1 事象・確率

1.1 標本空間と事象

確率的な判断が妥当と思われる実験や観測などを考えよう．そこにおいて, 起こるか起こらないか不確定なものを**事象 (event)** と呼ぶことにする．これを「起こりうる現象」の集まったものと見なすことにする．では, その「起こりうる現象」を全て集めたものは一番大きな事象ということになるが, これを**標本空間 (sample space)** と呼び, Ω(オメガと読む) で表ことにする．これは**全事象 (total event)** とも呼ばれる.「何も起こらない」というようなものも事象の仲間に入れておくと便利なので, それを**空事象 (empty event)** と呼び, $\emptyset$(ファイと読む) で表すことにする.

“起こりうる現象” を最小単位まで分解して記号化しておくと, 標本空間はそれらの記号の全ての集まり (**集合 (set)**) とみなせる．このことにより, 事象は標本空間の部分集合として表現できるようになる．　事象と集合とを同一視できるならば, 集合についての演算をそのまま事象の演算として借用できることになる．A, B を事象とするとき,「A または B が起こる」という事象を記号

$$A \cup B$$

で表し, **和事象 {sum (union, join) event}** と呼ぶ.「A も B も起こる」という事象を記号

$$A \cap B$$

で表し, **積事象 {product (intersection, meet) event}** と呼ぶ.

$A_1, A_2, \ldots, A_n$ を事象とする.「$A_i,\ i = 1, 2, \ldots, n$ のうち少なくとも一つが起こる」 という事象を記号

$$\bigcup_{i=1}^{n} A_i$$

で表し, 同様に,「全ての $A_i,\ i = 1, 2, \ldots, n$ が起こる」という事象を記号

$$\bigcap_{i=1}^{n} A_i$$

で表す．可算無限個の $A_i,\ i = 1, 2, \ldots$ に関しても同様に

$$\bigcup_{i=1}^{\infty} A_i, \quad \bigcap_{i=1}^{\infty} A_i$$

を定義する．これらに関しては, **分配法則**が成立つ.

$$B \cap \left(\bigcup_{i=1}^{n} A_i\right) = \bigcup_{i=1}^{n} (B \cap A_i), \quad B \cap \left(\bigcup_{i=1}^{\infty} A_i\right) = \bigcup_{i=1}^{\infty} (B \cap A_i),$$
$$B \cup \left(\bigcap_{i=1}^{n} A_i\right) = \bigcap_{i=1}^{n} (B \cup A_i), \quad B \cup \left(\bigcap_{i=1}^{\infty} A_i\right) = \bigcap_{i=1}^{\infty} (B \cup A_i).$$

積事象については

$$A_1 A_2 \cdots A_n = \bigcap_{i=1}^{n} A_i, \quad A_1 A_2 \cdots = \bigcap_{i=1}^{\infty} A_i$$

と表す事もある.

「A は起こらない」という事象を

$$A^c$$

で表し, A の**余事象** **(complementary event)** と呼ぶ. また,「事象 A は起ったが, 事象 B は起こらなかった」ならば, これを**差事象** **(difference event)** と呼び, 次のように表す.

$$A - B = A \cap B^c.$$

「事象 A が起こる時は必ず事象 B も起こる」ならば, このことを

$$A \subset B$$

で表す. また, **ド・モルガンの法則**が成立つ.

$$(A \cup B)^c = A^c \cap B^c, \quad \text{一般的には} \quad \left(\bigcup_{i=1}^{n} A_i\right)^c = \bigcap_{i=1}^{n} A_i^c, \quad \left(\bigcup_{i=1}^{\infty} A_i\right)^c = \bigcap_{i=1}^{\infty} A_i^c.$$

$$(A \cap B)^c = A^c \cup B^c, \quad \text{一般的には} \quad \left(\bigcap_{i=1}^{n} A_i\right)^c = \bigcup_{i=1}^{n} A_i^c, \quad \left(\bigcap_{i=1}^{\infty} A_i\right)^c = \bigcup_{i=1}^{\infty} A_i^c.$$

また, $A \cap B = \emptyset$ が成り立つとき, 事象 A と B は**排反** **(exclusive)** または互いに**素** **(disjoint)** であるという. そのときの和事象を $A + B$ と表わす. 同様に, $A_1, A_2, \ldots$ が互いに排反ならば, 和事象に対して和の記号を用いる.

$$\sum_{i=1}^{n} A_i = \bigcup_{i=1}^{n} A_i, \quad \sum_{i=1}^{\infty} A_i = \bigcup_{i=1}^{\infty} A_i.$$

1.2 可測集合族の定義

次の 3 つの条件 (集合の演算に関して閉じている) を満たす集合族のことを可測集合族といい，それに属している集合は**可測** **(measurable)** であるという．我々が普通に考える集合は可測であるから可測性を特に意識する必要はない.

(M1) $\emptyset \in \mathcal{A}$, $\Omega \in \mathcal{A}$.

(M2) $A_n \in \mathcal{A},\ n = 1, 2, \ldots \Longrightarrow \bigcup_{n=1}^{\infty} A_n \in \mathcal{A}$.

(M3) $A \in \mathcal{A} \Longrightarrow A^c \in \mathcal{A}$

そのとき，組 $(\Omega, \mathcal{A})$ を**可測空間** **(measurable space)** という.

1.3 確率の定義

事象の起こり易さの程度をある規則に従って 0 以上 1 以下の数で表した数値を，その事象の **(確率 probability)** という. 事象に数値を対応させるような写像 (集合関数) P に対し，以下の 3 つの性質を満たすとき，P を**確率測度** **(probability measure)**，$P(A)$ の値を事象 A の確率という．ロシアの数学者コルモゴロフにより定義 (1933 年) されたものであり，公理論的確率の定義と呼ばれている.

(P1) 任意の事象 $A \in \mathcal{A}$ に対して,

$$0 \le P(A) \le 1.$$

(P2) 全事象 Ω に対して,

$$P(\Omega) = 1.$$

(P3) **(可算加法性)** $A_n \in \mathcal{A},\ n = 1, 2, \ldots$ が互いに素 $(A_m \cup A_n = \phi\,(m \neq n))$ ならば,

$$P\left(\bigcup_{n=1}^{\infty} A_n\right) = \sum_{n=1}^{\infty} P(A_n).$$

そのとき，組 $(\Omega, \mathcal{A}, P)$ を**確率空間 (probability space)** という.

1.4 確率測度に関する基本的性質

定理 0.1 確率測度に関する基本的性質は，(P1)–(P3) から導かれる.

(1) 空事象 $\emptyset$ に対して，$P(\emptyset) = 0$.

(2) **(有限加法性)** 有限個の排反な $A_1, A_2, \ldots, A_n$ に対して，$P\left(\sum_{i=1}^n A_i\right) = \sum_{i=1}^n P(A_i)$.
とくに,

$$P(A + B) = P(A) + P(B).$$

(3) **(単調性)** $A \subset B \Longrightarrow P(A) \leq P(B)$.

(4) **(余事象の確率)** 任意の事象 A に対して,

$$P(A^c) = 1 - P(A).$$

(5) **(加法定理)** 任意の事象 A, B に対して,

$$P(A \cup B) = P(A) + P(B) - P(A \cap B).$$

(6) **(ポアンカレの包除展開公式)** n 個の事象 $A_1, A_2, \ldots, A_n$ に対して, その和集合の確率は次で与えられる.

$$P\left(\bigcup_{i=1}^n A_i\right) = S_1 - S_2 + S_3 - S_4 + \cdots + (-1)^{n+1} S_n\,.$$

ただし,

$$\begin{aligned}
S_1 &= \sum_{i=1}^n P(A_i), \\
S_2 &= \sum_{1 \leq i < j \leq n} P(A_i \cap A_j), \\
S_3 &= \sum_{1 \leq i < j < k \leq n} P(A_i \cap A_j \cap A_k), \\
&\vdots \\
S_n &= P\left(\bigcap_{i=1}^n A_i\right).
\end{aligned}$$

確率の加法性に関する性質は, ある事象の確率を計算したいならば, その事象を (確率がすでに明らかなあるいはその確率を簡単に計算できる) 事象に分解することが重要であることを示している.

(7) **(連続性)** $A_n \in \mathcal{A},\ (n = 1, 2, \ldots),\ A_1 \subset A_2 \subset \cdots$ (単調増大) であるとき，$A = \bigcup_{n=1}^\infty A_n$ とすれば,

$$P(A) = \lim_{n \to infty} P(A_n).$$

(8) **(連続性)** $A_n \in \mathcal{A},\ (n = 1, 2, \ldots),\ A_1 \supset A_2 \supset \cdots$ (単調減少) であるとき，$A = \bigcap_{n=1}^\infty A_n$ とすれば,

$$P(A) = \lim_{n \to infty} P(A_n).$$

1.5 条件付き確率と独立性

1.5.1 条件付き確率と乗法定理

標本空間 $\Omega = \{\omega_1, \omega_2, \ldots, \omega_N\}$ が 2 つの事象 A, B によって次のように分割されているとする.

	B	B^c	計
A	a	b	$a+b$
A^c	c	d	$c+d$
計	$a+c$	$b+d$	$N\,(=a+b+c+d)$

このとき, 等確率の仮定で確率を与えると

$$P(A) = (a+b)/N, \quad P(B) = (a+c)/N, \quad P(A \cap B) = a/N$$

となる. これより, 事象 A が起こったということがわかっているとして, 仮の標本空間を A であるとみなしてみよう. このとき, 事象 B が起こる確率を $P(B \,|\, A)$ で表すと,

$$P(B \,|\, A) = \frac{a}{a+b} = \frac{a/N}{(a+b)/N} = \frac{P(A \cap B)}{P(A)}$$

と書ける. つまり, 事象 A が起こったという情報が与えられたときの B の起こる確率が, 標本空間 Ω の上の等確率の仮定で与えられた確率から求められることを意味している.

条件付き確率の定義　A, B を事象とするとき,

$$P(B \,|\, A) = \frac{P(A \cap B)}{P(A)}, \quad \text{ただし } P(A) \neq 0.$$

を事象 A が起きたことが与えられたときの事象 B の**条件付き確率 (conditional probability)** という.

注意 条件付き確率も確率というからには, 確率が満たすべき 3 条件 (P1)–(P3) を満足する.

定理 0.2 (乗法定理) $P(A) \neq 0,\ P(B) \neq 0$ とする.

$$P(A \cap B) = P(A)P(B \,|\, A) \tag{1.1}$$

$$= P(B)P(A \,|\, B). \tag{1.2}$$

一般的には次の形で与えられる. 事象 $A_1, A_2, \ldots, A_n$ において, $P(\cap_{i=1}^{n-1} A_i) > 0$ のとき

$$P(\cap_{i=1}^{n} A_i) = P(A_1)P(A_2 \,|\, A_1)P(A_3 \,|\, A_1 \cap A_2) \cdots P(A_n \,|\, \cap_{i=1}^{n-1} A_i).$$

定理 0.3 (全確率の公式) 事象系 $A_1, A_2, \ldots, A_k$ は全事象を尽くし, かつ互いに排反とする. すなわち

$$\bigcup_{i=1}^{k} A_i = A_1 \cup A_2 \cup \cdots \cup A_k = \Omega, \quad A_i \cap A_j = \emptyset \quad (i \neq j) \tag{1.3}$$

を満たすとき,

$$P(B) = \sum_{i=1}^{k} P(A_i)P(B \,|\, A_i).$$

1.5.2 事象の独立性

事象 A, B において, $P(A) \neq 0$ のとき,

$$P(B \,|\, A) = P(B) \tag{1.4}$$

が成り立つとする．このとき

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

が成り立つから, 事象 A が起きたか否かが事象 B が起こる確率に影響を与えない．つまり, A についての情報が与えられても, B についての確率は変化しないので, A は B についての (価値ある) 情報を全く含まない事象であるということが分かる．このとき, 事象 B は事象 A に**独立 (independent)** であるといい,

$$A \perp\!\!\!\perp B$$

で表す．(1.4) 式が成り立つとき (1.1) 式より

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

この式は $P(A) = 0$ または $P(B) = 0$ の時にも定義できるので「独立性」は (1.4) 式よりも一般にこちらで定義される．このとき, $P(B) \neq 0$ の時には, 条件付き確率の定義より

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

となる．ゆえに,

$$P(B \,|\, A) = P(B) \iff P(A \cap B) = P(A)P(B) \iff P(A \,|\, B) = P(A)$$

となることがわかる．つまり, A と B が独立ならば, これらは互いに相手に対する価値ある情報を含まないということになる．

明らかに次のことが成り立つ．

$$A \perp\!\!\!\perp B \iff A \perp\!\!\!\perp B^c \iff A^c \perp\!\!\!\perp B^c \iff A^c \perp\!\!\!\perp B.$$

一般に n 個の事象系 $A_1, A_2, \ldots, A_n$ は任意の部分事象系 $A_{i_1}, A_{i_2}, \ldots, A_{i_k}$ $(1 \leq i_1 < i_2 < \cdots < i_k; k = 2, \ldots, n)$ に対して

$$P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}) = P(A_{i_1})P(A_{i_2}) \cdots P(A_{i_k})$$

が成り立つとき互いに独立であるという．これを次のように表記する．

$$\perp\!\!\!\perp\{A_1, A_2, \ldots, A_n\}.$$

例えば $n = 3$ の場合, $\perp\!\!\!\perp\{A_1, A_2, A_3\}$ は次に等しい．

$$\begin{aligned}
P(A_1 \cap A_2) &= P(A_1)P(A_2), \\
P(A_1 \cap A_3) &= P(A_1)P(A_3), \\
P(A_2 \cap A_3) &= P(A_2)P(A_3), \\
P(A_1 \cap A_2 \cap A_3) &= P(A_1)P(A_2)P(A_3).
\end{aligned}$$

このことから

$$\perp\!\!\!\perp\{A_1, A_2, A_3\} \iff A_1 \perp\!\!\!\perp A_2, A_2 \perp\!\!\!\perp A_3, A_1 \perp\!\!\!\perp A_3, A_1 \cap A_2 \perp\!\!\!\perp A_3$$

であることも明らかであろう．すると, 次も容易に得られる．

$$\perp\!\!\!\perp\{A_1, A_2, A_3\} \iff \perp\!\!\!\perp\{A_1, A_2, A_3^c\}.$$

例題 1.1 次を証明せよ.

$$\perp\!\!\!\perp\{A, B, C\} \implies A \cup B \perp\!\!\!\perp C.$$

(解)

$$\begin{aligned}\perp\!\!\!\perp\{A, B, C\} &\iff \perp\!\!\!\perp\{A^c, B, C\}\\ &\iff \perp\!\!\!\perp\{A^c, B^c, C\}\\ &\implies A^c \cap B^c \perp\!\!\!\perp C\\ &\iff A \cup B \perp\!\!\!\perp C. \qquad \square\end{aligned}$$

この例題により, $\perp\!\!\!\perp\{A_1, A_2, \ldots, A_n\}$ のとき, 事象 $A_1, A_2, \ldots, A_{n-1}$ に対して和事象・積事象・余事象を任意に適用して事象 D が作られているとすると, この D と A_n とは独立であることがわかる.

