

1. 確率

1.1 確率について

ラプラス流の確率の定義は次のようになる.

定義 1. (i) 起こり得る場合の数が n 通りあり、どの場合も起こるのが同様に確からしいとする.

(ii) ある事象 A の場合の数が r 通りであるとき、事象 A の起こる確率を $p = \frac{r}{n}$ と定義する.

例 2. どの目の出るのも同様に確からしいサイコロを 2 回投げる. 起こり得る根元事象全体は (1 回目に出た目が i , 2 回目に出た目が j のとき (i, j) と書く)

$$\Omega = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}.$$

の 36 通り. 少なくとも 1 回は 1 の目が出るという事象 A を考える.

$$A = \{(i, j) \mid i = 1 \text{ または } j = 1\}$$

である. $\#A = 36 - 5^2 = 11$ なので ($\#A$ で集合 A の要素の個数を表すことが多い) A の起こる確率 $p = 11/36$.

しかし、ラプラス流に

$$\frac{\text{事象 } A \text{ の場合の数}}{\text{起こり得るすべての場合の数}}$$

で確率を与えられない場合もあるため、ラプラス流の確率の定義を離れ、確率のもつ性質を抽出した確率空間というものを考える必要がある. 以下の定義は Kolmogorov が著書「確率論の基礎概念」で導入している. この定義は参考のために述べるもので、完全な理解を求めるものではありません.

定義 3. 確率空間とはある集合 Ω , Ω の部分集合の集まり $\mathcal{F}$, 確率 P の 3 つの組み $(\Omega, \mathcal{F}, P)$ で以下をみたすものである.

(1) $\mathcal{F}$ は次の性質をみたす:

- (a) $A_1, A_2, \dots, A_i, \dots$ が $\mathcal{F}$ の要素ならば $\cup_{i=1}^{\infty} A_i$ も $\mathcal{F}$ の要素.
- (b) $A \in \mathcal{F}$ ならば $A^c \in \mathcal{F}$
- (c) $\Omega \in \mathcal{F}$.

(2) $A \in \mathcal{F}$ に対して実数 $P(A)$ が定まり、次をみたす.

- (a) $0 \leq P(A) \leq 1$
- (b) $P(\Omega) = 1$
- (c) [可算加法性] $A_1, A_2, \dots, A_i, \dots$ が $\mathcal{F}$ の要素で $i \neq j$ のとき $A_i \cap A_j = \emptyset$ ならば

$$P(\cup_{i=1}^{\infty} A_i) = \sum_{i=1}^{\infty} P(A_i)$$

上記定義で $\cup_{i=1}^{\infty} A_i$ は集合の和を表し,

$$\cup_{i=1}^{\infty} A_i = \{\omega \in \Omega \mid \text{ある } i \text{ が存在して } \omega \in A_i\}$$

と定義されています.

注意 4. (1) Ω の要素 ω 1 つからなる集合 $\{\omega\}$ を**根元事象**という.

(2) $\mathcal{F}$ は Ω の部分集合 A で確率 $P(A)$ が定まっているもの全体である. $\mathcal{F}$ の要素は**事象**と呼ばれる.

(3) 事象 $A \in \mathcal{F}$ に対して A の補集合 A^c を A の**余事象**という. (A の補集合は英語で complement of A と言う. その頭文字 c をつけるのである)

(4) $A \cap B$ を事象 A, B の**積事象**という. $A \cap B = \emptyset$ となるとき事象 A, B は互いに**排反**という.

例 5. (1) 2 回のサイコロ投げの確率空間

$\Omega = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$. $\mathcal{F}$ は Ω の部分集合全体, $P(A) = \frac{A \text{ の要素の個数}}{36}$ と定義すれば $(\Omega, \mathcal{F}, P)$ は確率空間の一例である.

もう少し一般に有限集合 $\Omega = \{\omega_1, \dots, \omega_n\}$, $\mathcal{F}$ を Ω の部分集合全体, 各根元事象 $\{\omega_i\}$ の確率を p_i とし, $P(A) = \sum_{\{\omega_i \mid \omega_i \in A\}} p_i$ と定義すると $(\Omega, \mathcal{F}, P)$ は確率空間である.

(2) 無限回のサイコロ投げ

何回もサイコロ投げを続けるとその試行の結果として、1~6 の数字の無限列が現れる. これが根元事象とみなせる. すなわち $\Omega = \{(a_1, a_2, \dots, a_n, \dots) \mid a_i = 1, \dots, 6\}$. $\mathcal{F}$ と P の定義は簡単ではないが、うまく定義することができる.

(3) $\Omega = [a, b]$ とする. $\mathcal{F}$ として長さが定まるような $[a, b]$ の部分集合とし $P(A) = \frac{A \text{ の長さ}}{b-a}$ と定めれば、確率空間になる. (ただし、 $[a, b]$ の任意の部分集合に対して長さが決まるわけではないということを注意しておきます. 長さが定まる集合をルベグ可測集合と言います.)

1.2 事象の独立性

定義 6. (1) 事象 A, B が独立とは

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

となるときに言う.

(1) 条件付き確率 A, B を事象とし, $P(A) > 0$ とする. 条件 A の下での B の確率を

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

と定義する.

命題 7. $0 < P(A) < 1, 0 < P(B) < 1$ とする. A と B が独立とは次の (1), (2) のいずれかが成立することと同値である.

(1) $P(B|A) = P(B|A^c) = P(B)$. (2) $P(A|B) = P(A|B^c) = P(A)$.

定義 8. n 個の事象 $A_1, \dots, A_n$ が独立とは次のときに言う:

$A_1, \dots, A_n$ から有限個とってきた $A_{i_1}, \dots, A_{i_p}$ ($i_1 < \dots < i_p$) について

$$P(\cap_{k=1}^p A_{i_k}) = P(A_{i_1}) \cdots P(A_{i_p})$$

がつねに成り立つ.

