* = \frac{Cov(X, Y)}{V(Y)} \tag{63}$$

よって $a = a^*$ を用いた場合の X_C を X_C^* とすると X_C^* の分散は

$$V(X_C^*) = V(X) - \frac{(Cov(X, Y))^2}{V(Y)} = (1 - \phi_{X,Y}^2)V(X) \tag{64}$$

で与えられる．ただし $\phi_{X,Y}$ は式 (7) で定義された X と Y の相関関数である．式 (64) より，X と Y の間に強い相関があれば $\phi_{X,Y}^2 \approx 1$ であり，結果として，分散 $V(X_C^*)$ は限りなく 0 に近付く．

実際にこの考え方を適用する際には，式 (63) で与えられる $a^* = Cov(X, Y)/V(Y)$ の推定値が必要となる．分散 $V(Y)$ の推定量に関しては式 (36) で与えられる標本分散 $S_Y^2(n)$ が利用できる．一方，$Cov(X, Y)$ の不変推定量 $C_{X,Y}(n)$ は

$$C_{X,Y}(n) = \frac{1}{n-1} \sum_{j=1}^{n} (X_j - \overline{X}(n))(Y_j - \overline{Y}(n)) \tag{65}$$

で与えられる．これらを用いて最適なパラメタ a^* の推定量 $\hat{a}^*(n)$ は

$$\hat{a}^*(n) = \frac{C_{X,Y}(n)}{S_Y^2(n)}$$

となり[31]，X_C の制御された推定量 $\overline{X}_C^*(n)$ は

$$\overline{X}_C^*(n) = \overline{X}(n) - \hat{a}^*(n)(\overline{Y}(n) - \nu)$$

で与えられる．このような手法を制御変量法 (control variates: CV) という．

表 8 は $\rho = 0.95$，平均サービス時間が 1 の $D/M/1$[32] の独立な 25 回のシミュレーションから得られた標本平均を X_j $(j = 1, \ldots, 25)$ とし，各シミュレーションから得られたサービス時間の標本平均を Y_j としたときの平均待ち時間の推定量を示している．この表において，サイズは各シミュレーションに含まれる待ち時間の標本数を示している．表より，CV を用いることでより良い推定量が得られていることが分かる．

問 **6.1** 客が到着した順にサービスされる単一サーバ待ち行列 $G/G/1$ において，n 番目と $n+1$ 番目の客の到着時間間隔を A_{n+1}，n 番目の客のサービス時間を S_n とし，n 番目の客の待ち時間を W_n とすると

$$W_{n+1} = \max(0, W_n + S_n - A_{n+1})$$

という関係が成り立つ．すなわち，計算機上で二つの乱数列 $\{A_n;\ n = 1, 2, \}$ と $\{S_n;\ n = 1, 2, \ldots\}$ を生み出せば，待ち時間の列 $\{W_n;\ n = 1, 2, \ldots\}$ は簡単に生み出すことができる．

[31] $C_{X,Y}(n)$ と $S_Y^2(n)$ の間には相関があることに注意する．それゆえ，これらの比をそのまま取ると偏りがあるので，なるべく大きな標本数 n を用意することが望ましい．少ない標本しか得られない場合はジャックナイフ推定量を用いる．

[32] $D/M/1$ とは客の到着間隔が一定で，サービス時間が指数分布に従う単一サーバ待ち行列を指す．

表 8: 制御変量法 (CV) による効果

サイズ	制御変量法	通常の複製 / 消去法
5000	8.7223	8.2991
10000	9.1090	8.5038
15000	9.0166	8.9084
20000	9.1917	9.0879
25000	9.0765	8.9663
理論値	9.1724	9.1724

(1) 利用率 ρ が 0.99 , 平均サービス時間が 1 の定常状態にある $M/M/1$, $M/D/1$ ならびに $D/M/1$ の平均待ち時間をそれぞれ AV あるいは CV を援用したシミュレーションによって評価せよ .