例題 1.2 2 つずつは互いに独立であるが, 3 つは独立でない例

2 枚の正常なコインを投げるとすると, その標本空間は

$$\{(H, H),\ (H, T),\ (T, H),\ (T, T)\}$$

とおけて, 根元事象の確率はそれぞれ 1/4 と考えられる. このとき, 次のように事象を定義する.

$$\begin{aligned}A &= \{(H, H),\ (H, T)\}\\ B &= \{(H, H),\ (T, H)\}\\ C &= \{(H, H),\ (T, T)\}\end{aligned}$$

これらのそれぞれ 2 つずつは互いに独立であるが, $\perp\!\!\!\perp\{A, B, C\}$ ではない.

1.6 ベイズの定理 (Bayes' theorem)

事象系 $A_1, A_2, \ldots, A_k$ は互いに排反であり, その和集合は全事象であるとする. すなわち,

$$\bigcup_{i=1}^{k} A_i = A_1 \cup A_2 \cup \cdots \cup A_k = \Omega, \tag{1.5}$$

$$A_i \cap A_j = \emptyset \quad (i \neq j). \tag{1.6}$$

このとき, 次の命題が成り立つ.

定理 0.4 ((ベイズの定理)) 任意の事象 B と (1.5), (1.6) を満たす $A_i,\ i = 1, 2, \ldots, k$ に対して

$$P(A_i \,|\, B) = \frac{P(A_i)P(B \,|\, A_i)}{P(A_1)P(B \,|\, A_1) + P(A_2)P(B \,|\, A_2) + \cdots + P(A_k)P(B \,|\, A_k)}.$$

条件つき確率の定義と全確率の公式から, 証明は容易に得られる. これを**ベイズの定理 (Bayes' theorem)** という. $P(Ai)$ 及び $P(A_i \,|\, B)$ をそれぞれ事象 A_i の**事前確率 (prior probability)**, **事後確率 (posterior probability)** という.

例題 1.3 患者がある症状を訴えるとき, 5% が A 疾患であることが知られている. ある精密検査 B は真の A 疾患者に対して, 70% が陽性反応を示し, A 疾患でない人に対しても 10% の陽性反応を示す. その症状を訴えた患者の精密検査 B の結果が陽性反応を示したとき, その患者が A 疾患である確率を求めよ.

(解) ある症状を訴える患者について, 事象を

$A =$「症状を訴える患者が A 疾患である」

とすると, 題意より $P(A)=0.05, P(A^c)=0.95$ である．また

$$B = \text{「精密検査 B に陽性反応がでる」}$$

とすると, 題意より $P(B\,|\,A)=0.7, P(B\,|\,A^c)=0.1$ であるから, ある症状を訴えた患者で精密検査 B の結果が陽性反応を示した患者が A 疾患である確率は

$$\begin{aligned} P(A\,|\,B) &= \frac{P(A\cap B)}{P(B)} = \frac{P(A)P(B\,|\,A)}{P(A)P(B\,|\,A)+P(A^c)P(B\,|\,A^c)} \\ &= \frac{0.05\ \times\ 0.7}{0.05\times 0.7+0.95\times 0.1} \fallingdotseq 0.270. \end{aligned} \qquad \square$$

練習問題1

問題 1.1 任意の事象 A, B に対して次を示せ.

$$P(A \cap B) - P(A)P(B) = P(A^c)P(B) - P(A^c \cap B) = P(A)P(B^c) - P(A \cap B^c).$$

問題 1.2 (**Poincare の包除展開公式**) n 個の事象 $A_1, A_2, \ldots, A_n$ に対して次を定義する.

$$S_k = \sum_{1 \le i_1 < i_2 < \cdots < i_k \le n} P(A_{i_1} \cap A_{i_2} \cap \cdots \cap A_{i_k}).$$

ただし, $1, 2, \ldots, n$ の中から k 個の数 $i_1, i_2, \ldots, i_k$ を取り出すときの全ての組み合わせについて和を取るものとする. このとき次が成り立つことを示せ.

$$P(A_1 \cup \cdots \cup A_n) = S_1 - S_2 + \cdots + (-1)^{n+1} S_n\,.$$

問題 1.3 対称差 $A \triangle B$ に対して, $d(A, B) = P(A \triangle B)$ と定義するとき, d は次の性質を満たすことを示せ.

(1) $d(A, B) \ge 0.$

(2) $A = B \implies d(A, B) = 0.$

(3) $d(A, B) = d(B, A).$

(4) $d(A, B) \le d(A, C) + d(C, B).$

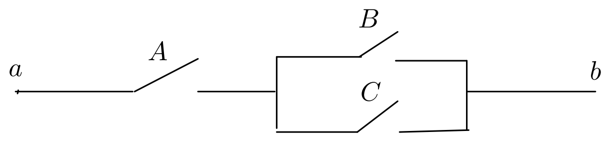
問題 1.4 次の命題が正しいならば証明せよ. 誤っているならば反例を示せ.

(1) $E \perp\!\!\!\perp F,\ E \perp\!\!\!\perp G \implies E \perp\!\!\!\perp (F \cup G).$

(2) $E \perp\!\!\!\perp F,\ E \perp\!\!\!\perp G,\ F \cap G = \emptyset \implies E \perp\!\!\!\perp (F \cup G).$

(3) $E \perp\!\!\!\perp F,\ F \perp\!\!\!\perp G,\ E \perp\!\!\!\perp (F \cap G) \implies G \perp\!\!\!\perp (E \cap F).$

問題 1.5 下の図の回路で ab の導通する確率を求めよ．ここで，A, B, C は，それぞれスイッチ A, B, C が閉じている事象とし，$P(B) = 1/2,\ P(A|B) = 1/3,\ P(A \cap C|B^c) = 1/4$ とする．



問題 1.6 (飛行機の墜落原因に関する例)

飛行機の故障が, 機体の故障, エンジンの故障, 無線機の故障, 操縦の失敗の 4 つに分けられるとする. ある距離を飛ぶ間にこれらの故障が起こる確率をそれぞれ 0.002, 0.002, 0.010, 0.001 とし, その故障が起きたとき, 飛行機が墜落する確率をそれぞれ 0.25, 0.30, 0.01, 0.90 とする. 今, 一機の飛行機が墜落したが, 原因は全く分からないとする. このとき, エンジンの故障が原因で墜落した確率はいくらか.

問題 1.7 クイズ番組での話である. 3 通の封筒があり, その中の 1 つだけにアタリくじが入っている. あなたはその中の 1 つを選びました. すると, 司会者は残りの 2 つの封筒の中から 1 通を選び「これを選ばなくてよかったですね, これはハズレです」と言ってハズレであることを示して捨てた (司会者はどの封筒にアタリくじが入っているか知っている). 司会者曰く「ここで提案があります. この残った封筒とあなたの選んだ封筒を交換する気はありませんか?」あなたはどうしますか?

問題 1.8 あるギャンブラーはポケットに正常な硬貨と両面表である硬貨を 1 枚ずつ持っている. 無作為に 1 枚選び, 放り投げたところ表がでた. この硬貨が正常である確率はいくらか. この硬貨をもう 1 度投げたところまた表であった. このとき, 硬貨が正常である確率はいくらか. さらに投げたら今度は裏であった. 硬貨が正常である確率はいくらか.

2 確率変数とその確率分布

母集団の特性を表すカテゴリーや数量を変数と考えるとき，この変数の取る値は偶然変動を伴う**試行 (trial)** (観測, 実験) の結果であるので，これを**確率変数 (random variable)** といい, 通常 $X, Y, Z, \ldots$ などで表す．確率変数 X の取るすべての値 (標本値，観測値，実現値) の集まりを**観測空間または標本空間 (sample space)** といい, $\mathbb{R}_X$ で表す．確率変数の可能な実現値とその確率を対にしたものをその確率変数の**確率分布 (probability distribution)** という．確率変数は大別して離散型と連続型の 2 つの確率変数 (分布) に分類される．

2.1 累積分布関数 (cumulative distribuiton function)

X を確率変数とし，

$$F(x) = P(X \leq x), \quad x \in \mathbb{R} \text{ (実数全体の集合)} \tag{2.7}$$

を x の関数とみるとき, $F(x)$ を X の**累積分布関数 (cumulative distribuiton function) (c.d.f.)** という．累積分布関数は次の性質をもつ．

1. 任意の実数 x に対し， $0 \leq F(x) \leq 1$.
2. $\lim_{x \to -\infty} F(x) = 0, \quad \lim_{x \to \infty} F(x) = 1$.
3. $F(x)$ は単調非減少関数である．すなわち， $x_1 < x_2 \Longrightarrow F(x_1) \leq F(x_2)$.
4. $F(x)$ は右連続関数である．すなわち， $\lim_{x \to a+0} F(x) = F(a), \quad \forall a \in \mathbb{R}$.

2.2 離散型確率変数 (discrete random variable)

確率変数 X の標本空間 $\mathbb{R}_X$ が，有限集合 $\mathbb{R}_X = \{x_1, x_2, \ldots, x_n\}$ または可算無限集合 $\mathbb{R}_X = \{x_1, x_2, \ldots\}$ であるとき, X を**離散型確率変数 (discrete random variable)** という．

$$p(x_i) = P(X = x_i) \quad (i = 1, 2, \ldots, n)$$

を離散型確率変数 X の**確率関数 (probability function)** という．確率関数について, 常に次が成り立つ．逆に，これらの性質をもつものを確率関数と定義してもよい．

(1) $p(x_i) \geq 0 \quad (i = 1, 2, \ldots)$.

(2) $\displaystyle\sum_{i=1}^{\infty} p(x_i) = 1$.

このとき分布関数は， $F(x) = \displaystyle\sum_{\{i | x_i \leq x\}} p(x_i)$ と表される．ここで， $\displaystyle\sum_{\{i \,:\, x_i \leq x\}}$ は $x_i \leq x$ であるような i についての和を表す．

2.3 連続型確率変数 (continuous random variable)

確率変数 X の標本空間 $\mathbb{R}_X$ を $\mathbb{R}$ とする．適当な関数 $f(x)$ が存在して，

(1) $f(x) \geq 0, \quad x \in \mathbb{R}$

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

が成立し，かつ任意の $a < b$ に対して

$$P(a < X < b) = \int_a^b f(x)\,dx \tag{2.8}$$

であるとき，X を**連続型確率変数 (continuous random variable)** といい，$f(x)$ を X の**確率密度関数 (probability density function)** (**p.d.f.**) という．X が連続型確率変数のとき, 常に $P(X = x) = 0$ であり, 連続型確率変数に対しては区間に対し確率を考える．

X が連続型確率変数のとき, その c.d.f. $F(x)$ はたかだか有限個の点を除き微分可能となり,

$$f(x) = F'(x) \tag{2.9}$$

が成り立つ． (2.9) から直ちに次が得られる．

$$F(x) = \int_{-\infty}^x f(u)\,du$$

(1), (2) は $f(x)$ が p.d.f. であるための条件である．(2.8) から, 連続型確率変数の分布は確率密度関数によって与えられることがわかる．

2.4 確率変数 (分布) の特性値

2.4.1 期待値 (expectation) (平均 (mean))

- 離散型確率変数の期待値

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

  を離散型確率変数 X の**期待値 (expectation)** または**平均 (mean)** という．分布の “中心的位置” を表す特性値である．

- 任意の関数 $\varphi(x)$ に対して, 確率変数 $\varphi(X)$ の期待値を次で定義する．

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

連続型確率変数 X に対しては期待値を次で定義する．

- 連続型確率変数の期待値

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

  を連続型確率変数 X の期待値という．

- 任意の関数 $\varphi(x)$ に対して, 確率変数 $\varphi(X)$ の期待値を次で定義する．

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

2.4.2 分散 variance, 標準偏差 standarad deviation

分布の散らばり具合を表す特性値を定義しよう.

確率変数 X に対して, $X-\mu$ を期待値からの**偏差 (deviation)** という. 偏差の 2 乗の期待値

$$V(X)=\sigma^2=E((X-\mu)^2)=\begin{cases}\displaystyle\sum_{i=1}^{n}(x_i-\mu)^2\,p(x_i) & (\text{離散型の場合})\\ \displaystyle\int_{-\infty}^{\infty}(x-\mu)^2 f(x)\,dx & (\text{連続型の場合})\end{cases}$$

を X の**分散 variance** という. 常に正または 0 であって, 0 となるのは分布が 1 点に集中するときに限る. 分散の非負の平方根 $\sigma=\sqrt{V(X)}$ を**標準偏差 (standard deviation)** という. 共に分布の "バラツキ" の程度を表す特性値である. なお, 標準偏差の単位は確率変数の単位と一致する.

$$\text{分散 (標準偏差) が大きい} \quad \Longleftrightarrow \quad \text{期待値の周りの散らばりが大きい}$$

分散について次が成り立つ.

$$V(X)=E(X^2)-[E(X)]^2 \tag{2.10}$$

2.4.3 歪度 (skewness), 尖度 (kurtosis)

偏差の 3 乗を相対化した値

$$\gamma_1=\frac{E\{(X-\mu)^3\}}{\sigma^3}$$

を確率変数 X の**歪度 (skewness)** という. 分布の "非対称度" を表す特性値である. 偏差の 4 乗を相対化した値

$$\gamma_2=\frac{E\{(X-\mu)^4\}}{\sigma^4}$$

を確率変数 X の**尖度 (kurtosis)** という. 分布の中央での "(非) 尖り具合", "分布の裾の長さ" を表す特性値である. γ_2-3 を尖度として定義する場合もある.

注意 歪度 γ_1, 尖度 γ_2 の定義における分子はそれぞれ次で計算される.