例えば A, B, C が独立とは次の 4 つの等式が成立することと同値である.

$$P(A \cap B \cap C) = P(A)P(B)P(C), P(A \cap B) = P(A)P(B), P(A \cap C) = P(A)P(C), P(B \cap C) = P(B)P(C)$$

(i) 「 A, B, C が独立」 と (ii) 「 A と B が独立」 かつ 「 B と C が独立」 かつ 「 C と A が独立」

は違うことを注意せよ. 当然 (i) ならば (ii) だが逆は成立しない.

1. 確率

1.2 事象の独立について (続き)

注意 9. 事象 $A_1, \dots, A_n$ が独立ということを条件付き確率を用いて定義することもできる. すなわち次が成立することと同値である.

(*) 「すべての i に対して A_i の確率とそれ以外の事象および余事象

$$A_1, \dots, A_{i-1}, A_{i+1}, \dots, A_n, A_1^c, \dots, A_{i-1}^c, A_{i+1}^c, \dots, A_n^c$$

による積事象で条件付けした確率が一致する」

例えば

$$P(A_i) = P(A_i | A_1 \cap \dots \cap A_{i-1} \cap A_{i+1}, \dots, A_n)$$

$$P(A_i) = P(A_i | A_1 \cap \dots \cap A_{i-1} \cap A_{i+1}^c, \dots, A_n^c)$$

などが成立する (他に $2^{n-1} - 2$ 個の同様な式が成立することを要求していることに注意) ということである.

2. 確率分布

2.1 確率変数、分布関数

定義 10. $(\Omega, \mathcal{F}, P)$ を確率空間とする. Ω 上の関数を確率変数という. すなわち Ω の各根元事象 $\omega \in \Omega$ に対して数値 $X(\omega)$ が対応しているものを言う. 確率変数 X に対して関数

$$F(x) = P(\{\omega \in \Omega \mid X(\omega) \leq x\}) \quad -\infty < x < \infty$$

を X の分布関数という. $P(\{\omega \in \Omega \mid X(\omega) \leq x\})$ を $P(X \leq x)$ のように略記することが多い.

例 11. (1) サイコロを 2 回投げる場合の確率空間上の確率変数

$\Omega = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$ である. $\omega = (i, j)$ という根元事象は 1 回目に i , 2 回目に j が出るという事象に対応している.

$$X_1(\omega) = i, X_2(\omega) = j \quad \omega = (i, j) \text{ のとき}$$

と定めれば X_1, X_2 はそれぞれ 1 回目、2 回目の目を表す確率変数である. $X_1(\omega) + X_2(\omega)$ も確率変数であり $2 \leq r \leq 7$ のとき $P(X_1 + X_2 = r) = \frac{r-1}{36}$, $8 \leq r \leq 12$ のとき $P(X_1 + X_2 = r) = \frac{13-r}{36}$ となる.

この例では、 X_1 は $\{1, \dots, 6\}$, $X_1 + X_2$ は $\{2, \dots, 12\}$ のように離散的な値しか取らない. これを離散型確率変数と言う.

一般的に確率空間の全事象の空間 Ω が有限集合ならば、確率変数 $X(\omega)$ の取る値は有限個であり、離散型確率変数になる.

(2) 確率変数 X が飛び飛びの値ではなく、連続的に値を取り得る場合、例えば取り得る可能性のある値が実数全体や $[a, b]$ のような区間に広がっている場合がある. これを連続型の確率変数と言う.

$$\bullet f(x) \geq 0 \ x \in \mathbb{R}, \quad \bullet \int_{-\infty}^{\infty} f(x) dx = 1$$

をみたす関数 f が存在して

$$P(X \leq x) = \int_{-\infty}^x f(t) dt = F(x)$$

となる確率変数 X を確率密度関数 f をもつ連続型確率変数と言う.

定義 12. 確率変数の値がどのように散らばっているかを知るには、 $P(a < X < b)$, $P(a < X \leq b)$ などの確率がわかればよいであろう。

$$P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$$

だから分布関数がわかっていれば十分とも言える。このように確率変数の散らばり具合を総称して確率分布あるいは単に分布と言う。(X の確率分布をもう少しきちんと定義すると次のようになる：確率変数 X に対して $\mu_X(A) = P(X \in A)$ $A \subset \mathbb{R}$ で定まる $\mathbb{R}$ 上の確率 μ_X を X の確率分布, 確率法則と言います。)

(1) のような離散型の確率変数の分布を離散型の確率分布、(2) のように確率密度関数を持つ分布関数を持っている分布を連続型の確率分布と言う。

(a) 以下の確率密度関数を持つ分布を $[a, b]$ 上の一様分布と言う。

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \text{ のとき} \\ 0 & x \leq a, x \geq b \text{ のとき} \end{cases}$$

(b) $\int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx = 1$ はよく知られている。 $f(x) = \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}}$ を密度関数にもつ確率分布を標準正規分布という。

2.2 確率変数の平均、期待値

確率変数 (または確率分布) の期待値、平均を離散型、連続型の場合に分けて定義する。

定義 13. (1) 離散型のとき

X の確率分布が $P(X = a_i) = p_i$ ($1 \leq i \leq N$) とする。ただし、 $\{a_1, \dots, a_N\}$ はすべて異なる数とする。このとき X の平均、期待値 EX を

$$EX = \sum_{i=1}^N a_i p_i$$

と定める。

(2) 連続型のとき

X の確率分布が密度関数 f をもつとき

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

と定義する。

命題 14. X を確率変数とする。 g を $\mathbb{R}$ 上の関数とすると $g(X)$ も確率変数である。

(1) X が定義 13 (1) の離散型のとき

$$Eg(X) = \sum_{i=1}^N g(a_i) p_i.$$

(2) X が定義 13 (2) のように密度関数をもつとき

$$Eg(X) = \int_{-\infty}^{\infty} g(x) f(x) dx.$$

注意 15. X, Y を確率変数, a, b を定数とするととき

$$E[aX + bY] = aE[X] + bE[Y].$$

このことは、後で X と Y の結合分布を学んでから示すことにする。 $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$ などは命題 14 から直ちにわかる。