(2) 上記と同じ設定で , $M/M/1$ と $M/D/1$ の平均待ち時間の差 , ならびに $M/M/1$ と $D/M/1$ の待ち時間の差を CRN を援用したシミュレーションによって評価せよ .

□

A 定理の証明

A.1 大数の弱法則の証明

この節では定理 3.7 の証明を行う . 互いに独立で同じ分布に従い , 有限の平均 μ_X をもつ確率変数列 $\{X_n;\ n = 1, 2, \}$ と任意の正数 $\epsilon > 0$ が与えられたとき , ある正数 $b > 0$ を用いて新しい確率変数列 $\{Y_n;\ n = 1, 2, \ldots\}$ を以下で定義する .

$$Y_n = \begin{cases} X_n, & |X_n - \mu_X| \le b \\ \mu + b, & X_n - \mu_X > b \\ \mu - b, & X_n - \mu_X < -b \end{cases}$$

確率変数列 $\{Y_n;\ n = 1, 2, \ldots\}$ は互いに独立で同一の分布に従っており , その平均を μ_Y とする . Y_1 の分散 σ_Y^2 は

$$E((Y_1 - \mu_x)^2) = E(\{(Y_1 - \mu_Y) + (\mu_Y - \mu_X)\}^2) = \sigma_Y^2 + (\mu_Y - \mu_X)^2$$

であるので , $\sigma_Y^2 \le E((Y_1 - \mu_x)^2)$ である . さらに

$$\sigma_Y^2 \le E((Y_1 - \mu_x)^2) = \int_{-\infty}^{\infty} (x - \mu_x)^2 dF_Y(x) \le b \int_{-\infty}^{\infty} |x - \mu_x| dF_Y(x)$$

を得る . ただし $F_Y(x)$ は Y_1 の分布関数である . 最後の不等号は $|Y - \mu_X| \le b$ を用いた . さらに $\mu_X - b \le x \le \mu_X + b$ では $F_Y(x)$ は X_1 の分布関数に等しいことに注意すると ,

$$\sigma_Y^2 \le b \int_{-\infty}^{\infty} |x - \mu_x| dF_Y(x) \le b \int_{-\infty}^{\infty} |x - \mu_x| dF_X(x)$$

である .

さて , $T_n = Y_1 + \cdots + Y_n$ とし , 式 (4) のチェビシェフの不等式を ϵ を $\epsilon/2$ に変更して適用すると

$$P\left(\left|\frac{T_n}{n} - \mu_Y\right| \ge \frac{\epsilon}{2}\right) \le \frac{4\sigma_Y^2}{n\epsilon^2} \le \frac{4b\alpha}{n\epsilon^2}$$

となる . ただし $\alpha = \int_{-\infty}^{\infty} |x - \mu_x| dF_X(x) < \infty$ とした . $|\mu_Y - \mu_X|$ は $b \to \infty$ としたとき 0 へ近付くので , 任意の $\epsilon > 0$ に対して , 十分大きな b が存在し $|\mu_Y - \mu_X| < \epsilon/2$ とすることができる . よって , そのような b に対して

$$P\left(\left|\frac{T_n}{n} - \mu_X\right| \ge \epsilon\right) \le \frac{4b\alpha}{n\epsilon^2} \tag{66}$$

を得る[33]．

S_n と T_n は全ての k $(k = 1, \ldots, n)$ で $|X_k - \mu_X| \leq b$ となるような標本に対しては同じ値を取る．よって A_k で事象 $|X_k - \mu_X| > b$ を表すと

$$P(T_n \neq S_n) \leq P\left(\cup_{k=1}^{n} A_k\right) = nP(A_1) = nP(|X_1 - \mu_X| > b)$$

を得る．さらに $G(x) = P(|X_1 - \mu_X| \leq x)$ とすると

$$P(|X_1 - \mu_X| > b) = \int_b^\infty dG(x) \leq \frac{1}{b}\int_b^\infty x dG(x)$$