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

3 多変量確率分布 (multivariate probability distribution)

3.1 確率分布

簡単のため 2 変量連続型確率分布の場合を中心に考える. 2 つの連続型確率変数 X, Y の組 (X, Y) の標本空間を $\mathbb{R}^2$(平面上の点全体) とする. 適当な関数 $f(x,y)$ が存在して,

(1) $f(x,y)\geq 0, \quad (x,y)\in\mathbb{R}^2.$

(2) $\displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dx\,dy=1.$

が成立し，かつ任意の $a < b,\, c < d$ に対して

$$P(a < X < b,\, c < Y < d) = \int_a^b \int_c^d f(x,y)\,dx\,dy \tag{3.1}$$

であるとき，(X,Y) を **2 次元連続型確率変数 (2 次元連続型確率ベクトル) (two-dimensional continuous random variable)** といい，$f(x,y)$ を (X,Y) の**同時 (確率) 密度関数 (joint (probability) density function)** という．このとき，任意の 2 次元領域 E に対して事象 $\{(X,Y) \in E\}$ の確率が

$$P((X,Y) \in E)) = \iint_E f(x,y)\,dx\,dy$$

と表される． (1), (2) は $f(x,y)$ が 2 次元の確率密度関数であるための条件である．

任意の $a < b,\, c < d$ に対して

$$\begin{aligned}
P(a \le X \le b) &= P(a \le X \le b, -\infty < Y < \infty) = \int_a^b \int_{-\infty}^{\infty} f(x,y)\,dy\,dx \\
P(c \le Y \le d) &= P(-\infty < X < \infty, c \le Y \le d) = \int_c^d \int_{-\infty}^{\infty} f(x,y)\,dx\,dy
\end{aligned}$$

であり，

$$\begin{aligned}
f_1(x) &= \int_{-\infty}^{\infty} f(x,y)\,dy \\
f_2(y) &= \int_{-\infty}^{\infty} f(x,y)\,dx
\end{aligned}$$

とおくと

$$\begin{aligned}
P(a \le X \le b) &= \int_a^b f_1(x)\,dx \\
P(c \le Y \le d) &= \int_c^d f_2(y)\,dy
\end{aligned}$$

が成り立つ．$f_1(x), f_2(y)$ をそれぞれ同時密度関数 $f(x,y)$ より定まる X, Y の**周辺 (確率) 密度関数 (marginal (probability) density function)** という．このとき，明らかに次式が成り立つ．

$$\int_{-\infty}^{\infty} f_1(x)\,dx = \int_{-\infty}^{\infty} f_2(y)\,dy = \int_{-\infty}^{\infty}\int_{-\infty}^{\infty} f(x,y)\,dxdy = 1$$

すなわち，1 変量のときの密度関数の性質を持つ．

- X, Y の同時密度関数と周辺密度関数が，すべての x, y で

$$f(x,y) = f_1(x) f_2(y) \tag{3.2}$$

となるとき，X と Y は**独立 (independent)** であるといい，$X \perp\!\!\!\perp Y$ と表す．

X と Y が独立ならば，X のみで表される事象と Y のみで表される事象は，事象の独立性の意味で独立である．すなわち，

$$\begin{aligned}
P(a \le X \le b,\, c \le Y \le d) &= \int_a^b \int_c^d f(x,y)\,dy\,dx \\
&= \int_a^b f_1(x)\,dx \int_c^d f_2(y)\,dy \\
&= P(a \le X \le b) P(c \le Y \le d)
\end{aligned}$$

なので，事象 $A = \{a \le X \le b\}$ と $B = \{c \le Y \le d\}$ は独立である．

X と Y が独立でなければその間になんらかの関係があるが，その関係を与える密度関数を次のように定義する．$h > 0$ として条件付き確率は

$$\begin{aligned} P(a \le X \le b \mid c-h \le Y \le c+h) &= \frac{P(a \le X \le b,\, c-h \le Y \le c+h)}{P(c-h \le Y \le c+h)} \\ &= \frac{1}{2h}\int_a^b \int_{c-h}^{c+h} f(x,y)\,dy\,dx \Big/ \frac{1}{2h}\int_{c-h}^{c+h} f_2(y)\,dy \\ &\longrightarrow \int_a^b \frac{f(x,c)}{f_2(c)},dx,\ h \to 0 \end{aligned}$$

となる．

$$f_1(x|y) = \frac{f(x,y)}{f_2(y)} \tag{3.3}$$

とおくと，密度関数の定義を満たす.

密度関数 $f_1(x|y)$ を $Y = y$ という条件の下での X の**条件付密度関数 (conditional density function)** という．同じように，$X = x$ という条件の下での Y の条件付き密度関数は

$$f_2(y|x) = \frac{f(x,y)}{f_1(x)} \tag{3.4}$$

で与えられる．条件付き密度関数を考えることにより，X, Y の同時密度関数は

$$f(x,y) = f_1(x|y)f_2(y) = f_2(y|x)f_1(x) \tag{3.5}$$

と表される．また，X と Y の独立性の定義 (3.2) は

$$f_1(x|y) = f_1(x) \quad \text{または} \quad f_2(y|x) = f_2(y) \tag{3.6}$$

としもてよい.

多次元分布

n 個の確率変数 $Z_1, Z_2, ..., Z_n$ を準備し， これらを並べて一つの確率的変動をもつものとして定義した n 次元確率ベクトル $Z = (Z_1, Z_2, ..., Z_n)$ について考えよう（しばらく，記号 Z を用いる）．これが確率分布として同時密度関数 $f_Z(z_1, z_2, ..., z_n)$ をもつとする．ここでは連続型について考える（離散型も同様なので，その構成は読者に任せる）．このとき，多次元区間 $\prod_{i=1}^n J_i$ における確率は次で与えられる．

$$P\left(Z \in \prod_{i=1}^n J_i\right) = \int_{J_n}\int_{J_{n-1}} \cdots \int_{J_1} f_Z(z_1, z_2, ..., z_n)dz_1 \cdots dz_{n-1}dz_n$$

今，ベクトル Z を 2 つの部分ベクトル $Z = (X', Y')'$ に分ける．つまり，

$$Z = \begin{pmatrix} X \\ Y \end{pmatrix}$$

と置き，$X = (X_1, X_2, ..., X_l)'$ ，$Y = (Y_1, Y_2, ..., Y_m)'$ とすると，$n = l + m$ である．このとき，任意の多次元区間 $A = \prod_{i=1}^l J_i$ に対して

$$\begin{aligned} P(X \in A) &= P(X \in A, Y \in R^m) \\ &= P(Z \in A \times R^m) \\ &= \int_{A \times R^m} f_Z(x,y)dz \\ &= \int_A \left[\int_{R^m} f_Z(x,y)dy\right] dx \end{aligned}$$

が成り立つ．ただし，$z = (x, y)$, $x = (x_1, ..., x_l)$, $y = (y_1, ..., y_m)$, $dz = dxdy$, $dx = dx_1 \cdots dx_l$, $dy = dy_1 \cdots dy_m$ と略記する．X の密度関数は f_Z を $y = (y_1, ..., y_k)$ について積分した

$$f_X(x) = \int_{R^m} f_Z(x, y)dy$$

で得られることが分かる．これは X の周辺密度関数 (周辺分布) と呼ばれる．また，同時密度関数 f_Z を周辺密度関数 f_X で割った関数 $f_{Y|X}$ は，条件付密度関数を与える．すなわち

$$f_{Y|X}(y_1, \cdots, y_m | x_1, \cdots, x_l) = \frac{f_Z(z_1, \cdots, z_n)}{f_X(x_1, \cdots, x_l)}$$

は，$X = x$ という条件の下での Y の条件付密度関数を与える．次の同値関係がすぐに導かれる．

定理 3.1 以下の命題は同値である（確かめよ）．

(1) $X \perp\!\!\!\perp Y$（X と Y は独立）

(2) $\forall x \in R^l \ \forall y \in R^m \quad F_Z(x, y) = F_X(x)F_X(y)$

(3) $\forall x \in R^l \ \forall y \in R^m \quad f_Z(x, y) = f_X(x)f_X(y)$

3.2 2 次元確率変数 (分布) の特性値

(X, Y) を 2 次元の確率変数とする．X, Y の関数 $\varphi(X, Y)$ に対して

$$E\{\varphi(X,Y)\} = \begin{cases} \displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty} \varphi(x_i, y_j)\, p(x_i, y_j) & \text{(離散的な場合)} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty} \varphi(x, y)\, f(x, y)\, dx\, dy & \text{(連続的な場合)} \end{cases}$$

を確率変数 $\varphi(X, Y)$ の**期待値 (expectation)** という．ただし，$p(x_i, y_j) = P\{((X, Y) = (x_i, y_j)\}\,; i = 1, 2, \cdots, \ j = 1, 2, \cdots,$ である．

偏差の積の期待値

$$\begin{aligned} Cov(X, Y) = \sigma_{xy} &= E((X - \mu_x)(Y - \mu_y)) \\ &= \begin{cases} \displaystyle\sum_{i=1}^{\infty}\sum_{j=1}^{\infty}(x_i - \mu_x)(y_j - \mu_y)p(x_i, y_j) & \text{離散的な場合} \\ \displaystyle\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x - \mu_x)(y - \mu_y)f(x, y)\, dx\, dy & \text{連続的な場合} \end{cases} \end{aligned}$$

を X と Y の**共分散 (covariance)** という．ここで，$\mu_x = E(X)$, $\mu_y = E(Y)$ とする．また X, Y の標準偏差をそれぞれ σ_x, σ_y とするとき

$$\rho_{xy} = \frac{\sigma_{xy}}{\sigma_x \sigma_y}$$

を X と Y の**相関係数 (correlation coefficient)** という．常に $-1 \leq \rho_{xy} \leq 1$ である．$\rho_{xy} = 0$ のとき, X と Y は**無相関 (uncorrelated)** であるという．共分散の単位は X と Y の単位の積であり, 相関係数は無名数 (単位がない数) である．

条件付分布による期待値を次のように定義する．

X, Y の関数 $\varphi(X, Y)$ に対して

$$E(\varphi(X,Y)|Y=y) = \begin{cases} \displaystyle\sum_{i=1}^{\infty} \varphi(x_i, y)\, p_1(x_i|y) & \text{(離散的な場合)} \\ \displaystyle\int_{-\infty}^{\infty} \varphi(x, y)\, f_1(x|y)\, dx & \text{(連続的な場合)} \end{cases}$$

を $Y = y$ のときの $\varphi(X, Y)$ の**条件付き期待値 (conditional expectation)** という．ただし，$p_1(x_i|y_j) = p(x_i, y_j)/p_2(y_j)$ である．特に，$E(X|Y=y)$ を X の Y への**回帰関数 (regression function)** という．$X = x$ のときの条件付き期待値や，Y の X への回帰関数は X と Y の役割を交換すればよい．

回帰関数は X が平均的に Y にどのように関係しているかを表している．X と Y が独立ならば，$E(X|Y = y) = E(X)$ となることは明らかである．次の定理は，条件付き分布が簡単な場合には，条件付き期待値を考えれば期待値の計算が便利になることを示している．証明は簡単であるので，各自に任せる．

定理 3.2

$$E(\varphi(X,Y)) = E[E(\varphi(X,Y)|Y)]$$

が成り立つ．ここで，$E(\varphi(X,y)|y) = E(\varphi(X,Y)|Y=y)$ は y の関数となる．特に，$\varphi(x,y) = \varphi_1(x)\varphi_2(y)$ ならば

$$E(\varphi_1(X)\varphi_2(Y)) = E[\varphi_2(Y)E(\varphi_1(X)|Y)]$$

が成り立つ．

また，確率ベクトル Z の期待値ベクトルを次で定義する．

$$E(Z) = (E(Z_1), E(Z_2), ..., E(Z_n))'$$

これは Z の平均ベクトルとも呼ばれる．明らかに，

$$E(Z) = (E(X)', E(Y)')'$$

また，Z の要素である Z_i，Z_j に対して，

$$\begin{aligned} Cov(Z_i, Z_j) &= \sigma_{ij} \\ &= E(Z_i - E(Z_i))(Z_j - E(Z_j)) \\ &= E(Z_i Z_j) - E(Z_i)E(Z_j) \end{aligned}$$

を定義する．$i = j$ の時は Z_i の**分散**なので，$\sigma_i^2 = \sigma_{ii}$ とも書かれる．$i \neq j$ の時は Z_i と Z_j の**共分散**と呼ばれる．これらを行列に並べた

$$V(Z) = \Sigma = \begin{pmatrix} \sigma_{11} & \cdots & \sigma_{1n} \\ \vdots & \ddots & \vdots \\ \sigma_{n1} & \cdots & \sigma_{nn} \end{pmatrix}$$

は Z の**分散 (共分散) 行列**と呼ばれる．ところで，$W = (W_{ij})$ を $n \times m$ **確率行列** (確率変数を要素に持つ行列) とする．これは nm 次元のベクトルと考えられるので，上と同様に，平均を

$$EW = \begin{pmatrix} EW_{11} & \cdots & EW_{1m} \\ \vdots & \ddots & \vdots \\ EW_{n1} & \cdots & EW_{nm} \end{pmatrix}$$

で定義する．分散行列はその定義から明らかなように次のように表現できる．

$$V(Z) = E(Z - E(Z))(Z - E(Z))' = E(ZZ') - E(Z)E(Z)'$$

定理 3.3 $a \in R^n$, $b \in R^n$, A を $k \times n$ 行列, B を $m \times l$ 行列とするとき次が成り立つ (**レポート問題 1**).

(1) $E(a'Z) = a'E(Z)$

(2) $E(AZ) = AE(Z)$

(3) $E(AWB) = A(E(W))B$

(4) $Cov(a'Z, b'Z) = a'V(Z)b$

(5) $V(AZ) = AV(Z)A'$

$Z = (X', Y')$ の場合,

$$Cov(X, Y) = E(X - EX)(Y - EY)'$$

と定義して, X と Y の共分散行列と呼ぶことにすると, 明らかに次が成り立つ.

$$V(Z) = \begin{pmatrix} V(X) & Cov(X, Y) \\ Cov(Y, X) & V(Y) \end{pmatrix}$$

a を任意の n 次元ベクトルとすると,

$$a'V(Z)a = V(a'Z) \geq 0$$

が成り立つので, 分散行列は非負定値行列である.

3.3 2 変量正規分布 (bivariate normal distribution)

(X, Y) を 2 次元の連続型確率変数とする. (X, Y) 同時確率密度関数が

$$f(x, y) = \frac{1}{2\pi\sigma_x\sigma_y\sqrt{1-\rho^2}} e^{-Q(x,y)/2} \quad (-\infty < x < \infty,\ -\infty < y < \infty), \tag{3.7}$$

$$Q(x, y) = \frac{1}{1-\rho^2}\left\{\frac{(x-\mu_x)^2}{\sigma_x^2} - 2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y} + \frac{(y-\mu_y)^2}{\sigma_y^2}\right\} \tag{3.8}$$

で与えられるとき, (X, Y) は **2 次元正規分布 (bivariate normal distribution)** に従うといい

$$(X, Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$$

で表す. $\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho$ が 2 次元正規分布のパラメータであり, $\sigma_x > 0, \sigma_y > 0, -1 < \rho < 1$ である.