2.3 分散・標準偏差・積率母関数

定義 16. 確率変数 X に対して

$$\begin{aligned} V[X] &= E[(X - m)^2] \text{ を } X \text{ の分散} \\ \sigma[X] &= \sqrt{V(X)} \text{ を } X \text{ の標準偏差} \end{aligned}$$

という。ただし m は X の期待値。また $E[X^n]$ を X の n 次モーメントという。

$$(1) P(X = a_i) = p_i \text{ (}\{a_1, \dots, a_N\} \text{ は相異なる数) のとき } V[X] = \sum_{i=1}^N (a_i - m)^2 p_i.$$

$$(2) X \text{ の分布が確率密度関数 } f \text{ をもつ連続型確率変数のとき } V[X] = \int_{\mathbb{R}} (x - m)^2 f(x) dx.$$

注意 17. 離散型確率変数で有限個の値しか取らない場合は期待値、分散とも有限和なので、確定するが、

(i) 離散型だが無限個の値を取り得る場合

(ii) 連続型の確率変数の場合

は期待値、分散が発散して定義できない場合がある。確率密度関数 $f(x) = \frac{1}{\pi(1+x^2)}$ をもつ連続型の分布はコーシー分布と呼ばれる重要な分布だが、この分布の平均値は定義できない。

定理 18. (1) $V[X] = E[X^2] - E[X]^2$. (2) $V[aX + b] = a^2 V[X]$.

分散は確率変数 X が平均値 m からどれだけ離れて分布しているかを表している。実際次の評価が成り立つ。

定理 19 (Chebyshev の不等式). X の分散を σ^2 , 期待値を m とすると

$$P(|X - m| \geq r) \leq \frac{\sigma^2}{r^2}.$$

この証明は後で述べる。確率分布の研究には次の積率母関数がよく用いられる。

定義 20. 確率変数 X に対して $t \in \mathbb{R}$ の関数

$$M_X(t) = E[e^{tX}]$$

を積率母関数という。

積率母関数は期待値や分散と同様、すべての確率変数、確率分布に対して定義できるわけでは無いことに注意してほしい。 $t \in \mathbb{R}$ の関数 $\varphi_X(t) = E[e^{tX\sqrt{-1}}]$ は特性関数と呼ばれ、やはりよく使われるがこの講義では論じない。

定理 21. M_X が $t = 0$ の近傍で滑らかな関数ならば $M_X^{(n)}(0) = E[X^n]$, $V[X] = M_X''(0) - (M_X'(0))^2$.

注意 22. (1) $M_X(t) = M_Y(t)$ ならば X と Y の分布は同じ

(2) X と Y の分布が同じならば $E[X^n] = E[Y^n] \forall n$ (ただしすべての次数のモーメントが存在すれば)。しかし $E[X^n] = E[Y^n] \forall n$ だとしても X と Y の分布が同じと結論できるわけでは無い。

3. 重要な確率分布

3.1 離散型確率分布

(1) 二項分布

結果が二つの試行、(例えば、コイン投げで表が出るか裏が出るかなど)で成功の確率が p , 失敗の確率が $q (= 1 - p)$ とする. この試行を n 回独立に繰り返したとき、成功の回数を表す確率変数 X の従う分布を二項分布と言い、 $B(n, p)$ と表す. 具体的には X は $\{0, 1, \dots, n\}$ のいずれかの値を取る離散型確率変数で

$$P(X = k) = {}_nC_k p^k q^{n-k} \quad 0 \leq k \leq n$$

となる分布を言う.

定理 23. X が二項分布 $B(n, p)$ に従うとする.

$$(1) E[X] = np, \quad (2) V[X] = npq, \quad (3) M_X(t) = (pe^t + q)^n = (1 + p(e^t - 1))^n.$$

証明.

$$M_X(t) = \sum_{k=0}^n e^{kt} {}_nC_k p^k q^{n-k} = \sum_{k=0}^n {}_nC_k (pe^t)^k q^{n-k} = (pe^t + q)^n.$$

したがって $M'_X(t) = npe^t(pe^t + q)^{n-1}$ これは $n \geq 1$ で成り立つ. さらに微分して $M''_X(t) = npe^t(pe^t + q)^{n-1} + n(n-1)p^2e^{2t}(pe^t + q)^{n-2}$. これも $n \geq 1$ で成立する. 従って M_X を用いて

$$\begin{aligned} E[X] &= M'_X(0) = np \\ V[X] &= E[X^2] - E[X]^2 \\ &= M''_X(0) - M'_X(0)^2 \\ &= np + n(n-1)p^2 - (np)^2 = npq. \end{aligned}$$

□

直接計算することもできる. 例えば $k {}_nC_k = \frac{n!}{(k-1)!(n-k)!} = n \frac{(n-1)!}{(k-1)!(n-k)!} = n {}_{n-1}C_{k-1}$ ($k \geq 1$) より

$$\begin{aligned} E[X] &= \sum_{k=0}^n k {}_nC_k p^k q^{n-k} \\ &= \sum_{k=1}^n n {}_{n-1}C_{k-1} p^k q^{n-k} \\ &= np \sum_{k=1}^n {}_{n-1}C_{k-1} p^{k-1} q^{n-1-(k-1)} \\ &= np(p+q)^{n-1} = np \end{aligned}$$

(2) ポアソン分布

ポアソン分布とは、それぞれの起こる確率は小さいが、考えている期間または回数が大きいため、一定の比率である現象が起こると考えられる現象の回数の従う分布である.