である．α が有限であるので最後の積分は $b \to \infty$ の極限で 0 へ収束する．よって，任意の $\delta > 0$ に対して十分大きな b が存在し $P(|X_1 - \mu_X| > b) \leq \delta^2/b$ とすることが出来る．すなわち，十分大きな b に対して

$$P(T_n \neq S_n) \leq nP(|X_1 - \mu_X| > b) \leq \frac{n\delta^2}{b} \tag{67}$$

事象 $|S_n/n - \mu_X| \geq \epsilon$ は事象 $|T_n/n - \mu_X| \geq \epsilon$ あるいは $S_n \neq T_n$ のいずれかが起こったときのみ起こり得るので，式 (66) と式 (67) を組合せて，$\delta = b/n$ とすると

$$P\left(\left|\frac{S_n}{n} - \mu_X\right| \geq \epsilon\right) \leq \frac{4\delta\alpha}{\epsilon^2} + \delta \tag{68}$$

を得る．式 (66) と式 (67) は任意の n と十分大きな b について成立するので，式 (68) は任意の正数 $\delta > 0$ と十分大きな n について成立する．よって，任意の正数 $\epsilon > 0$ に対して，適当な δ を選ぶことで式 (68) の右辺第 1 項，第 2 項共に，いくらでも小さくすることが出来る．よって定理が成立する．

A.2 中心極限定理の証明

この節では定理 3.10 の証明を行う．最初に，式 (5) に従って $Y_n = (X_n - \mu)/\sigma$ とおくと，$\{Y_n;\ n = 1, 2, \ldots$ は平均 0，分散 1 をもつ確率変数列である．標準正規分布 $N(0, 1)$ の特性関数は $\exp(-\eta^2/2)$ であるので，$Z_n = (S_n - n\mu)/(\sqrt{n}\sigma) = \sum_{k=1}^{n} Y_k/\sqrt{n}$ としたとき

$$\lim_{n\to\infty} E(\exp(i\eta Z_n)) = \exp\left(-\frac{\eta^2}{2}\right)$$

を示すことが出来れば，定理 3.5 より題意が示される．まず，

$$E(\exp(i\eta Z_n)) = E\left(\exp\left(i\frac{\eta}{\sqrt{n}}\sum_{k=1}^{n} Y_k\right)\right) = E\left(\prod_{k=1}^{n}\exp\left(i\frac{\eta}{\sqrt{n}}Y_k\right)\right)$$

であり Y_n は互いに独立で同一な分布に従うので

$$E(\exp(i\eta Z_n)) = E\left(\exp\left(i\frac{\eta}{\sqrt{n}}Y_1\right)\right)^n$$

となる．

一方，関数 $f(x)$ を点 c でテーラー展開すると

$$f(x) = \sum_{k=0}^{m}\frac{f^{[k]}(c)}{k!}(x-c)^k + \frac{1}{m!}\int_c^x (x-s)^m f^{[m+1]}(s)ds$$

[33] $|\mu_Y - \mu_x| < \epsilon/2$ のとき，$\{|T_n/n - \mu_X| \geq \epsilon\} = \{|T_n/n - \mu_Y + \mu_X - \mu_Y| \geq \epsilon\} \Rightarrow \{|T_n/n - \mu_Y| + |\mu_X - \mu_Y| \geq \epsilon\} \Rightarrow \{|T_n/n - \mu_Y| \geq \epsilon/2\}$ であるので，$A \Rightarrow B$ ならば $P(A) \leq P(B)$ に注意すると式 (66) を得る．