定理 3.4 (2 次元正規分布の周辺分布) $(X, Y) \sim N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ のとき

$$X \sim N(\mu_x, \sigma_x^2), \quad X \sim N(\mu_y, \sigma_y^2)$$

(証明) (3.8) より

$$Q(x, y) = \frac{1}{\sigma_y^2(1-\rho^2)}\left[y - \left\{\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x - \mu_x)\right\}\right]^2 - \frac{(x-\mu_x)^2}{\sigma_x^2}$$

だから, X の周辺確率密度関数 $f_1(x)$ は

$$\begin{aligned} f_1(x) &= \int_{-\infty}^{\infty} f(x, y)\, dy \\ &= \frac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right\}\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma_y\sqrt{1-\rho^2}} e^{-Q^*(y)/2}\, dy, \qquad (3.9) \\ Q^*(y) &= \frac{1}{\sigma_y^2(1-\rho^2)}\left[y - \left\{\mu_y + \rho\frac{\sigma_y}{\sigma_x}(x-\mu_x)\right\}\right]^2 \end{aligned}$$

ここで (3.9) における被積分関数は

$$N\left(\mu_y+\rho\frac{\sigma_y}{\sigma_x}(x-\mu_x),\sigma_y^2(1-\rho^2)\right)$$

の確率密度関数とみなせるから，その積分値は 1 となる．したがって

$$f_1(x)=\frac{1}{\sqrt{2\pi}\sigma_x}\exp\left\{-\frac{(x-\mu_x)^2}{2\sigma_x^2}\right\}$$

を得る．$f_1(x)$ は $N(\mu_x,\sigma_x^2)$ の確率密度関数だから $X\sim N(\mu_x,\sigma_x^2)$．同様にして $Y\sim N(\mu_y,\sigma_y^2)$ も得られる．□

例題 3.1 $(X,Y)\sim N(\mu_x,\mu_y,\sigma_x^2,\sigma_y^2,\rho)$ のとき，X と Y の共分散 $Cov(X,Y)$，相関係数 ρ_{xy} を求めよ．

解 定理 3.4 より

$$\begin{aligned}E(X)&=\mu_x,\quad V(X)=\sigma_x^2,\\ E(Y)&=\mu_y,\quad V(Y)=\sigma_y^2\end{aligned}$$

したがって

$$\begin{aligned}Cov(X,Y)&=E[(X-E(X)(Y-E(Y)]\\ &=\int_{-\infty}^{\infty}\int_{-\infty}^{\infty}(x-\mu_x)(y-\mu_y)\frac{1}{2\pi\sqrt{1-\rho^2}}e^{-Q(x,y)/2}\,dx\,dy \qquad (3.10)\\ Q(x,y)&=\frac{1}{1-\rho^2}\left\{\frac{(x-\mu_x)^2}{\sigma_x^2}-2\rho\frac{(x-\mu_x)(y-\mu_y)}{\sigma_x\sigma_y}+\frac{(y-\mu_y)^2}{\sigma_y^2}\right\}\end{aligned}$$

ここで

$$\frac{x-\mu_x}{\sigma_x}=u,\quad \frac{y-\mu_y}{\sigma_y}=v$$

とおくと

$$\frac{\partial(x,y)}{\partial(u,v)}=\begin{vmatrix}\dfrac{\partial x}{\partial u}&\dfrac{\partial x}{\partial v}\\ \dfrac{\partial y}{\partial u}&\dfrac{\partial y}{\partial v}\end{vmatrix}=\begin{vmatrix}\sigma_x&0\\0&\sigma_y\end{vmatrix}=\sigma_x\sigma_y$$

$$\begin{aligned}\therefore\quad Cov(X,Y)&=\sigma_x\sigma_y\int_{-\infty}^{\infty}u\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\int_{-\infty}^{\infty}v\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left\{-\frac{(v-\rho u)^2}{1-\rho^2}\right\}dv\,du\\ &=\sigma_x\sigma_y\rho\int_{-\infty}^{\infty}u^2\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du=\sigma_x\sigma_y\rho\end{aligned}$$

ここで

$$\int_{-\infty}^{\infty}v\frac{1}{\sqrt{2\pi}\sqrt{1-\rho^2}}\exp\left\{-\frac{(v-\rho u)^2}{1-\rho^2}\right\}dv=E(V)=\rho u,\quad V\sim N(\rho u,1-\rho^2)$$

$$\int_{-\infty}^{\infty}u^2\frac{1}{\sqrt{2\pi}}e^{-u^2/2}\,du=E(U^2)=1,\quad U\sim N(0,1)$$

を用いた．

$$\rho_{xy}=\frac{Cov(X,Y)}{\sqrt{V(X)}\sqrt{V(Y)}}=\frac{\sigma_x\sigma_y\rho}{\sigma_x\sigma_y}=\rho$$

□

一般の多変量正規分布は，(3.7) 式の $f(x,y)$ における楕円の式 (3.8) を多変量に拡張して与えられる．

$\boldsymbol{X}=(X_1,\cdots,X_k)'$ が k 変量正規分布に従うとすると，平均ベクトル $\boldsymbol{\mu}=(\mu_1,\cdots,\mu_k)'$ と分散共分散行列 $\Sigma=(\sigma_{ij})$ により定まる分布であり，同時確率密度関数は

$$f(\boldsymbol{x})=\frac{1}{\sqrt{|2\pi\Sigma|}}\exp\{-\frac{1}{2}(\boldsymbol{x}-\boldsymbol{\mu})'\Sigma^{-1}(\boldsymbol{x}-\boldsymbol{\mu})\}$$

である．ここで，” ′ ”は転置の記号とする．

3.4 確率過程

確率空間 $(\Omega,\mathcal{A},P)$ のほかに時間 t という空間 $\mathcal{T}$ を考え，時間で添え字付けられた確率変数 X_t を**確率過程** **(stochastic process)** または**時系列** **(time series)** という．無限個の確率変数の集合として

$$\boldsymbol{X}=\{X_t\mid t\in\mathcal{T}\}=\{X_t(\omega)\mid t\in\mathcal{T},\omega\in\Omega\}$$

と表したり，t の関数であることを強調して

$$\boldsymbol{X}=\{X(t)\mid t\in\mathcal{T}\}=\{X(t,\omega)\mid t\in\mathcal{T},\omega\in\Omega\}$$

と表し，**見本関数** **(sample function)** という．特に，有限集合 $\mathcal{T}=\{1,2,\cdots,n\}$ のときは $\boldsymbol{X}=(X_1,\cdots,X_n)$ で n 次元確率ベクトルとなり，離散時間 $\mathcal{T}=\{1,2,\cdots\}$(自然数) のときは $\boldsymbol{X}=\{X_1,X_2,\cdots\}$ で確率変数列となる．

時間についての任意の有限集合 $\{t_1,\cdots,t_n\}$ に対する $(X_{t_1},\cdots,X_{t_n})$ の同時分布関数

$$F_{t_1\cdots t_n}(x_1,\cdots,x_n)=P(X_{t_1}\leq x_1,\cdots,X_{t_n}\leq x_n)$$

を**有限次元分布** **(finite dimensional distribution)** といい，確率過程 $\boldsymbol{X}$ の分布を決定する．2 次までのモーメントとして次のものが定義される．

平均値関数 **(mean function)**：$\mu_t=E(X_t)$,

分散関数 **(variance function)**：$\sigma_t=V(X_t)$,

共分散関数 **(covariance function)**：$R_{st}=Cov(X_s,X_t)$.

別の確率過程 $\boldsymbol{Y}$ を考えるとき，$R_{XY}(s,t)=Cov(X_s,Y_t)$. を**相互共分散関数** **(cross-covariance function)** といい，それに対して，$R_{st}=Cov(X_s,X_t)$ を**自己共分散関数** **(auto-covariance function)** という．

確率過程の例として，**ガウス過程** **(Gaussian process)** を紹介する．

確率過程 $\boldsymbol{X}$ の任意の有限次元分布が正規分布であるとき，$\boldsymbol{X}$ を**ガウス過程** **(Gaussian process)** または**正規過程** **(normal process)** という．正規過程の代表的なものとしてウイナー過程を考える．時間空間は非負の実数空間とする ($\mathcal{T}=\mathbb{R}^+$)．次の性質を満たす確率過程 $\boldsymbol{W}=\{W(t)\mid t\in\mathbb{R}^+\}$ を**ウィナー過程** **(Wiener process)** または**ブラウン運動** **(Brownian motion)** という．

1. 始点固定：$W(0)=0$.
2. 正規性：$W(t)-W(s)$ $(s\leq t)$ は正規分布 $N(0,t-s)$ に従う．
3. 独立増分性：任意の分割 $t_1<t_2<t_3<\cdots<t_n$ に対して，変動の増分 $W(t_2)-W(t_1),W(t_3)-W(t_2),\cdots,W(t_n)-W(t_{n-1})$ は独立である．

平均値関数：$\mu_t=E\{W(t)\}=E\{N(0,t)\}=0$,

分散関数：$V\{W(t)\}=V\{N(0,t)\}=t$,

共分散関数：$R_{st}=E\{W(s)W(t)\}=\min(s,t)$ である．

レポート問題２． 以下の各問に答えよ．

確率変数 X, Y の同時確率分布 $P\,(X = i, Y = j)$, $i = 0, 2, 3\,;\ j = -1, 0, 1, 3,$ が次の表のように与えられている．

$X \setminus Y$	-1	0	1	3
0	0.12	0.1	0.1	0.08
2	0.1	0.1	0.05	0.04
3	0.08	0.1	0.1	0.03

(1) X, Y の周辺確率関数をそれぞれ求めよ．

(2) X と Y が独立かどうかを調べよ．

(3) $Y = 1$ が与えられたときの X の条件付確率関数を求めよ．またこのとき，$P(X \leq 2 \,|\, Y = 1)$ の値を求めよ．

(4) $E(X|Y = 1)$, $Var(X|Y = 1)$ の値を求めよ．

練習問題２．

1. 確率変数 X, Y の同時密度関数が

$$f(x, y) = cx^3 \exp\{-x(1 + y)\}, \quad 0 < x,\, y < \infty$$

であるとき，次の各問に答えよ．

(1) 定数 c の値を求めよ．

(2) X, Y の周辺密度関数 $f_1(x), f_2(y)$ を求めよ．

(3) X, Y の平均，分散を求めよ．X と Y の相関係数を求めよ．

2. 確率変数 X, Y の同時密度関数が

$$f(x, y) = \frac{1}{\alpha\beta} \exp\{-\frac{x}{\alpha} - \frac{y - x}{\beta}\}, \quad 0 < x \leq y < \infty$$

であるとする．ここで，α, β は，α, $\beta > 0$, $\alpha \neq \beta$ なる定数である．次の各問に答えよ．

(1) X, Y の周辺密度関数 $f_1(x), f_2(y)$ を求めよ．

(2) $X = x$ が与えられたときの Y の条件付密度関数 $f_2(y|x)$ を求めよ．

(3) X, Y の平均，分散を求めよ．X と Y の相関係数を求めよ．

3. 区間 $[0, 1]$ からランダムに 1 点をとりその点の座標を X とする．次に，区間 $[X, 1]$ からランダムに 1 点をとりその点の座標を Y とする．このとき，(X, Y) の同時分布を求めよ．x と Y それぞれの平均と分散，X と Y の相関係数を求めよ．

4. (X, Y) は 2 変量正規分布 $N(\mu_x, \mu_y, \sigma_x^2, \sigma_y^2, \rho)$ に従うとする．

(1) X の周辺分布は，平均 μ_x, 分散 σ_x^2 の正規分布 $N(\mu_x, \sigma_x^2)$ に従うことを示せ．

(2) $X = x$ を与えたときの Y の条件付分布を求めよ．

4 母関数 (generating function)

確率分布の特性値，期待値や分散などを定義に従って計算するのは面倒なことが多い．簡単な計算を可能にし，また独立な確率変数の和の分布などを求めるときなど，確率論の基本的な道具として広く用いられる母関数について紹介する．

4.1 確率母関数 (probability generating function)

X を離散型確率変数とし，標本空間 R_X を $R_X = \{0, 1, \ldots, n, \ldots\}$ とする． θ を実変数として

$$P(\theta) = E(\theta^X) = \sum_{x=0}^{\infty} \theta^x P(X = x) \tag{4.11}$$

で定義される関数 $P(\theta)$ を X の**確率母関数 (probability generating function)** という．$P(\theta)$ の θ についての 1 次, 2 次の導関数を計算すると

$$\begin{aligned} P'(\theta) &= \sum_{x=0}^{\infty} x\,\theta^{x-1} P(X = x) \\ P''(\theta) &= \sum_{x=0}^{\infty} x(x-1)\,\theta^{x-2} P(X = x) \end{aligned}$$

となる．ここで， $\theta = 1$ とおくと

$$\begin{aligned} P(1) &= \sum_{x=0}^{\infty} P(X = x) = 1 \\ P'(1) &= \sum_{x=0}^{\infty} xP(X = x) = E(X) \\ P''(1) &= \sum_{x=0}^{\infty} x(x-1)P(X = x) = E(X(X-1)) \end{aligned}$$

となる．分散の計算式

$$V(X) = E(X^2) - [E(X)]^2$$

において

$$E(X^2) = E(X(X-1)) + E(X)$$

となるから, $P(\theta)$ が求まるならば, これを微分することにより期待値や分散が簡単に計算できることになる.

例題 4.1 X が二項分布 $B(n, p)$ に従うとき, X の確率母関数を求め, これを用いて二項分布 $B(n, p)$ の期待値, 分散を求めよ.

解 $X \sim B(n, p)$ のとき, $R_X = \{0, 1, \ldots, n\}$

$$P(X = x) = \binom{n}{x} p^x q^{n-x} \quad (q = 1 - p\,; x = 0, 1, \ldots, n)$$

だから

$$\begin{aligned} P(\theta) &= \sum_{x=0}^{n} \theta^x P(X = x) = \sum_{x=0}^{n} \theta^x \binom{n}{x} p^x q^{n-x} \\ &= \sum_{x=0}^{n} \binom{n}{x} (p\,\theta)^x q^{n-x} = (p\,\theta + q)^n \end{aligned}$$

を得る．ここで，上式の最後の等式は二項定理による．$P(\theta)$ の θ についての 1 次と 2 次の導関数を求めると

$$\begin{aligned} P'(\theta) &= np(p\theta+q)^{n-1}, \\ P''(\theta) &= n(n-1)p^2(p\theta+q)^{n-2} \end{aligned}$$

となる．これより

$$\begin{aligned} E(X) &= P'(1) = np \\ V(X) &= E(X(X-1)) + E(X) - [E(X)]^2 \\ &= P''(1) + P'(1) - (P'(1))^2 \\ &= n(n-1)p^2 + np - (np)^2 = np(1-p) = npq \end{aligned}$$

4.2 積率 (モーメント)(moment)

X を確率変数，r を自然数とするとき

$$\mu_r' = E(X^r) \quad (r = 1, 2, \ldots)$$

を X の**原点のまわりの r 次の積率 (モーメント) (r-th moment about origin)** という．μ_1' は X の期待値である．また

$$\mu_r = E\left\{(X-\mu_1')^r\right\} \quad (r = 1, 2, \ldots)$$

を X の**期待値のまわりの r 次の積率 (モーメント) (r-th moment about mean)** という． μ_2 は X の分散である． X が離散型の確率変数で

$$R_X = \{x_1, x_2, \ldots\}, \quad p(x_i) = P(X = x_i)\ (i = 1, 2, \ldots)$$

のとき

$$\begin{aligned} \mu_r' &= E(X^r) = \sum_{i=1}^{\infty} x_i^r p(x_i) \\ \mu_r &= E((X-\mu_1')) = \sum_{i=1}^{\infty} (x_i - \mu_1')^r p(x_i) \end{aligned}$$

また，X が連続型確率変数で，$f(x)$ をその確率密度関数とするとき

$$\begin{aligned} \mu_r' &= E(X^r) = \int_{-\infty}^{\infty} x^r f(x)\,dx \\ \mu_r &= E((X-\mu_1')^r) = \int_{-\infty}^{\infty} (x-\mu_1')^r f(x)\,dx \end{aligned}$$

確率変数 X の期待値のまわりの r 次の積率 μ_r と原点のまわりの r 次の積率を μ_r' との間に，次の関係が成り立つ.