定義 24. $\lambda > 0$ とする.

(1) 0 以上の整数 $\{0, 1, 2, \dots\}$ 上の確率分布で $\{k\}$ の確率が

$$\mu(\{k\}) = e^{-\lambda} \frac{\lambda^k}{k!} \quad k = 0, 1, \dots$$

と与えられるものをパラメータ λ のポアソン分布と言う。

(2) (確率変数で言い換えると) 0 以上の整数に値を取る確率変数 X が

$$P(X = k) = e^{-\lambda} \frac{\lambda^k}{k!}$$

を満たすとき X はパラメータ λ のポアソン分布に従うという。

例 25. 次のデータはポアソン分布に従うと考えられる。

- (1) 馬に蹴られて死亡した 1 年間の兵士の数
- (2) ある地方で 1 年間で宝くじで一等が出た件数

ポアソン分布は二項分布の極限として得られる。 X を二項分布 $B(n, p)$ に従う確率変数とする。 $np = \lambda$ (期待値一定) の下で $n \rightarrow \infty$ としてみよう。

$$\begin{aligned} P(X = k) &= {}_n C_k p^k (1-p)^{n-k} \\ &= \frac{n!}{k!(n-k)!} \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{n-k} \\ &= \frac{n(n-1)\cdots(n-k+1)}{n^k} \left(1 - \frac{\lambda}{n}\right)^{-k} \frac{\lambda^k}{k!} \left(1 - \frac{\lambda}{n}\right)^n \rightarrow \frac{\lambda^k}{k!} e^{-\lambda}, \end{aligned}$$

ここで $\lim_{|x| \rightarrow \infty} \left(1 + \frac{1}{x}\right)^x = e$ を用いた。

定理 26. X をパラメータ λ のポアソン分布に従う確率変数とする。

- (1) $M_X(t) = e^{\lambda(e^t - 1)}$.
- (2) $E[X] = V[X] = \lambda$. またすべての n について $E[X^n] < \infty$.

(3) 幾何分布

表の出る確率が p ($0 < p < 1$) の硬貨を何回も投げる独立試行を繰り返す。 $k+1$ 回目の試行で初めて表が出る確率は $(1-p)^k p$ である。この確率分布をパラメータ p の幾何分布と言う。すなわち離散型確率変数 X がパラメータ p の幾何分布に従うとは

$$P(X = k) = (1-p)^k p, \quad k = 0, 1, 2, \dots$$

のときに言う。初めて表が出るまでの間に裏の出た回数 X の従う分布である。

定理 27. X がパラメータ p の幾何分布に従うとする。このとき

- (1) $M_X(t) = E[e^{tX}]$ は $t < -\log(1-p)$ ($-\log(1-p)$ は正数であることに注意) のとき有限で

$$M_X(t) = \frac{p}{1 - e^t(1-p)}.$$

- (2) $E[X] = \frac{1-p}{p}$, $V[X] = \frac{1-p}{p^2}$. すべての n について $E[X^n] < \infty$.

積率母関数の微分を計算しても平均値、分散は求まるが、次のように直接計算もできる。分散はどのようにして計算できるか各自考えてみて下さい。

$$\begin{aligned} E[X] &= \sum_{k=0}^{\infty} k(1-p)^k p = p(1-p) \sum_{k=0}^{\infty} k(1-p)^{k-1} = p(1-p) \left\{ -\frac{d}{dp} \left(\sum_{k=0}^{\infty} (1-p)^k \right) \right\} \\ &= p(1-p) \left\{ -\frac{d}{dp} \left(\frac{1}{p} \right) \right\} = \frac{1-p}{p}. \end{aligned}$$

命題 28 (幾何分布の無記憶性). X がパラメータ p の幾何分布に従うとする。任意の自然数 s, t について

$$P(X \geq t+s \mid X \geq s) = P(X \geq t).$$

3.2 連続型確率分布

(1) 一様分布 $[a, b]$ 上の一様分布とは密度関数が

$$f(x) = \begin{cases} \frac{1}{b-a} & a \leq x \leq b \text{ のとき} \\ 0 & x < a, x > b \text{ のとき} \end{cases}$$

で与えられるものを言う。これはすでに述べた。密度関数の $x = a, b$ における値を 0 にするか $1/(b-a)$ にするかは確率分布の定義にまったく影響を与えないことに注意。また、この分布の平均は $\frac{a+b}{2}$, 分散は $\frac{(b-a)^2}{12}$ 。

(2) 指数分布

指数分布は幾何分布の極限として得られる。時刻 $[0, \infty)$ を $\frac{1}{n}$ の長さに分割し、時刻 $t = \frac{k}{n}$ $k = 0, 1, \dots$ で独立にある事象 A が確率 $p = \frac{\lambda}{n}$ で起こるか $1 - \frac{\lambda}{n}$ で起こらないかのいずれかとする。 $\lambda > 0$ は定数である。 $np = \lambda$ (一定) で $n \rightarrow \infty$ を考えることにする。 T を初めて事象 A が起こる時刻とする。 $P(T = k/n) = (1 - \frac{\lambda}{n})^k \frac{\lambda}{n}$ (パラメータ λ/n の幾何分布) である。このとき時刻 t が区間 $[k_t/n, (k_t + 1)/n)$ に属すとする。

$$\begin{aligned} P(t \text{ を含む長さ } 1/n \text{ の時間のうちで } A \text{ が起こる}) &= \left(1 - \frac{\lambda}{n}\right)^{k_t} \frac{\lambda}{n} \\ &\doteq \left(1 - \frac{\lambda}{n}\right)^{nt} \frac{\lambda}{n} \quad (n \text{ が大きい時}) \\ &\doteq \lambda e^{-\lambda t} \frac{1}{n} \quad (n \text{ が大きい時}) \end{aligned}$$