である．ただし $f^{[k]}(x)$ は $f(x)$ の k 階導関数である．ここで $c=0$，$m=1$ として $f(x)=e^{ix}$ へ適用すると

$$e^{ix}=1+ix-\int_0^x(x-t)e^{it}dt=1+ix-x^2\int_0^1(1-s)e^{isx}ds$$

を得る．ここで $x=\eta Y_1/\sqrt{n}$ とすると

$$\begin{aligned}\exp\left(i\frac{\eta}{\sqrt{n}}Y_1\right)&=1+i\frac{\eta}{\sqrt{n}}Y_1-\frac{\eta^2}{n}Y_1^2\int_0^1(1-s)\exp\left(i\frac{s\eta}{\sqrt{n}}Y_1\right)ds\\&=1+i\frac{\eta}{\sqrt{n}}Y_1-\frac{\eta^2}{n}Y_1^2+\frac{\eta^2}{n}Y_1^2\int_0^1(1-s)\left\{1-\exp\left(i\frac{s\eta}{\sqrt{n}}Y_1\right)\right\}ds\end{aligned}$$

よって $E(Y_1)=0$, $E(Y_1^2)=1$ に注意すると

$$E\left(\exp\left(i\frac{\eta}{\sqrt{n}}Y_1\right)\right)=1-\frac{\eta^2}{n}E\left(Y_1^2\right)+\frac{\eta^2}{n}E\left(Y_1^2\int_0^1(1-s)\left\{1-\exp\left(i\frac{s\eta}{\sqrt{n}}Y_1\right)\right\}ds\right)$$

を得る．

ここで確率変数 W_n を

$$W_n=Y_1^2\int_0^1(1-s)\left\{1-\exp\left(i\frac{s\eta}{\sqrt{n}}Y_1\right)\right\}ds$$

で定義する．

$$|W_n|=Y_1^2\int_0^1(1-s)\left|1-\exp\left(i\frac{s\eta}{\sqrt{n}}Y_1\right)\right|ds\le Y_1^2<\infty$$

であるので，$\varepsilon(n)=E(W_n)$ とすると，定理 2.3 の優収束定理ならびに $\lim_{n\to\infty}\exp(is\eta Y_1/\sqrt{n})=1$ より

$$\lim_{n\to\infty}\varepsilon(n)=E\left(\lim_{n\to\infty}Y_1^2\int_0^1(1-s)\left\{1-\exp\left(i\frac{s\eta}{\sqrt{n}}Y_1\right)\right\}ds\right)=0$$

を得る．以上より $\lim_{n\to\infty}\varepsilon(n)=0$ なる $\varepsilon(n)$ に対して

$$\lim_{n\to\infty}E(\exp(i\eta Z_n))=\lim_{n\to\infty}\left(1-\frac{\eta^2}{2n}+\frac{\eta^2}{n}\varepsilon(n)\right)^n=\exp\left(-\frac{\eta^2}{2}\right)$$

を示せば良い．

ここで

$$\left(1-\frac{\eta^2}{2n}+\frac{\eta^2}{n}\varepsilon(n)\right)^n=\left(1-\frac{\eta^2}{2n}\right)^n\left(1+\frac{\frac{\eta^2}{n}\varepsilon(n)}{1-\frac{\eta^2}{2n}}\right)$$

であり，右辺第 1 項は $n\to\infty$ のとき $\exp(-\eta^2/2)$ へ収束する．よって以下では右辺第 2 項が 1 へ収束することを示す．

$$\begin{aligned}\left(1+\frac{\frac{\eta^2}{n}\varepsilon(n)}{1-\frac{\eta^2}{2n}}\right)^n-1&=\sum_{k=1}^n\frac{n!}{k!(n-k)!}\left(\frac{2\eta^2\varepsilon(n)}{2n-\eta^2}\right)^k\\&=n\left(\frac{2\eta^2\varepsilon(n)}{2n-\eta^2}\right)\sum_{k=0}^{n-1}\frac{(n-1)!}{(k+1)!(n-1-k)!}\left(\frac{2\eta^2\varepsilon(n)}{2n-\eta^2}\right)^k\\&\le n\left(\frac{2\eta^2\varepsilon(n)}{2n-\eta^2}\right)\sum_{k=0}^{n-1}\frac{(n-1)!}{k!(n-1-k)!}\left(\frac{2\eta^2\varepsilon(n)}{2n-\eta^2}\right)^k\end{aligned}$$