定理 4.1

$$\mu_r = \mu_r' - \binom{r}{1}\mu_{r-1}'\mu_1' + \binom{r}{2}\mu_{r-2}'(\mu_1')^2 - \cdots + (-1)^{\ell}\binom{r}{\ell}\mu_{r-\ell}'(\mu_1')^{r-\ell} + \cdots + (-1)^r(\mu_1')^r \qquad (4.12)$$

(証明) 二項定理により

$$\begin{aligned}(X-\mu_1')^r \;=\; \{X+(-\mu_1')\}^r = X^r - \binom{r}{1}X^{r-1}\mu_1' + \binom{r}{2}X^{r-2}(\mu_1')^2 - \cdots \\ + (-1)^\ell\binom{r}{\ell}X^{r-\ell}(\mu_1')^\ell + \cdots + (-1)^r(\mu_1')^r\end{aligned}$$

したがって,

$$\begin{aligned}\mu_r \;&=\; E((X-\mu_1')^r) \\ &=\; E(X^r) - \binom{r}{1}E(X^{r-1})\mu_1' + \binom{r}{2}E(X^{r-2})(\mu_1')^2 - \cdots + (-1)^\ell\binom{r}{\ell}E(X^{r-\ell})(\mu_1')^\ell + \cdots + (-1)^r(\mu_1')^r \\ &=\; \mu_r' - \binom{r}{1}\mu_{r-1}'\,\mu_1' + \binom{r}{2}\mu_{r-2}'\,(\mu_1')^2 - \cdots \quad + (-1)^\ell\binom{r}{\ell}\mu_{r-\ell}'\,(\mu_1')^\ell + \cdots + (-1)^r(\mu_1')^r\end{aligned}$$

ここで, 期待値の線形性

$$E(a_1X_1 + a_2X_2 + \cdots + a_nX_n + b) = a_1E(X_1) + a_2E(X_2) + \cdots + a_nE(X_n) + b$$

(ただし, a_i, b は定数) を用いた. □

とくに, $r = 2, 3, 4$ に対して (4.12) 式は次で表される.

$$\mu_2 \;=\; \mu_2' - (\mu_1')^2 \tag{4.13}$$

$$\mu_3 \;=\; \mu_3' - 3\mu_2'\mu_1' + 2(\mu_1')^3 \tag{4.14}$$

$$\mu_4 \;=\; \mu_4' - 4\mu_3'\mu_1' + 6\mu_2'(\mu_1')^2 - 3(\mu_1')^4 \tag{4.15}$$

また, 3 章で与えた歪度 γ_1 および尖度 γ_2 は積率を用いると次のように表せる.

歪度・尖度

$$\gamma_1 \;=\; \frac{E((X-\mu)^3)}{\sigma^3} = \frac{\mu_3}{(\mu_2)^{3/2}} \tag{4.16}$$

$$\gamma_2 \;=\; \frac{E((X-\mu)^4)}{\sigma^4} = \frac{\mu_4}{(\mu_2)^2} \tag{4.17}$$

4.3 積率 (モーメント) 母関数 (moment generating function)

本章 1 節で, 離散型確率分布の期待値・分散が確率母関数を用いることにより容易に求められることを述べた. 本節では離散型のみならず連続型の分布に対しても定義できる母関数について考える.

X を確率変数, θ を実変数とするとき

$$M(\theta) = E(e^{\theta X})$$

を確率変数 X の**積率 (モーメント) 母関数 (moment generating function)** (**m.g.f.** と略記) という. X が離散型の確率変数で

$$R_X = \{x_1, x_2, \ldots\}, \quad p(x_i) = P(X = x_i)\ (i = 1, 2, \ldots)$$

のとき

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

また, X が連続型確率変数で, $f(x)$ をその確率密度関数とするとき

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

この積率母関数については次が成り立つ.

定理 4.2 確率変数 X の積率母関数を $M(\theta)$, a, b を定数とするとき, 確率変数 $aX + b$ の積率母関数は

$$\psi(\theta) = e^{b\theta} M(a\theta)$$

で与えられる.
(証明)

$$\psi(\theta) = E(e^{\theta(aX+b)}) = e^{b\theta} E(e^{(a\theta)X}) = e^{b\theta} M(a\theta) \qquad \square$$

定理 4.3 互いに独立な n 個の確率変数 $X_i\ (i = 1, 2, \ldots, n)$ の積率母関数を $M_i(\theta)$, $a_i\ (i = 1, 2, \ldots, n)$ を定数とするとき, 確率変数 $a_1X_1 + a_2X_2 + \ldots + a_nX_n + b$ の積率母関数は

$$\psi(\theta) = e^{b\theta} M_1(a_1\theta) M_2(a_2\theta) \cdots M_n(a_n\theta)$$

で与えられる.
(証明)

$$\begin{aligned}
\psi(\theta) &= E(e^{\theta(a_1X_1 + a_2X_2 + \cdots + a_nX_n + b)}) \\
&= e^{b\theta} E(e^{\theta(a_1X_1)} e^{\theta(a_2X_2)} \cdots e^{\theta(a_nX_n)}) \\
&= e^{b\theta} E(e^{\theta(a_1X_1)}) E(e^{\theta(a_2X_2)}) \cdots E(e^{\theta(a_nX_n)}) \\
&= e^{b\theta} M_1(a_1\theta) M_2(a_2\theta) \cdots M_n(a_n\theta) \qquad \square
\end{aligned}$$

ここで, $X_1, X_2, \ldots, X_n$ が互いに独立であることから, $a_1X_1, a_2X_2, \ldots, a_nX_n$ も互いに独立であることを用いた.

定理 4.4 $|\theta| \leq \theta_0$ (θ_0 はある正の定数) に対して $M(\theta)$ が存在するとき, X の原点のまわりの積率 $\mu'_r\ (r = 1, 2, \ldots)$ はすべて存在し

$$M(\theta) = 1 + \frac{\mu'_1}{1!}\,\theta + \frac{\mu'_2}{2!}\,\theta^2 + \cdots + \frac{\mu'_r}{r!}\,\theta^r + \cdots \quad (|\theta| \leq \theta_0) \tag{4.18}$$

が成り立つ. したがって

$$M^{(r)}(0) = \mu'_r \quad (r = 1, 2, \ldots) \tag{4.19}$$

がいえる.

この定理が積率母関数の名のゆえんである. 定理 4.4 の数学的に厳密な証明は他書に譲り[1], ここでは直感的に導いてみよう.

指数関数 e^x のマクローリン展開

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \cdots + \frac{1}{r!}x^r + \cdots$$

より

$$\begin{aligned}
M(\theta) = E(e^{\theta X}) &= E\left(1 + \frac{1}{1!}\theta X + \frac{1}{2!}(\theta X)^2 + \cdots + \frac{1}{r!}(\theta X)^r + \cdots\right) \\
&= 1 + \frac{E(X)}{1!}\theta + \frac{E(X^2)}{2!}\theta^2 + \cdots + \frac{E(X^r)}{r!}\theta^r + \cdots \\
&= 1 + \frac{\mu'_1}{1!}\,\theta + \frac{\mu'_2}{2!}\,\theta^2 + \cdots + \frac{\mu'_r}{r!}\,\theta^r + \cdots
\end{aligned}$$

[1]例えば, 国沢清典編 確率統計演習 1 確率

が成り立つ. (4.19) 式は (4.18) 式の両辺を θ で微分した式に $\theta = 0$ を代入することを順次繰り返して得られる.

例題 4.2 X が正規分布 $N(\mu, \sigma^2)$ に従うとき, X の積率母関数を求め, これを用いて正規分布 $N(\mu, \sigma^2)$ の期待値, 分散, 歪度, 尖度を求めよ.

解 $X \sim N(\mu, \sigma^2)$ のとき, X の確率密度関数は

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

だから, その積率母関数は

$$\begin{aligned}
M(\theta) &= E(e^{\theta X}) = \int_{-\infty}^{\infty} e^{\theta x} f(x)\,dx \\
&= \int_{-\infty}^{\infty} e^{\theta x}\frac{1}{\sqrt{2\pi}\sigma}\exp\left\{-\frac{(x-\mu)^2}{2\sigma^2}\right\}dx = \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}e^{-Q}\,dx \\
Q &= \frac{(x-\mu)^2}{2\sigma^2} - \theta x = \frac{1}{2\sigma^2}\{x^2 - 2(\mu+\sigma^2\theta)x + \mu^2\} \\
&= \frac{1}{2\sigma^2}[\{x-(\mu+\sigma^2\theta)\}^2 - (\mu+\sigma^2\theta)^2 + \mu^2] \\
&= \frac{1}{2\sigma^2}\{x-(\mu+\sigma^2\theta)\}^2 - \left(\mu\theta + \frac{\sigma^2}{2}\theta^2\right) \\
\therefore\quad M(\theta) &= \int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\{x-(\mu+\sigma^2\theta)\}^2 + \left(\mu\theta+\frac{\sigma^2}{2}\theta^2\right)\right]dx \\
&= \exp\left(\mu\theta+\frac{\sigma^2}{2}\theta^2\right)\int_{-\infty}^{\infty}\frac{1}{\sqrt{2\pi}\sigma}\exp\left[-\frac{1}{2\sigma^2}\{x-(\mu+\sigma^2\theta)\}^2\right]dx
\end{aligned}$$

ここで最後の積分における被積分関数は $N(\mu+\sigma^2\theta, \sigma^2)$ の確率密度関数と考えられるから, その積分の値は 1 に等しい.

$$\therefore\quad M(\theta) = \exp\left(\mu\theta + \frac{\sigma^2}{2}\theta^2\right)$$

指数関数の Maclaurin 展開

$$e^x = 1 + \frac{1}{1!}x + \frac{1}{2!}x^2 + \frac{1}{3!}x^3 + \cdots + \frac{1}{r!}x^r + \cdots$$

を用いて $M(\theta)$ を展開すると

$$\begin{aligned}
M(\theta) &= \exp\left(\mu\theta+\frac{\sigma^2}{2}\theta^2\right) = \exp(\mu\theta)\exp\left(\frac{\sigma^2}{2}\theta^2\right) \\
&= \left\{1 + \frac{1}{1!}\mu\theta + \frac{1}{2!}(\mu\theta)^2 + \frac{1}{3!}(\mu\theta)^3 + \frac{1}{4!}(\mu\theta)^4 + \cdots\right\}\cdot\left\{1 + \frac{1}{1!}\frac{\sigma^2}{2}\theta^2 + \frac{1}{2!}\left(\frac{\sigma^2}{2}\theta^2\right)^2 + \cdots\right\} \\
&= 1 + \frac{1}{1!}\mu\theta + \frac{1}{2!}(\sigma^2+\mu^2)\theta^2 + \frac{1}{3!}(3\mu\sigma^2+\mu^3)\theta^3 + \frac{1}{4!}(3\sigma^4+6\mu^2\sigma^2+\mu^4)\theta^4 + \cdots
\end{aligned}$$

これより, 原点のまわりの r 次の積率 μ'_r $(r = 2, 3, 4)$ は

$$\begin{aligned}
E(X) = \mu'_1 &= \mu \\
\mu'_2 &= \sigma^2 + \mu^2 \\
\mu'_3 &= 3\mu\sigma^2 + \mu^3 \\
\mu'_4 &= 3\sigma^4 + 6\mu^2\sigma^2 + \mu^4
\end{aligned}$$

ゆえに, X の期待値のまわりの r 次の積率 μ_r $(r = 2, 3, 4)$ は (4.13), (4.14), (4.15) 式より

$$\begin{aligned} V(X) = \mu_2 &= \mu_2' - (\mu_1')^2 = (\sigma^2 + \mu^2) - \mu^2 = \sigma^2 \\ \mu_3 &= \mu_3' - 3\mu_2'\,\mu_1' + 2(\mu_1')^3 \\ &= (3\mu\sigma^2 + \mu^3) - 3(\sigma^2 + \mu^2)\mu + 2\mu^3 = 0 \\ \mu_4 &= \mu_4' - 4\mu_3'\,\mu_1' + 6\mu_2'\,(\mu_1')^2 - 3(\mu_1')^4 \\ &= (3\sigma^4 + 6\mu^2\sigma^2 + \mu^4) - 4(3\mu\sigma^2 + \mu^3)\mu + 6(\sigma^2 + \mu^2)\mu^2 - 3\mu^4 = 3\sigma^4 \end{aligned}$$

これより, X の歪度 γ_1, 尖度 γ_2 は (4.16) 式および (4.17) 式から

$$\begin{aligned} \gamma_1 &= \frac{\mu_3}{\mu_2^{3/2}} = \frac{0}{\sigma^3} = 0 \\ \gamma_2 &= \frac{\mu_4}{\mu_2^2} = \frac{3\sigma^4}{\sigma^4} = 3 \end{aligned}$$

例題 4.3 α, β をそれぞれ正の実数とする. 連続型確率変数 X の確率密度関数が

$$f(x) = \begin{cases} \dfrac{\beta^{-\alpha}}{\mathit{\Gamma}(\alpha)}\, x^{\alpha-1} e^{-x/\beta} & (x \geq 0) \\ 0 & (x < 0) \end{cases}$$

で与えられるとき, X は**ガンマ分布 (gamma distribution)** に従うといい, $X \sim Ga(\alpha, \beta)$ と表す. ここで, $\Gamma(\alpha) = \int_0^\infty e^{-x} x^{\alpha-1}\, dx$ はガンマ関数である. α, β がガンマ分布のパラメータであり, 密度関数の概形は α で定まり, α を shape parameter という. X の測定単位を変える (X を定数倍する) と β が変わり, β を scale parameter という.

X がガンマ分布 $G(\alpha, \beta)$ に従うとき, X の積率母関数を求め, これを用いてガンマ分布 $G(\alpha, \beta)$ の期待値, 分散, 歪度, 尖度を求めよ.

解 $X \sim G(\alpha, \beta)$ のとき

$$\begin{aligned} M(\theta) = E(e^{\theta X}) &= \int_0^\infty e^{\theta x} \frac{\beta^{-\alpha}}{\mathit{\Gamma}(\alpha)}\, x^{\alpha-1} e^{-x/\beta}\, dx \\ &= \frac{\beta^{-\alpha}}{\mathit{\Gamma}(\alpha)} \int_0^\infty x^{\alpha-1} e^{-(1-\beta\theta)x/\beta}\, dx \end{aligned}$$

$(1 - \beta\theta)x/\beta = u$ とおいて置換積分を行うと

$$\begin{aligned} M(\theta) &= \frac{\beta^{-\alpha}}{\mathit{\Gamma}(\alpha)} \int_0^\infty \left(\frac{\beta u}{1 - \beta\theta}\right)^{\alpha-1} e^{-u} \frac{\beta}{1 - \beta\theta}\, du \\ &= \frac{\beta^{-\alpha}}{\mathit{\Gamma}(\alpha)} \frac{\beta^\alpha}{(1 - \beta\theta)^\alpha} \int_0^\infty u^{\alpha-1} e^{-u}\, du = (1 - \beta\theta)^{-\alpha} \end{aligned}$$

ただし, $\theta < 1/\beta$. これから

$$\begin{aligned} \mu = E(X) &= M'(0) = \alpha\beta \\ E(X^2) &= M''(0) = \alpha(\alpha+1)\beta^2 \\ E(X^3) &= M'''(0) = \alpha(\alpha+1)(\alpha+2)\beta^3 \\ E(X^4) &= M^{(4)}(0) = \alpha(\alpha+1)(\alpha+2)(\alpha+3)\beta^4 \\ \therefore \quad \sigma^2 = V(X) &= E(X^2) - [E(X)]^2 = \alpha(\alpha+1)\beta^2 - (\alpha\beta)^2 = \alpha\beta^2 \end{aligned}$$

さらに

$$
\begin{aligned}
E((X-\mu)^3) &= E(X^3) - 3\mu E(X^2) + 2\mu^3 \\
&= \alpha(\alpha+1)(\alpha+2)\beta^3 - 3\alpha\beta\cdot\alpha(\alpha+1)\beta^2 + 2(\alpha\beta)^3 = 2\alpha\beta^3 \\
E((X-\mu)^4) &= E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 3\mu^4 \\
&= \alpha(\alpha+1)(\alpha+2)(\alpha+3)\beta^4 - 4\alpha\beta\cdot\alpha(\alpha+1)(\alpha+2)\beta^3 \\
&\quad + 6(\alpha\beta)^2\cdot\alpha(\alpha+1)\beta^2 - 3(\alpha\beta) = 3(\alpha^2\beta^4 + 2\alpha\beta^4)
\end{aligned}
$$

以上より

$$
\begin{aligned}
\gamma_1 &= \frac{E((X-\mu)^3)}{\sigma^3} = \frac{2\alpha\beta^3}{(\alpha\beta^2)^{3/2}} = \frac{2}{\sqrt{\alpha}} \\
\gamma_2 &= \frac{E((X-\mu)^4)}{\sigma^4} = \frac{3(\alpha^2\beta^4 + 2\alpha\beta^4)}{(\alpha\beta^2)^2} = 3 + \frac{6}{\alpha}
\end{aligned}
$$

とくに，$\alpha = k/2, \beta = 2$ のガンマ分布 $G_a(k/2, 2)$ は，**自由度 k の χ^2-分布 (chi-square distribution with k degrees of freedom)** といい，χ_k^2 と表す．このとき，上の結果に $\alpha = k/2, \beta = 2$ を代入して次を得る．

$$
\begin{aligned}
&X \sim \chi_k^2 \text{のとき} \\
&\quad M(\theta) = (1-2\theta)^{-k/2} \quad (\theta < 1/2), \\
&\quad E(X) = k, \quad V(X) = 2k, \\
&\quad \gamma_1 = \frac{2\sqrt{2}}{\sqrt{k}}, \quad \gamma_2 = 3 + \frac{12}{k}
\end{aligned}
$$

ここで積率母関数に関する定理を 2 つ証明なしにあげておく. 証明は脚注 1 を参照.