ゆえに $P(T \in dt) = \lambda e^{-\lambda t} dt$ となり n 大のとき T は密度関数 $f(x) = \lambda e^{-\lambda x}$ ($x \geq 0$) をもつ確率分布に従う。

定義 29. 密度関数

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \geq 0 \text{ のとき} \\ 0 & x < 0 \text{ のとき} \end{cases}$$

をもつ確率分布をパラメータ λ の指数分布と言う。

一方、時刻 $t \in [k_t/n, (k_t + 1)/n)$ までに k 回事象 A が起こる確率 p_k は $p_k = {}_{k_t+1}C_k \left(\frac{\lambda}{n}\right)^k \left(1 - \frac{\lambda}{n}\right)^{k_t+1-k}$ 。これは二項分布 $B(k_t + 1, p)$ ($p = \lambda/n$) に従う。 $(k_t + 1) \frac{\lambda}{n} \doteq \lambda t$ だから $n \rightarrow \infty$ (したがって $k_t \rightarrow \infty$ となることに注意) で $\lim_{n \rightarrow \infty} p_k = \frac{(\lambda t)^k}{k!} e^{-\lambda t}$ 。パラメータ λt のポアソン分布に収束する。まとめると

まとめ

(I) 一回の試行では起こるのがまれな事象を考える。独立な試行を多数繰り返すことにより、その事象がある一定回数起こる状況になっているとき、生起回数の分布はポアソン分布に従う。そのポアソン分布を特徴づけるパラメータは平均生起回数 λ である。

(II) (I) の状況で事象が起こる時間間隔の分布はパラメータ λ の指数分布に従う。指数分布は幾何分布の極限として得られる。

定理 30. X がパラメータ λ の指数分布に従うとき、 $M_X(t) = \frac{\lambda}{\lambda - t}$ ($t < \lambda$), $E[X] = \frac{1}{\lambda}$, $V[X] = \frac{1}{\lambda^2}$, $E[X^n] = \frac{n!}{\lambda^n}$ 。

幾何分布と同様、指数分布も無記憶性を持つことがわかる。

命題 31. X はパラメータ λ の指数分布に従うとする。すべての定数 t, s について

$$P(X \geq t + s \mid X \geq s) = P(X \geq t).$$

これは $P(X \geq t) = e^{-\lambda t}$ から直ちにわかる。

(3) 正規分布

定義 32. $m \in \mathbb{R}$, $\sigma > 0$ とする. 確率密度関数

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

をもつ連続型の $\mathbb{R}$ 上の確率分布を正規分布といい、 $N(m, \sigma^2)$ と表す. $m = 0, \sigma = 1$ の分布 $N(0, 1)$ を標準正規分布と言う.

注意 33. $\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = 1$ は大学 1 年のときに学んだ (はずの) $\int_{\mathbb{R}} e^{-x^2} dx = \sqrt{\pi}$ の式で適当に変数変換して示される.

命題 34. 確率変数 X が正規分布 $N(m, \sigma^2)$ に従うとする.

(1) $M_X(t) = e^{mt - \frac{\sigma^2 t^2}{2}}$.

(2) 確率変数 X が正規分布 $N(m, \sigma^2)$ に従うとする. このとき $E[X] = m$, $V[X] = \sigma^2$.

(3) p, q を実数とする. 確率変数 $pX + q$ の分布は正規分布 $N(pm + q, p^2\sigma^2)$ である.

(3) は正規分布に従う確率変数を線形変換してもやはり正規分布に従うことを示している (ただし, 平均、分散は一般には変わってしまうことに注意). これは例えば次のように示される. まず次に注意する.

命題 35. Y を確率変数とする. 次の (1), (2) は同値である.

(1) Y の確率分布は密度関数 f を持つ.

(2) 任意の有界関数 φ について $E[\varphi(Y)] = \int_{-\infty}^{\infty} \varphi(x)f(x)dx$.

命題 34 (3) を示すため, $\varphi(pX + q)$ の期待値を計算してみよう. 簡単のため $p > 0$ とする.

$$\begin{aligned} E[\varphi(pX + q)] &= \int_{-\infty}^{\infty} \varphi(px + q) \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx = \int_{-\infty}^{\infty} \varphi(t) \frac{1}{p\sqrt{2\pi}\sigma} e^{-\frac{(\frac{t-q}{p}-m)^2}{2\sigma^2}} dt \\ &= \int_{-\infty}^{\infty} \varphi(t) \frac{1}{\sqrt{2\pi p^2}\sigma} e^{-\frac{(t-q-m p)^2}{2p^2\sigma^2}} dt. \end{aligned}$$

従って $\frac{1}{\sqrt{2\pi p^2}\sigma} e^{-\frac{(t-q-m p)^2}{2p^2\sigma^2}}$ が密度関数になる. また, X が正規分布に従うとき $\frac{X-m}{\sigma}$ は標準正規分布に従う. このことから標準正規分布に従う確率変数 T について分布関数 $F(x) = P(T \leq x)$ がわかれば X の分布もわかることになる. というのは

$$\begin{aligned} P(a \leq X \leq b) &= P\left(\frac{a-m}{\sigma} \leq \frac{X-m}{\sigma} \leq \frac{b-m}{\sigma}\right) \\ &= P\left(\frac{a-m}{\sigma} \leq T \leq \frac{b-m}{\sigma}\right) \\ &= F\left(\frac{b-m}{\sigma}\right) - F\left(\frac{a-m}{\sigma}\right) \end{aligned}$$

となるから.

$$F(z) = \int_{-\infty}^z \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$$