となるので

$$\begin{aligned}\left|\left(1+\frac{\frac{\eta^2}{n}\varepsilon(n)}{1-\frac{\eta^2}{2n}}\right)^n-1\right|&\le\left|n\left(\frac{2\eta^2\varepsilon(n)}{2n-\eta^2}\right)\right|\cdot\left|1+\left|\frac{2\eta^2\varepsilon(n)}{2n-\eta^2}\right|\right|^{n-1}\\&=2\eta^2|\varepsilon(n)|\left|\frac{1}{2-\eta^2/n}\right|\cdot\left|1+\frac{1}{n}\left|\frac{2\eta^2\varepsilon(n)}{2-\eta^2/n}\right|\right|^{n-1}\end{aligned}$$

$n \to \infty$ のとき $\varepsilon(n) \to 0$ であるので，$|\varepsilon(n)| \leq 1$ かつ $n > \eta^2$ となるような十分に大きな n が存在し，そのとき

$$\left|\frac{1}{2-\eta^2/n}\right| \leq 1, \qquad \left|\frac{2\eta^2\varepsilon(n)}{2-\eta^2/n}\right| \leq 2\eta^2$$

となる．よって，そのような十分大きな n に対して

$$\left|\left(1+\frac{\frac{\eta^2}{n}\varepsilon(n)}{1-\frac{\eta^2}{2n}}\right)^n - 1\right| \leq 2\eta^2|\varepsilon(n)|\left(1+\frac{2\eta^2}{n}\right)^{n-1} \leq 2\eta^2|\varepsilon(n)|\left(1+\frac{2\eta^2}{n}\right)^n$$

となり，

$$\lim_{n\to\infty}\left(1+\frac{2\eta^2}{n}\right)^n = e^{2\eta^2} < \infty$$

であるので，$n \to \infty$ の極限において右辺第 2 項は 1 へ収束する．

参考文献

[1] Brémaud, P. (1998) *Markov Chains, Gibbs Fields, Monte Carlo Simulation, and Queues*. Springer-Verlag, New York.

[2] Feller, W. (1971) *An Introduction to Probability Theory and Its Applications, Vol.II*. John Wiley & Sons, New York.

[3] Gallager, R. G. (1996) *Discrete Stochastic Processes*. Kluwer, Boston.

[4] Gross, D. and Harris, C. M. (1998) *Fundamentals of Queueing Theory, 3rd Ed.*. Wiley, New York.

[5] Law, A. M. and Kelton, W. D. (2000) *Simulation Modeling and Analysis*, 3nd Ed., McGraw-Hill, New York.

[6] 宮沢政清 (1993) 確率と確率過程. 近代科学社.

[7] 野田一雄，宮岡悦良 (1992) 数理統計学の基礎. 共立出版.

[8] Press, W. H., Teukolsky, S. A., Vetterling, W. T. and Flannery. B. P. (1988) *Numerical Recipies in C*. Cambridge Univ. Press, New York.
(邦訳：丹慶勝市，奥村晴彦，佐藤俊郎，小林誠 (1993) ニューメリカルレシピ・イン・シー，C 言語による数値計算のレシピ，技術評論社)

[9] 薩摩順吉 (1989) 確率・統計. 岩波書店.

[10] 佐藤 担 (1994) はじめての確率論，測度から確率へ. 共立出版.

[11] 滝根哲哉，伊藤大雄，西尾章治郎 (2001) ネットワーク設計理論. 岩波書店.

[12] Wolff, R. W. (1989) *Stochastic Modeling and the Theory of Queues*. Prentice Hall, Englewood Cliffs.