定理 4.5 確率変数 X, Y の積率母関数が ($|\theta| \le \theta_0$ (θ_0 はある正の定数) で存在して) 一致するときは, X, Y の分布関数は一致する.

定理 4.6 確率変数 $X_1, X_2, \ldots$ の分布関数および積率母関数をそれぞれ, $F_1(x), F_2(x), \ldots, M_1(\theta)$, $M_2(\theta), \ldots$ とし, 分布関数 $F(x)$ をもつ確率変数 X の積率母関数 $M(\theta)$ とする. $|\theta| \le \theta_0$ (θ_0 はある正の定数) に対して $M_n(\theta) \to M(\theta)$ $(n \to \infty)$ が成り立つならば

$$
F_n(x) \to F(x) \quad (n \to \infty)
$$

が ($F(x)$ の連続点で) 成り立つ. このとき確率変数列 $\{X_n\}$ は X に**法則収束する (converges in law)** または**分布収束する (converges in distribution)** という.

分布の再生性については, 積率母関数を使って簡単に証明することができる.

確率分布の再生性　X と Y は互いに独立であるとする.

(1) 二項分布の再生性

$$
X \sim B(m,p),\ Y \sim B(n,p) \text{ のとき}, \quad X+Y \sim B(m+n,p)
$$

(2) ポアソン分布の再生性

$$
X \sim Po(\lambda_x),\ Y \sim Po(\lambda_y) \text{ のとき}, \quad X+Y \sim Po(\lambda_x+\lambda_y)
$$

(3) 正規分布の再生性

$$X \sim N(\mu_x, \sigma_x^2),\ Y \sim N(\mu_y, \sigma_y^2) \text{ のとき,} \quad X+Y \sim N(\mu_x+\mu_y, \sigma_x^2+\sigma_y^2)$$

(4) カイ 2 乗分布の再生性

$$X \sim \chi_k^2,\ Y \sim \chi_l^2 \text{ のとき,} \quad X+Y \sim \chi_{k+l}^2$$

レポート問題 4.1 以下の各問に答えよ.

(1) 確率変数 X が二項分布 $B(n,p)$ に従うとき，X の積率母関数 $M_X(\theta)$ を求めよ．ここで，

$$P(X=k) = {}_nC_k\, p^k (1-p)^{n-k}, \quad k=0,1,\ldots,n.$$

(2) 確率変数 Y がポアソン分布 $Po(\lambda)$ に従うとき，Y の積率母関数 $M_Y(\theta)$ を求めよ．ここで，

$$P(Y=k) = \frac{\lambda^k e^{-\lambda}}{k!}, \quad k=0,1,\ldots.$$

(3) $np=\lambda$ の条件の下で，

$$\lim_{n\to\infty} M_X(\theta) = M_Y(\theta)$$

が成り立つことを示し，それが意味することを述べよ.

レポート問題 4.2 上に述べた確率分布の再生性を積率母関数を使って証明せよ.

[**参考**] (1) ラプラス変換 (Laplace transform) $L(\theta)$ は

$$L(\theta) = E\{\exp(-\theta X)\}$$

である。統計学では，$M(\theta)=L(-\theta)$ で定義される積率母関数を使う。積率母関数はどの分布についても必ず存在するわけではないが，次のように定義されるフーリエ変換 $\phi(t)$ は必ず存在する.

$$\phi(\theta) = E(e^{i\theta X}), \quad \text{ただし } i=\sqrt{-1}$$

これを統計学では X の特性関数という．積率母関数が存在するとき，特性関数は

$$\phi(\theta) = M(i\theta)$$

により求まる．また，確率母関数から積率母関数や特性関数が次のような関係で求まる.

$$M(\theta) = P(\exp(\theta)), \qquad \phi(\theta) = P(\exp(i\theta))$$

分布関数と特性関数は 1 対 1 であり，一方が求まれば他方も求まるという関係にある.

(2) 確率ベクトル $\boldsymbol{X} = (X_1,\cdots,X_n)'$ の積率母関数は，$\boldsymbol{\theta} = (\theta_1,\cdots,\theta_n)$ に対して

$$M(\boldsymbol{\theta}) = E\{\exp(\boldsymbol{\theta}\boldsymbol{X})\} = E\{\exp(\theta_1 X_1 + \cdots + \theta_n X_n)\}$$

で与えられる.

4.4 確率変数の関数の分布

X の密度関数を $f(x)$ とする．$y = g(x)$ は微分可能で，狭義単調関数であるとする． このとき，$Y = g(X)$ の確率密度関数は

$$h(y) = f\left(g^{-1}(y)\right)\left|\frac{1}{g'(g^{-1}(y))}\right|$$

である．

（例）対数正規分布 X が正規分布 $N(\mu, \sigma^2)$ に従っているとき，$Y = e^X$ の密度関数を求める．

$$h(y) = \frac{1}{\sqrt{2\pi}\sigma y}\exp\{-(\log y - \mu)^2/2\sigma^2\}, \quad y > 0$$

$\log Y$ は正規分布に従っているので，この確率分布は対数正規分布と呼ばれる．

(２つ以上の確率変数の関数の分布)

$X_1, \ldots, X_n$ と $Y_1, \ldots, Y_n$ が相互に１価な連続関数であり

$$\begin{cases} Y_1 &= \varphi_1(X_1, \ldots, X_n) \\ Y_2 &= \varphi_2(X_1, \ldots, X_n) \\ & \vdots \\ Y_n &= \varphi_n(X_1, \ldots, X_n) \end{cases}$$

で表されるとする．また，この変換の逆変換を

$$\begin{cases} X_1 &= \psi_1(Y_1, \ldots, Y_n) \\ X_2 &= \psi_2(Y_1, \ldots, Y_n) \\ & \vdots \\ X_n &= \psi_n(Y_1, \ldots, Y_n) \end{cases}$$

とする．このとき，$(X_1, \ldots, X_n)$ と $(Y_1, \ldots, Y_n)$ のそれぞれの同時密度関数 $f_X(x_1, \ldots, x_n)$ と $f_Y(y_1, \ldots, y_n)$ との間には

$$\begin{aligned} f_Y(y_1, \ldots, y_n) &= |J|\, f_X(x_1, \ldots, x_n) \\ &= |J|\, f_X(\psi_1(y_1, \ldots, y_n), \ldots, \psi_n(y_1, \ldots, y_n)) \end{aligned}$$

の関係がある．ここで，J はヤコビアンであり

$$\begin{aligned} J &= \frac{\partial(x_1, \ldots, x_n)}{\partial(y_1, \ldots, y_n)} \\ &= \begin{vmatrix} \dfrac{\partial\psi_1(y_1, \ldots, y_n)}{\partial y_1} & \cdots & \dfrac{\partial\psi_1(y_1, \ldots, y_n)}{\partial y_n} \\ \vdots & & \vdots \\ \dfrac{\partial\psi_n(y_1, \ldots, y_n)}{\partial y_1} & \cdots & \dfrac{\partial\psi_n(y_1, \ldots, y_n)}{\partial y_n} \end{vmatrix} \end{aligned}$$

で表される．

（独立な確率変数の和の分布）

X_1 と X_2 は互いに独立で，同時密度関数 $f(x_1, x_2) = f_{X_1}(x_1)f_{X_2}(x_2)$ を持つとする．

$$Y = X_1 + X_2$$

で表される Y の確率密度関数 $f_Y(y)$ は

$$f_Y(y) = \int_{-\infty}^{\infty} f_{X_1}(y - x_2) f_{X_2}(x_2)\, dx_2$$

で与えられる．

（証明）

$$Y_1 = X_1 + X_2, \qquad Y_2 = X_2$$

とすると，X_1, X_2 と Y_1, Y_2 の間で１対１の対応がある．したがって

$$f_Y(y_1, y_2) = |J| f_X(x_1, x_2)$$

が成立し，ヤコビアンは

$$|J| = \begin{vmatrix} \dfrac{\partial x_1}{\partial y_1} & \dfrac{\partial x_1}{\partial y_2} \\ \dfrac{\partial x_2}{\partial y_1} & \dfrac{\partial x_2}{\partial y_2} \end{vmatrix} = \begin{vmatrix} 1 & -1 \\ 0 & 1 \end{vmatrix} = 1$$

となる．X_1 と X_2 は独立であるので

$$f_Y(y_1, y_2) = f_X(x_1, x_2) = f_{X_1}(x_1) f_{X_2}(x_2) \qquad x_1 = y_1 - y_2,\ x_2 = y_2$$

で与えられる．一方，同時密度関数より周辺密度関数を求めるには

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_Y(y_1, y_2)\, dy_2$$

の関係を用いればよい．したがって，独立な２つの確率変数の和の確率密度関数は，おのおのの確率密度関数の**たたみ込み**

$$f_{Y_1}(y_1) = \int_{-\infty}^{\infty} f_{X_1}(y_1 - y_2) f_{X_2}(y_2)\, dy_2$$

で与えられる．

注意． たたみ込み演算により独立な確率変数の和の確率分布を求めるには，煩雑な計算を必要とする．モーメント母関数 (特性関数) という概念を使うと，これらの計算を簡単な微分，積分の演算で置き換えることができる．

練習問題　確率変数 X_1, X_2 の同時確率密度関数 $f(x_1, x_2)$ を

$$f(x_1, x_2) = \begin{cases} \dfrac{1}{4} \exp\left(-\dfrac{x_1 + x_2}{2}\right), & 0 < x_1 < \infty,\ 0 < x_2 < \infty \\ 0, & \text{その他} \end{cases}$$

とする．このとき，確率変数

$$Y = \frac{1}{2}(X_1 - X_2)$$

の確率密度関数を求めよ．

5　大数の法則と中心極限定理

5.1　大数の法則

5.1.1 大数の弱法則

独立な確率変数列 $X_1, X_2, \ldots$ は，期待値 μ, 分散 σ^2 をもつ同一の分布に従うとし，

$$S_n = X_1 + X_2 + \cdots + X_n$$

とおく.

定理 5.1 (大数の弱法則) 上の条件の下で, 任意の正数 ε に対して,

$$\lim_{n\to\infty} P\left(\left| \frac{S_n}{n} - \mu \right| \geq \varepsilon \right) = 0$$

が成り立つ.

このとき S_n/n は μ に**確率収束 (convergence in probability)** するといい

$$\frac{S_n}{n} \xrightarrow{P} \mu \quad (n \to \infty)$$

と表す. この結果を**大数の弱法則 (weak law of large numbers)** という.

この定理の結果よりもっと強い結論がもっと弱い条件で得られる. それが以下の大数の強法則である.

5.1.2 大数の強法則

独立な確率変数列 $X_1, X_2, \ldots$ が期待値 μ をもつ同一の分布に従うとし,

$$S_n = X_1 + X_2 + \cdots + X_n$$

とおく.

定理 5.2 (大数の強法則) 上の条件の下で

$$P\left(\lim_{n\to\infty} \frac{S_n}{n} = \mu \right) = 1$$

が成り立つ.

このとき S_n/n は μ に**概収束 (convergence almost everywhere)** する, **ほとんど確実に (almost surely)** 収束する, または**確率 1 で (with probability 1)** で収束するといい

$$\lim_{n\to\infty} \frac{S_n}{n} = \mu \quad \text{a.s.}$$

または

$$\frac{S_n}{n} \xrightarrow{\text{a.s.}} \mu \quad (n \to \infty)$$

と表す. この結果を**大数の強法則 (strong law of large numbers)** という.

5.2　中心極限定理

独立な確率変数列 $X_1, X_2, \ldots$ が期待値 μ, 分散 σ^2 をもつ同一の分布に従うとし,

$$S_n = X_1 + X_2 + \cdots + X_n$$

とおく．このとき, S_n を標準化した確率変数

$$\frac{S_n - n\mu}{\sqrt{n}\sigma}$$

の分布は $n \to \infty$ のとき標準正規分布に収束する．これは中心極限定理の最も単純な場合である.

定理 5.3 (中心極限定理) 上記の条件の下で, 任意の x に対して

$$\lim_{n\to\infty} P\left(\frac{S_n - n\mu}{\sqrt{n}\sigma} \le x\right) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz$$

が成り立つ.

このとき, $(S_n - n\mu)/\sqrt{n}\sigma$ は $N(0,1)$ に**法則収束 (convergence in low)** する, または**分布収束 (convergence in distribution)** するという.

大数の法則は古くから「法則」として認識されていたのに対して，中心極限定理はラプラスによって「定理」として最初に証明された.