であるが、この積分は簡単な関数では表されないことが知られている。しかし、近似値は計算することは可能。それをまとめたものは正規分布表と言うもので、例えば教科書の巻末 139 ページに見られる。教科書の巻末では $I(z) = \int_0^z \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} dx$ の表がのっているが $z > 0$ なら

$$F(z) = \frac{1}{2} + I(z)$$

だから $F(z)$ の近似値も簡単にわかる。

注意 36. 命題 34 の証明においては次の式変形を何回も用いていることに注意せよ。 $t = px + q$ ($p > 0$) と変数変換すると

$$\int_{-\infty}^{\infty} f(px + q)g(x)dx = \int_{-\infty}^{\infty} f(t)\frac{1}{p}g\left(\frac{t - q}{p}\right)dt.$$

4 多次元の確率分布

4.1 2 次元の確率分布

定義 37. (1) X, Y を同じ確率空間 $(\Omega, \mathcal{F}, P)$ で定義された確率変数とする。 $\mathbb{R}^2$ 上の値を取る確率変数の組 (確率ベクトルという) $\mathbb{X} = {}^t(X, Y)$ について $A \subset \mathbb{R}^2$ の確率を

$$\mu_{\mathbb{X}}(A) = P(\mathbb{X} \in A)$$

で定めることができる。この $\mathbb{R}^2$ 上の確率を $\mathbb{X} = {}^t(X, Y)$ の確率分布と言う。また X, Y の同時分布 (または結合分布) とも言う。これに対して X, Y のそれぞれの $\mathbb{R}$ 上の分布を周辺分布と言う。

上の定義で ${}^t(X, Y)$ は転置を取って縦ベクトルに直していることに注意して下さい。後で行列 A との積 $A\mathbb{X}$ を考えるため、縦ベクトルを考えています。 X, Y が離散型、連続型の場合にこの定義 37 を見直してみよう。

(1) X, Y が離散型のとき

$P(X = a_i) = p_i$ ($1 \leq i \leq M$), $P(Y = b_j) = q_j$ ($1 \leq j \leq N$) とする。 ${}^t(X, Y)$ は $\{{}^t(a_i, b_j)\}_{1 \leq i \leq M, 1 \leq j \leq N}$ のいずれかの点の値を取る。

$$P({}^t(X, Y) = {}^t(a_i, b_j)) = p_{ij}$$

とする。 $\sum_{1 \leq i \leq M, 1 \leq j \leq N} p_{ij} = 1$ であり ${}^t(a_i, b_j)$ に確率 p_{ij} があるような $\mathbb{R}^2$ 上の離散型の確率分布が同時分布となる。

命題 38 (同時分布と周辺分布の関係)。

上の (1) のような状況で $\sum_{j=1}^N p_{ij} = p_i$ ($1 \leq i \leq M$), $\sum_{i=1}^M p_{ij} = q_j$ ($1 \leq j \leq N$) が成り立つ。

周辺分布が同じでも同時分布は異なるような分布が存在する。各自例を考えて見よ。

(2) X, Y の同時分布が密度関数を持つ連続型のとき

${}^t(X, Y)$ の同時分布が $f(x, y)$ という密度関数を持つ場合を考える。これは任意の $-\infty < a < b < +\infty, -\infty < c < d < +\infty$ に対して

$$P(a \leq X \leq c, c \leq Y \leq d) = \int_a^b \int_c^d f(x, y) dx dy$$

となる時に言う。命題 38 に対応して次の命題が成立する。

命題 39. $t(X, Y)$ の同時分布が密度関数 $f(x, y)$ をもつ連続型のとき X の分布, Y の分布もそれぞれ密度関数 $f_1(x), f_2(y)$ をもち

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

で与えられる.

同時分布の概念を用いると確率変数の期待値 (平均) の線形性を証明できる.

定理 40. 確率変数 X, Y と実数 α, β について $E[\alpha X + \beta Y] = \alpha E[X] + \beta E[Y]$.

証明. X, Y が離散型で $P(X = a_i) = p_i$ ($1 \leq i \leq M$), $P(Y = b_j) = q_j$ ($1 \leq j \leq N$),

$$P(t(X, Y) = t(a_i, b_j)) = p_{ij}$$

とする. このとき $\alpha X + \beta Y$ は $\{\alpha a_i + \beta b_j \mid 1 \leq i \leq M, 1 \leq j \leq N\}$ のいずれかの値を取る離散型の確率変数である. 簡単のため $\{\alpha a_i + \beta b_j \mid 1 \leq i \leq M, 1 \leq j \leq N\}$ がすべて異なる値とする. このとき $P(\alpha X + \beta Y = \alpha a_i + \beta b_j) = P(X = a_i, Y = b_j) = p_{ij}$ ゆえ

$$\begin{aligned} E[\alpha X + \beta Y] &= \sum_{1 \leq i \leq M, 1 \leq j \leq N} (\alpha a_i + \beta b_j) p_{ij} \\ &= \sum_{1 \leq i \leq M} \alpha a_i \left(\sum_{j=1}^N p_{ij} \right) + \sum_{1 \leq j \leq N} \beta b_j \left(\sum_{i=1}^M p_{ij} \right) \\ &= \alpha \sum_{1 \leq i \leq M} a_i p_i + \beta \sum_{1 \leq j \leq N} b_j q_j \\ &= \alpha E[X] + \beta E[Y]. \end{aligned}$$

□

注意 41. 上の証明の式の第一行目で $\{\alpha a_i + \beta b_j\}_{1 \leq i \leq M, 1 \leq j \leq N}$ がすべて相異なる数とは限らない場合でもこのように書いてもよいことに注意してほしい. 従って上記の証明は常に正しい. 理由はわかりますか?

4.2 確率変数の独立性

定義 42. 同じ確率空間 $(\Omega, \mathcal{F}, P)$ で定義された確率変数 X, Y が独立であるとはすべての実数 $-\infty < a < b < +\infty, -\infty < c < d < +\infty$ に対して

$$P(a \leq X \leq b, c \leq Y \leq d) = P(a \leq X \leq b)P(c \leq Y \leq d)$$

が成立するときに言う.