（中心極限定理の例：二項分布の正規近似）$X_1, \cdots, X_n$ は互いに独立, 同一分布で二項分布 $B(1,p)$ に従うとする．このとき，$S_n = \sum_{i=1}^{n} X_i$ の分布は二項分布 $B(n,p)$ に従う．np が一定で n が大きいとき，S_n はポアソン分布に近ずく．一般には正規分布に近づくことを示す.

$$E(S_n) = np, \quad V(S_n) = npq \ \ (p+q=1), \ \ M_{S_n}(\theta) = (pe^{\theta} + q)^n$$

であるので，$Y_n = \dfrac{S_n - np}{\sqrt{npq}}$ の積率母関数 $M_{Y_n}(\theta)$ は,

$$\begin{aligned}
M_{Y_n}(\theta) &= E\left\{\exp\left(\theta \cdot \frac{S_n - np}{\sqrt{npq}}\right)\right\} \\
&= \exp\left(-\frac{np\theta}{\sqrt{npq}}\right) E\left\{\exp\left(\theta \cdot \frac{S_n}{\sqrt{npq}}\right)\right\} \\
&= \exp\left(-\sqrt{\frac{np}{q}}\theta\right)\left\{p \cdot \exp\left(\frac{\theta}{\sqrt{npq}}\right) + q\right\}^n
\end{aligned}$$

である．ここで，マクローリン展開により

$$\exp\left(\frac{\theta}{\sqrt{npq}}\right) = 1 + \frac{\theta}{\sqrt{npq}} + \frac{\theta^2}{2npq} + O(n^{-3/2})$$

である．ただし，$a_n = O(n^{-3/2})$ は $n^{3/2}a_n$ が有界を意味する．従って，$\log(1+x)$ のマクローリン展開を使うと

$$\begin{aligned}
\log M_{Y_n}(\theta) &= -\sqrt{\frac{np}{q}}\theta + n\log\left\{1 + \frac{\sqrt{p}}{\sqrt{nq}}\theta + \frac{\theta^2}{2nq} + O(n^{-3/2})\right\} \\
&= -\sqrt{\frac{np}{q}}\theta + n\left\{\frac{\sqrt{p}}{\sqrt{nq}}\theta + \frac{\theta^2}{2nq} - \frac{p}{2nq}\theta^2 + O(n^{-3/2})\right\} \\
&= \frac{\theta^2}{2} + O(n^{-1/2})
\end{aligned}$$

これより

$$\lim_{n\to\infty} M_{Y_n}(\theta) = \exp(\theta^2/2)$$

したがって

$$\lim_{n\to\infty} P\left(\frac{S_n - np}{\sqrt{npq}} \le x\right) = \Phi(x), \quad \Phi(x) \text{ は標準正規分布関数}$$

連続補正　中心極限定理から n が十分に大きければ, S_n の分布は正規分布 $N(n\mu, n\sigma^2)$ で近似できる．もし, S_n が整数値ならば上端および下端の値をそれぞれ 0.5 だけ広げれば, より良い近似ができる．すなわち, 整数 $a \le b$ に対して

$$\begin{aligned}
P(a \le S_n \le b) &\fallingdotseq P(a - 0.5 \le Y \le b + 0.5), \quad Y \sim N(n\mu, n\sigma^2) \\
&= P\left(\frac{a - 0.5 - np}{\sqrt{npq}} \le Z \le \frac{b + 0.5 - np}{\sqrt{npq}}\right), \quad Z \sim N(0,1) \\
&= \int_\alpha^\beta \frac{1}{\sqrt{2\pi}} e^{-z^2/2}\, dz
\end{aligned}$$

ここで

$$\alpha = \frac{a - 0.5 - np}{\sqrt{npq}}, \quad \beta = \frac{b + 0.5 - np}{\sqrt{npq}}$$

である．これを**連続補正 (correction for continuity)** という．

例題 6.3　$X \sim B(100, 0.3)$ のとき

$$P(25 \le X \le 40) = 0.873931 \text{ (正確な値)}$$

$np = 30, npq = 21$ だから

$$\begin{aligned}
P(25 \le X \le 40) &\fallingdotseq P\left(\frac{25 - 0.5 - 30}{\sqrt{21}} \le Z \le \frac{40 + 0.5 - 30}{\sqrt{21}}\right) \\
&= P(-1.20 \le Z \le 2.29) = 0.3849 + 0.4890 = 0.8739
\end{aligned}$$

連続補正を行わなければ

$$\begin{aligned}
P(25 \le X \le 40) &\fallingdotseq P\left(\frac{25 - 30}{\sqrt{21}} \le Z \le \frac{40 - 30}{\sqrt{21}}\right) \\
&= P(-1.09 \le Z \le 2.18) = 0.3621 + 0.4854 = 0.8475
\end{aligned}$$

補足： 簡単な漸近理論

確率・統計で利用する収束

確率変数列 $\{X_n\}$ が X に収束するというとき，次のようにいろいろな収束の仕方がある．

概収束：確率 1 で収束する，あるいはほとんど至るところで収束するともいう．定義は次で与えられる．

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

このとき，$X_n \xrightarrow{\text{a.s.}} X$ と書く．

確率収束：任意の $\epsilon > 0$ に対して,

$$\lim P(|X_n - X| < \epsilon) = 1$$

が成り立つことが定義である．このとき，$X_n \xrightarrow{\text{P}} X$ と書く．

分布収束：法則収束ともいう．F_n を X_n の分布関数，F を X の分布関数とするとき，F の任意の連続点において

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

が成り立つことが定義である．このとき，$X_n \xrightarrow{\mathrm{L}} X$ あるいは $X_n \xrightarrow{\mathrm{L}} F$ と書く．また，X が名前をもった分布に従うとき，例えば $N(0,1)$ に従っているとすると，$X_n \xrightarrow{\mathrm{L}} N(0,1)$ と書くこともある．

X が 1 点 c に退化しているとき（$P(X=c)=1$，本質的に X は c そのもの）であっても上記の定義は意味をもつので，そのときは

$$X_n \xrightarrow{\mathrm{a.s.}} X, \quad X_n \xrightarrow{\mathrm{P}} X, \quad X_n \xrightarrow{\mathrm{L}} X$$

などと表記する．

収束間の簡単な関係

$$X_n \xrightarrow{\mathrm{a.s.}} X \implies X_n \xrightarrow{\mathrm{P}} X \implies \xrightarrow{\mathrm{L}} X$$

ただし，$X=c$ の場合は，確率収束と分布収束は同値である．

収束と連続関数との関係

g が R 上の連続関数であるとき，g による確率変数の変換において各収束は遺伝する．つまり，次が成り立つ．

(i) $X_n \xrightarrow{\mathrm{a.s.}} \implies g(X_n) \xrightarrow{\mathrm{a.s.}} X$

(ii) $X_n \xrightarrow{\mathrm{P}} X \implies g(X_n) \xrightarrow{\mathrm{P}} X$

(ii) $X_n \xrightarrow{\mathrm{L}} X \implies g(X_n) \xrightarrow{\mathrm{L}} X$

スラツキーの定理　X_n, Y_n はそれぞれ確率変数列である．$X_n \xrightarrow{\mathrm{L}} X$ および $Y_n \xrightarrow{\mathrm{P}} c$ であるとき次が成り立つ．

(i) $X_n + Y_n \xrightarrow{\mathrm{L}} X + c$

(ii) $X_n Y_n \xrightarrow{\mathrm{L}} cX$

この 2 つの事実を用いると，下のことは自明である．

(i) $X_n - Y_n \xrightarrow{\mathrm{P}} 0,\ X_n \xrightarrow{\mathrm{L}} X \implies Y_n \xrightarrow{\mathrm{L}} X$

(ii) $X_n \xrightarrow{\mathrm{L}} X,\ Y_n \xrightarrow{\mathrm{P}} c = 0 \implies X_n Y_n \xrightarrow{\mathrm{P}} 0$

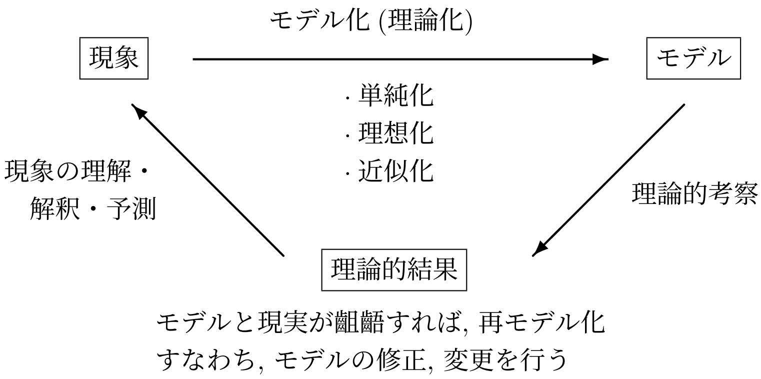
(iii) $X_n \xrightarrow{\mathrm{L}} X,\ Y_n \xrightarrow{\mathrm{P}} c \neq 0 \implies X_n / Y_n \xrightarrow{\mathrm{L}} X/c$

6　統計的推測の背景

6.1　現象とモデル

現象には大別して 2 通りある．一つは**決定論的現象** (**determinitic phenomenon**) と呼ばれるもので, 例えば, 星の運行のように任意の時刻におけるその位置が完全に決定できる現象である．もう一つは**確率** (**的**) (**probabilistic**), または**統計** (**的**) **現象** (**statisitical phenomenon**) と呼ばれるもので, 例えば, サイコロ投げのようにその結果が偶然性に左右されて予測することが困難な現象である．

確率論や統計学が対象とするのは後者の確率的現象である．ところで, 現実の現象はそのままでは非常に複雑なので, その中から副次的なものを無視して**理想的なモデル** (**ideal model**) を考える．このことを**現象のモデル化** (**modelization of phenomenon**) という．したがって, モデル化とは現実の現象に対する一つの近似である．その際, **単純化** (**simplification**) や**理想化** (**idealization**) が必ず行われる．モデルに対して理論的な考察を加え, その結果から現象を理解したり, 解釈したりするわけである．図示すると下図のようになる．



6.2　確率モデル

偶然的な現象に対する理論モデルを**確率モデル** (**probabilistic model**)(**統計モデル** (**statstical model**)) という．確率モデル (統計モデル) における現象は偶然的変動に左右されるが，その偶然的変動にも一定の法則性が存在する (例えば，硬貨を投げ続ける場合，各回の結果を完全に予知することはできないが，多数回繰り返すうち，ほぼ半数ずつの表と裏が出現する)．この法則性は具体的には**確率分布** (**probability distribution**) として把握される．したがって，確率モデルに対するもっとも肝要な知識はその確率分布を知ることである．なお，本来は決定論的現象であっても，あまりにも複雑な場合には確率現象として扱うのが現実的であり，しかも有効であることが多い.

6.3　母集団と母数, 統計量と標本分布

統計学では個々のデータの背後に理論的な確率モデル，すなわち**母集団** (**population**) を想定する．データを通して母集団の確率分布 (**母集団分布** (**distribution of population**)) に関する知識を効率的に精度よく得るのが統計学の目標である．母集団分布に正規分布を仮定する場合，これを**正規母集団** (**normal population**) という．

母集団分布の期待値 (平均)，分散，標準偏差をそれぞれ**母平均** (**population mean**)，**母分散** (**population variance**)，**母標準偏差** (**population standard deviation**) という．このような母集団分布に関連した重要な数値を**母数** (**パラメータ** (**parameter**)) という．母数が分かれば母集団分布は確定する．例えば，正規母集団 $N(\mu, \sigma^2)$ は母平均 μ と母分散 σ^2 が分かれば完全に決まる．

実際に母数に対して何らかの判断を行おうとする場合，基となるのはデータであり，いま，データとして n 個の個体に対する測定値が得られているとする．この n 個の個体を母集団からの**標本** (**sample**) という．標

本の個数 n を**標本の大きさ (sample zize)** といい，大きさ n の標本に対する測定値 $(x_1, x_2, \ldots, x_n)$ を**標本値 (sample value)** というが，この標本値について, 以下の仮定をする. 母集団分布に従う確率変数を X とするとき, X と同じ分布に従い, 互いに独立な n 個の確率変数 $(X_1, X_2, \ldots, X_n)$ を仮想的に考え, これらの実現値 (実際に取った値) が標本値 $(x_1, x_2, \ldots, x_n)$ であると考える．$(X_1, X_2, \ldots, X_n)$ を**無作為標本 (random sample)** という. 無作為標本の関数を**統計量 (statistic)** という．統計量は確率変数の関数であるからまた確率変数であり，その分布を**標本分布 (sample distribution)** という.

主要な統計量として，大きさ n の無作為標本 $(X_1, X_2, \ldots, X_n)$ に対して

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \qquad S^2 = \frac{1}{n}\sum_{i=1}^{n}(X_i - \bar{X})^2, \qquad U^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$$

を定義し, それぞれ，**標本平均 (sample mean)**, **標本分散 (sample variance)**, **標本不偏分散 (sample unbiased variance)** という.

$$\begin{array}{lcl}
\text{統計量 (確率変数)} & & \text{実現値} \\
(X_1, X_2, \ldots, X_n) & \Longrightarrow & (x_1, x_2, \ldots, x_n) \\
\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i & \Longrightarrow & \bar{x} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} x_i \\
S^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2 & \Longrightarrow & s^2 = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2 \\
U^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(X_i - \bar{X})^2 & \Longrightarrow & u^2 = \dfrac{1}{n-1}\displaystyle\sum_{i=1}^{n}(x_i - \bar{x})^2
\end{array}$$

標本分布

上の標本平均，標本分散については次が成り立つ.

- 母平均が μ，母分散が σ^2，smaple size が n のとき

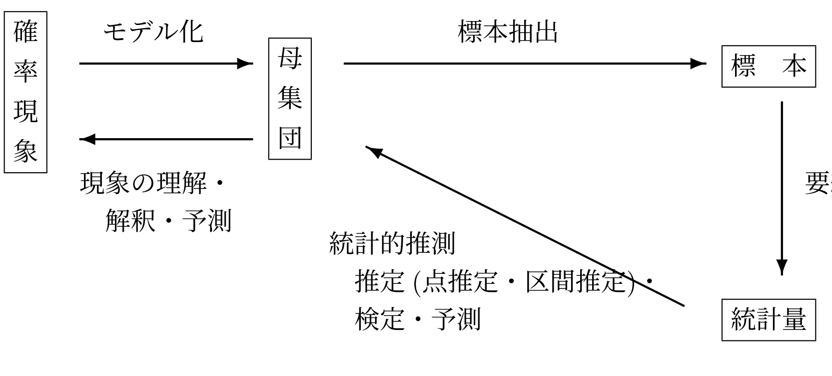
$$E(\bar{X}) = \mu, \quad V(\bar{X}) = \frac{\sigma^2}{n}$$

$$E(S^2) = \frac{n-1}{n}\sigma^2, \quad E(U^2) = \sigma^2$$

- 正規母集団 $N(\mu, \sigma^2)$ で，sample size が n のとき

$$\bar{X} \sim N\left(\mu, \frac{\sigma^2}{n}\right)$$

標本によって母集団についての判断を下すことを**統計的推測 (statistical inference)** という．統計的推論の構図は下図のようになる.



6.4 統計的推測の形式

前節で述べたように，母集団分布を知ることは母数を知ることに帰着する．母数が分かっていれば，母集団が確定するので統計的推測の必要はない．母数が分かっていない場合は，データの要約値である統計量の実現値を用いて母数を推測する．そしてその形式は，主なものとして**推定** (estimation)，**検定** (test)，**予測** (prediction) の三つがある．

6.4.1 推定

ここでは，母数を θ，推定のために用いられる統計量 (**推定量** (estimator) という)，推定量の実現値 (**推定値** (estimete) という) をそれぞれ

$$\begin{array}{ccc} \text{推定量 (確率変数)} & & \text{推定値 (実現値)} \\ \hat{\theta}(X_1, X_2, \ldots, X_n) & \Longrightarrow & \hat{\theta}(x_1, x_2, \ldots, x_n) \end{array}$$

とする．

[**点推定**]

母数 θ の値をただ 1 つの値 $\hat{\theta}(x_1, x_2, \ldots, x_n)$ で推定する．すなわち，$\hat{\theta}(x_1, x_2, \ldots, x_n)$ を θ の近似値とみなすわけである．このような手法を**点推定** (**法**) (point estimation) という．

[**区間推定**]

上の点推定法では，その推定がどれくらい正確であるか，については何も保証が得られない．他方，母数 θ に対して，2 つの統計量 $\hat{\theta}_L(X_1, X_2, \ldots, X_n), \hat{\theta}_U(X_1, X_2, \ldots, X_n)$ を

$$P\left(\hat{\theta}_L(X_1, X_2, \ldots, X_n) \leq \theta \leq \hat{\theta}_U(X_1, X_2, \ldots, X_n)\right) \geq 1 - \alpha$$

が成り立つように決める．ここで，α はあらかじめ定めた値で，$0 < \alpha < 1$ を満たす．

2 つの推定量 $\hat{\theta}_L(X_1, X_2, \ldots, X_n), \hat{\theta}_U(X_1, X_2, \ldots, X_n)$ の実現値による区間

$$\left(\hat{\theta}_L(x_1, x_2, \ldots, x_n), \hat{\theta}_U(x_1, x_2, \ldots, x_n)\right)$$

を α に対する**信頼係数** (confidence coefficient) $1 - \alpha$ の**信頼区間** (confidence interval)[2], $\hat{\theta}_L(x_1, x_2, \ldots, x_n), \hat{\theta}_U(x_1, x_2, \ldots, x_n)$ をそれぞれ，**信頼下限** (confidence lower limit)，**信頼上限** (confidence upper limit) という．このような手法を**区間推定** (**法**) (interval extimation) といい，通常 $\alpha = 0.05$ とした信頼係数 95% の区間推定がよく行われる．

6.4.2 検定

(**統計的仮説検定**) 次の例を通して検定の考え方を説明しよう．ある疾患の治療法に A，B の 2 法がある．患者 100 名を 2 群に分けて，70 名に A 法，30 名に B 法を実施したところ下表の結果を得た．これら 2 法に違いがあるかどうかを統計的に検討する．

	治癒	非治癒	計
A 法	25 (x_A)	45	70 (n_A)
B 法	20 (x_B)	10	30 (n_B)
計	45	55	100

研究者は両治療法間に差異があると想定しているとする．背理法的に考えて，まず「A 法と B 法との間に差異がない」と仮定する．この仮定が右表のデータと両立しなければ (すなわち，この仮定の下で上表のデータが出現する確率がきわめて小さいならば)，2 法の間に差異がない，という仮定を否定する．この仮定 (2 法の間に差異がない) のことを**帰無仮説** (null hypothesis) といい，通常 H あるいは H_0 で表す．データに基づいて帰無仮説を否定できないと判断するとき, 帰無仮説を**採択** (**受容**) **する** (accept) といい，帰無仮説を正しくないと判断するとき, 帰無仮説を**棄却する** (reject) という. 帰無仮説が棄却されるとき正しいと判断される状況を**対立仮説** (alternatve

[2] 母数 θ は未知だが定数であるので，個々の信頼区間は θ を含むか否かのいずれかである．すなわち，1 つの実現信頼区間が θ を含む "確率" は 1 か 0 のいずれかである．他方，信頼区間を多数作ると，そのうちのおよそ $100(1-\alpha)$% の信頼区間は θ を含み，およそ 100α% は θ を含まない．そこで 1 つの信頼区間が θ を含む "信頼係数" は $100(1-\alpha)$% であるという．

hypothesis) といい，通常 K あるいは H_1 で表す．

真＼決定	H	K
H	○	Type I error
K	Type II error	○

統計的仮説検定において, 本当は帰無仮説が正しいのにこれを棄却する誤りを**第一種の誤り (type I error)**, 本当は帰無仮説が正しくないのにこれを採択する誤りを**第二種の誤り (type II error)** という．統計的仮説検定において, 第一種, 第二種両方の誤りの確率を同時に小さくすることはできない. そこで統計的仮説検定では, 第一種の誤りの確率をあらかじめ定めた (通常は 5% あるいは 1%) 上で, 第二種の誤りの確率ができるだけ小さくなるような手法を考える．このあらかじめ定めておく第一種の誤りの確率を**有意水準 (level of significance)** または**危険率**といい，通常 α で表す．有意水準 α の検定で帰無仮説が棄却されるとき，**水準 α で有意 (significant on level α)** という．

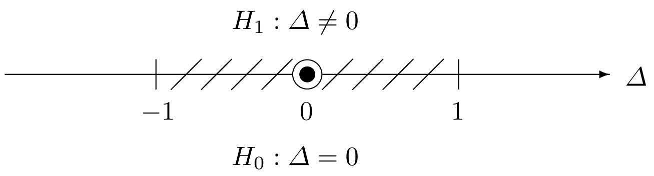
帰無仮説 H_0 の下で，上表のようなデータが出現する確率が α 以下であれば H_0 を棄却する．上の例において，未知である A 法の治癒率 (A 法の**母比率 (population proportion)**) を p_A，未知である B 法の治癒率 (B 法の母比率) を p_B とおくと，帰無仮説は

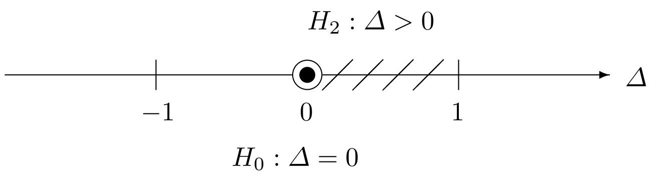
$$H_0 : p_A = p_B$$

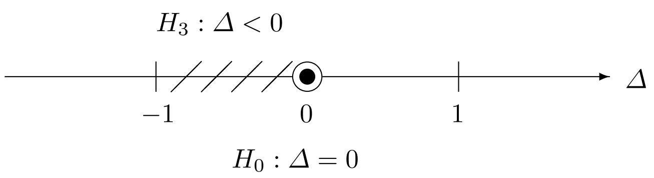
であり，対立仮説としては

$$H_1 : p_A \neq p_B, \qquad H_2 : p_A > p_B, \qquad H_3 : p_A < p_B$$

の 3 通りが考えられる．$\Delta = p_A - p_B$ とおいて図示すれば次のようになる．







ここで H_1 を**両側対立仮説 (two-sided alternative hypothesis)**, H_2, H_3 を**片側対立仮説 (one-sided alternative hypothesis)** という．どのような対立仮説を選ぶかは事前の情報による．A 法と B 法の優劣に関する事前の情報がなければ H_1 を選ぶ．もし A 法が旧治療法で B 法が新治療法ならば，B 法が優れていると期待されるので，対立仮説として H_3 を選ぶ．両側対立仮説のみを考える場合は，帰無仮説を H，対立仮説を K で表すことにする．今の例では

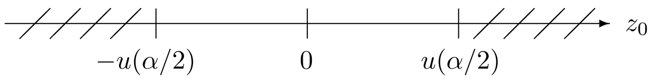
$$H : p_A = p_B \quad \text{vs.} \quad K : p_A \neq p_B$$

である．

A 法による治癒率 p_A のデータから計算される推定値 (母比率に対して**標本比率 (sample proportion)** という) を $\hat{p}_A = x_A/n_A$，B 法による治癒率 p_B のデータから計算される推定値を $\hat{p}_B = x_B/n_B$ とおくと，直感的にみて，$\hat{p}_A - \hat{p}_B$ の絶対値が大きければデータは H を否定し，K を支持していると考えることができる．実際の検定は標本変動を考慮して次のように行う (**参照).

$$|z_0| = \frac{|\hat{p}_A - \hat{p}_B|}{\sqrt{\hat{p}(1-\hat{p})\left(\dfrac{1}{n_A} + \dfrac{1}{n_B}\right)}} > u(\alpha/2)$$

のとき H を棄却し，K を採択する (有意水準 α)．ここで，$\hat{p} = (x_A + x_B)/(n_A + n_B)$ は患者全体の標本比率であり, $u(\alpha/2)$ は標準正規分布表の上側 $\alpha/2$ 点である[3]．z_0 を実現値とする統計量 (確率変数) Z_0 のように検定に用いられる統計量を**検定統計量 (test statistic)** という．上記の検定法は，z_0 が区間 $(-\infty, -u(\alpha/2))$ または $(u(\alpha/2), \infty)$ に入れば H を棄却 (K を採択) するということである．このような領域を H の**棄却域 (rejection region, critical region)** (K の**採択域 (acceptance region)**) という．



斜線部が H の棄却域

前記の治療データに当てはめれば

$$\hat{p}_A = 25/70, \hat{p}_B = 20/30, \hat{p} = 45/100, n_A = 70, n_B = 30 \quad \text{なので}$$

$$|z_0| = \frac{\left|\dfrac{25}{70} - \dfrac{20}{30}\right|}{\sqrt{\dfrac{45}{100}\left(1 - \dfrac{45}{100}\right)\left(\dfrac{1}{70} + \dfrac{1}{30}\right)}} \fallingdotseq 2.85 > 1.96$$

より, H は有意水準 5%で棄却される．なお, 検定統計量 Z_0 の絶対値が 2.85 より大きくなる確率は標準正規分布表から $2 \times (1 - 0.9978) = 0.0044$, すなわち 0.44%である．この確率のことを**有意確率 (significance probability)** または ***p* 値 (*p*-value)** という．

注意 1 $|z_0| > u(\alpha/2)$ と p-value $< \alpha$ は同値である．

なお, 検定問題 H_0 vs. H_2 に対しては, $z_0 > u(\alpha)$ のとき H_2 が採択 (立証) され, H_0 が棄却される (有意水準 α)．ここで, $u(\alpha)$ は標準正規分布表の上側 α 点である[4]．また, 検定問題 H_0 vs. H_3 に対しては, $z_0 < -u(\alpha)$ のときのとき H_3 が採択 (立証) され, H_0 が棄却される (有意水準 α).

検定の際の注意

① 統計的仮説検定の主眼は, 帰無仮説を棄却することにより対立仮説を間接的に立証することにあるが, 帰無仮説が棄却されないことによりこれを消極的に支持するために用いられることもある．

② n(標本の大きさ) が小さいと, 帰無仮説と対立仮説に大きな差異があっても対立仮説は立証されにくい．すなわち, 標本誤差が大きいので検査の結果は有意になりにくい．

③ n が大きいと, 帰無仮説と対立仮説の小さな差異まで立証されやすくなる．すなわち, 標本誤差が小さいので検定の結果は有意となりやすい．

④ ②, ③から分かるように統計的差異と実質的差異を区別して考える必要がある．

[3]付表の標準正規分布表を参照．例えば $\alpha = 0.05$ のとき，$u(0.025) = 1.96$.
[4]付表の標準正規分布表を参照．例えば $\alpha = 0.05$ のとき, $u(0.05) = 1.645$.

6.4.3 予測

過去のデータから母数の近似値を求めるのが母数の推定であったが, 将来の観測値の近似値を求めるのが将来値の**予測** **(prediction)** である．予測には**点予測** **(porint prediction)** と**区間予測** **(intervel prediction)** がある．予測に用いる統計量を**予測** (**統計**) **量** **(predictor)** という．母数は定数であるが, 将来値は確率変数である．将来値を Y とし, 無作為標本 $(X_1, X_2, \ldots, X_n)$ から 2 つの統計量 $l(X_1, X_2, \ldots, X_n), u(X_1, X_2, \ldots, X_n)$ を適当に作り

$$P(l(X_1, X_2, \ldots, X_n) \leq Y \leq u(X_1, X_2, \ldots, X_n)) \geq 1 - \alpha$$

となるようにする．このとき, 区間 $(l(x_1, x_2, \ldots, x_n), u(x_1, x_2, \ldots, x_n))$ を Y に関する信頼係数 $1 - \alpha\,(100(1 - \alpha)\%)$ の予測区間という．

単純な場合として, $X_1, X_2, \ldots, X_n$ を過去のデータ, $Y_1, Y_2, \ldots, Y_m$ を将来のデータとし, これらがすべて同一の正規母集団からの無作為標本である場合を考える．将来のデータの平均値 $\bar{Y} = \dfrac{1}{m}\displaystyle\sum_{j=1}^{m} Y_j$ に対する信頼係数 $100(1 - \alpha)\%$ の予測限界は, 予測量 $\bar{X} = \dfrac{1}{n}\displaystyle\sum_{i=1}^{n} X_i$ と $\bar{Y}$ の変動を考慮して

$$\bar{x} \pm t_{n-1}(\alpha/2)\sqrt{\left(\frac{1}{m} + \frac{1}{n}\right)}\, u$$

で与えられる[5]．ここで, $\bar{x}, u$ はそれぞれ

$$\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i, \quad U = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2}$$

の実現値である．

[5] $m \to \infty$ とすれば, この予測限界は母平均 μ に対する信頼係数 $100(1 - \alpha)\%$ 信頼限界 $\bar{x} \pm t_{n-1}(\alpha/2)\dfrac{u}{\sqrt{n}}$ に一致する．

確率論基礎演習問題 (補足)

$\boxed{\mathbf{1}}$ 確率変数 X は，平均 μ, 分散 σ^2 の正規分布 $N(\mu, \sigma^2)$ に従うとする．$Y = e^X$ とするとき，平均 $E(Y)$, 分散 $V(Y)$ を求めよ．

$\boxed{\mathbf{2}}$ 確率変数 X の確率密度関数が

$$f(x) = \begin{cases} \lambda e^{-\lambda x}, & 0 < x < \infty \qquad (\lambda \text{は正の定数}) \\ 0, & \text{その他の } x, \end{cases}$$

のとき，$E(X^n)$ を求めよ．

$\boxed{\mathbf{3}}$ 次の各問に答えよ．

(1) 確率変数 X が正規分布 $N(\mu, \sigma^2)$ に従うとき，X のモーメント (積率) 母関数 $M_X(t)$ を求めよ．ここで，正規分布 $N(\mu, \sigma^2)$ の確率密度関数は

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

である．

(2) X_1，X_2 は，それぞれ正規分布 $N(\mu_1, \sigma_1^2)$，$N(\mu_2, \sigma_2^2)$ に従う互いに独立な確率変数とするとき，$X_1 + X_2$ のモーメント (積率) 母関数 $M_{X_1+X_2}(t)$ を求め，$X_1 + X_2$ の確率分布を与えよ．

$\boxed{\mathbf{4}}$ 次の各問に答えよ

(1) 確率変数 Z が標準正規分布 $N(0,1)$ に従うとき，Z^2 の確率密度関数を求めよ．また，Z^2 のモーメント (積率) 母関数 $M_{Z^2}(t)$ は,

$$M_{Z^2}(t) = (1-2t)^{-\frac{1}{2}}, \quad t < \frac{1}{2}$$

であることを示せ.

(2) $Z_1, \cdots, Z_n$ は標準正規分布 $N(0,1)$ からの大きさ n の無作為標本とする．

$$\mathcal{X}^2 = \sum_{i=1}^{n} Z_i^2$$

とするとき，モーメント (積率) 母関数 $M_{\chi^2}(t)$ は

$$M_{\chi^2}(t) = (1-2t)^{-\frac{n}{2}}, \quad t < \frac{1}{2}$$

であり，$\mathcal{X}^2$ は 自由度 n のカイ 2 乗分布 $\mathcal{X}_n^2$ に従うことを示せ.