注意 43. X, Y が独立とはすべての実数 $-\infty < a < b < +\infty, -\infty < c < d < +\infty$ に対して事象 $\{\omega \mid a \leq X(\omega) \leq b\}$ と事象 $\{\omega \mid c \leq Y(\omega) \leq d\}$ とが独立になるということと同値である. 事象の独立性の定義を見よ.

例 44. サイコロを投げるという試行を独立に 2 回繰り返す. このとき確率空間は $\Omega = \{(i, j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$ となり, 確率は事象 $A \subset \Omega$ に対して $P(A) = A$ の要素の数/36 と定まる.

$$X_1(\omega) = i \quad (\omega = (i, j) \text{ のとき})$$

$$X_2(\omega) = j \quad (\omega = (i, j) \text{ のとき})$$

と定めると X_1 は 1 回目の目の数を表す確率変数, X_2 は 2 回目の目の数を表す確率変数で独立である. しかし, $X_1 + X_2$ と $X_1 - X_2$ は独立ではない.

命題 45. (1) X, Y を離散型の確率変数で $P(X = a_i) = p_i$ ($1 \leq i \leq M$), $P(Y = b_j) = q_j$ ($1 \leq j \leq N$),

$$P({}^t(X, Y) = {}^t(a_i, b_j)) = p_{ij}$$

とする. X, Y が独立である必要十分条件は $p_{ij} = p_i q_j$ ($1 \leq i \leq M, 1 \leq j \leq N$) となることである.

(2) X, Y がそれぞれ密度関数 $f_1(x), f_2(y)$ をもつ連続型確率分布に従う確率変数でその同時分布も密度関数 $f(x, y)$ をもつとする. X, Y が独立であるための必要十分条件はすべての (x, y) について $f(x, y) = f_1(x)f_2(y)$ となることである.

以上の命題を用いると次の定理が証明できる.

定理 46. (1) X, Y を独立な確率変数とすると $E[XY] = E[X]E[Y]$.

(2) X, Y を独立な確率変数とすると $V[\alpha X + \beta Y] = \alpha^2 V[X] + \beta^2 V[Y]$.

(3) X, Y の共分散を $\text{Cov}(X, Y) = E[(X - E[X])(Y - E[Y])]$ と定義する. X, Y が独立ならば $\text{Cov}(X, Y) = 0$.

注意 47. (1) 二つの確率変数 X, Y について $E[XY] = E[X]E[Y]$ だとしても X, Y は独立とは限らない.

(2) X, Y が独立ということと有界な関数 ϕ, ψ について常に $E[\phi(X)\psi(Y)] = E[\phi(X)]E[\psi(Y)]$ となるということとは同値である.

4.3 多次元の確率分布

定義 48. (1) $X_1, \dots, X_n$ を同じ確率空間で定義された確率変数とする. n 次元確率ベクトル $\mathbb{X} = {}^t(X_1, \dots, X_n)$ に対して $\mathbb{R}^n$ 上の確率分布を

$$\mu_{\mathbb{X}}(A) = P(\mathbb{X} \in A) \quad (A \subset \mathbb{R}^n)$$

のように定めることができる. $\mu_{\mathbb{X}}$ を $\mathbb{X}$ の確率分布、(確率) 法則と言う.

(2) n 次元確率ベクトル $\mathbb{X} = {}^t(X_1, \dots, X_n)$ の分布が密度関数 $f(x_1, \dots, x_n)$ を持つとは、任意の部分集合 $A \subset \mathbb{R}^n$ に対して

$$P(\mathbb{X} \in A) = \int_A f(x_1, \dots, x_n) dx_1 \cdots dx_n$$

となる時に言う. $\mathbb{X}$ の分布が離散型であるというのも確率変数 (1 次元確率ベクトルと言っても同じ) の場合と同様に定義する.

離散型の確率分布の代表例は多項分布、密度関数を持つ代表的な確率分布は多次元正規分布である.

定義 49. (1) 1 回の試行の結果、 $\{E_1, \dots, E_N\}$ のいずれかがそれぞれ確率 $p_1, \dots, p_N$ で起こるとする. n 回独立にこの試行を繰り返すとき状態 E_i の起こった回数を X_i とすると

$$P(X_1 = k_1, \dots, X_N = k_N) = \frac{n!}{k_1! \cdots k_N!} p_1^{k_1} \cdots p_N^{k_N}$$

となる ($n = \sum_{i=1}^N k_i$ に注意せよ). $\mathbb{X} = {}^t(X_1, \dots, X_N)$ の分布を多項分布と言う. これは二項分布 ($N = 2$ の場合に当たる) の一般化である.

$\mathbb{R}^N$ 上の分布であるが $\sum_{i=1}^N X_i = n$ であるから本質的に $N - 1$ 次元空間上の分布である. 二項分布の場合は二つの状態 (成功、失敗) しかないので、どちらかの状態の起こった回数のみ決めれば残りの状態の起こった回数も決まることに対応している.

(2) A を $n \times n$ -狭義正定値対称行列、 m を n 次元ベクトルとする. 密度関数 $f(x_1, \dots, x_n)$ が

$$f(x_1, \dots, x_n) = \frac{1}{\sqrt{2\pi^n} \sqrt{\det A}} \exp\left(-\frac{1}{2} (x - m, A^{-1}(x - m))\right)$$

で与えられる連続型の確率分布を多次元正規分布と言い $N(m, A)$ と表す. これは 1 次元の正規分布の拡張である.

定理 50. $\mathbb{X} = {}^t(X_1, \dots, X_n)$ を $N(m, A)$ に従う正規分布とする. このとき $E[X_i] = m_i$, $V[X_i] = a_{ii}$, $\text{Cov}(X_i, X_j) = a_{ij}$. ただし a_{ij} は A の (i, j) 成分である.

$$E[\mathbb{X}] = m, \quad E[(\mathbb{X} - m)^t(\mathbb{X} - m)] = A \quad (1)$$

と書くと印象的である. もちろんこの式の意味は左辺の期待値はベクトル (行列) の各成分ごとに期待値を計算していると理解するべきである. $\mathbb{X}, m$ は $n \times 1$ 行列だから行列の積 $(\mathbb{X} - m)^t(\mathbb{X} - m)$ は $n \times n$ 行列で (i, j) 成分は $(X_i - m_i)(X_j - m_j)$ ある.

命題 51. (1) $\mathbb{X} = {}^t(X_1, \dots, X_n)$ が離散型の確率変数で $\mathbb{X}$ の取る値 (ベクトル) を $\{a_1, \dots, a_N\}$ とし, $P(\mathbb{X} = a_i) = p_i$ ($1 \leq i \leq N$) とする. このとき $\mathbb{R}^n$ 上の有界関数 φ について

$$E[\varphi(\mathbb{X})] = \sum_{i=1}^N \varphi(a_i) p_i.$$

(2) $\mathbb{X} = {}^t(X_1, \dots, X_n)$ の結合分布が連続型の分布で確率密度関数 $f(x_1, \dots, x_n)$ を持つとき, $\mathbb{R}^n$ 上の有界関数 φ に対して

$$E[\varphi(\mathbb{X})] = \int_{\mathbb{R}^n} \varphi(x_1, \dots, x_n) f(x_1, \dots, x_n) dx_1 \cdots dx_n.$$

定義 52. (1) 確率変数 $X_1, \dots, X_n$ が独立とは任意の区間 $I_1, \dots, I_n$ に対して

$$P(X_1 \in I_1, \dots, X_n \in I_n) = P(X_1 \in I_1) \cdots P(X_n \in I_n)$$

となる時に言う.

(2) 確率変数の無限列 $\{X_1, X_2, \dots, X_n, \dots\}$ が独立とは任意の n について $\{X_1, \dots, X_n\}$ が独立の時に言う.

(3) 確率変数の無限列 $\{X_1, X_2, \dots, X_n, \dots\}$ が独立で同分布に従う

(*independent and identically distributed, i.i.d.* と略記する) とは各 X_i の分布がすべて同じかつ $\{X_i\}_{i=1}^\infty$ が独立であるときに言う.

命題 53. n 個の確率変数 $\{X_1, \dots, X_n\}$ を考える. 各 i に対して X_i が密度関数 $f_i(x)$ を持つ連続型確率変数であるとする. $\{X_1, \dots, X_n\}$ が独立であるための必要十分条件は n 次元確率ベクトル $\mathbb{X} = {}^t(X_1, \dots, X_n)$ の分布が密度関数 $f(x_1, \dots, x_n) = f_1(x_1) \cdots f_n(x_n)$ を持つ事である.

確率変数 X が正規分布に従うとその線形変換 $pX + q$ も正規分布になった. 多次元正規分布に従う多次元確率変数を線形変換してもやはり多次元正規分布に従う確率変数が得られる.

定理 54. $\mathbb{X}$ を正規分布 $N(m, A)$ に従う n 次元確率変数とする. P を (l, n) 行列で階数が l とする. このとき l 次元確率変数 $P\mathbb{X}$ は正規分布 $N(Pm, PA^tP)$ に従う. (tP は P の転置行列を表す)

共分散行列は

$$\begin{aligned} E[P(\mathbb{X} - m)^t(P(\mathbb{X} - m))] &= E[P(\mathbb{X} - m)^t(\mathbb{X} - m)^tP] \\ &= PE[(\mathbb{X} - m)^t(\mathbb{X} - m)]^tP \\ &= PA^tP. \end{aligned}$$

のように計算できる. 命題 53 および上の定理を用いると次の結果を得る.

定理 55. X_i ($1 \leq i \leq n$) をそれぞれ $N(m_i, \sigma_i^2)$ に従う独立な確率変数とするとその結合分布は n 次元正規分布 $N(m, A)$ に従う確率変数である. ただし $m = {}^t(m_1, \dots, m_n)$, A は対角行列で (i, i) 成分は σ_i^2 . また, $Y = \sum_{i=1}^n a_i X_i$ とすると Y の分布は正規分布 $N(\sum_{i=1}^n a_i m_i, \sum_{i=1}^n a_i^2 \sigma_i^2)$ に従う.

もう一つ正規分布に従う確率変数の大事な性質を述べる.

定理 56. 確率変数 X, Y の結合分布が正規分布に従うとする. X と Y が独立であるための条件は $\text{Cov}(X, Y) = 0$ である.

例 57. X_1, X_2 をそれぞれ標準正規分布に従う独立な確率変数とする. このとき $X_1 + X_2$ と $X_1 - X_2$ は独立である. これは $\text{Cov}(X_1 + X_2, X_1 - X_2) = E[(X_1 + X_2)(X_1 - X_2)] = 1 - 1 = 0$ と定理 56 からわかる. しかし例えばサイコロ投げで i 回目の出目を X_i としたとき X_1 と X_2 は独立だが $X_1 + X_2$ と $X_1 - X_2$ は独立ではないことを注意した. 正規分布に従うということが効いているのである.

定理 55 と定理 56 は William Gosset(1876–1937) による t -分布 (Gosset のペンネーム "Student" を冠して Student の t -分布とも言う, ギネス社は機密保持のため社員が論文を出版することを禁じていたため, ペンネームを用いた) の計算で重要な役割を果たす.

5 極限定理

5.1 大数の法則

確率変数 $X_1, X_2, \dots, X_n, \dots$ に対して

$$\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$$

と定め, 経験的な平均, 標本平均と呼ぶ.

定理 58. (1) 確率変数 $X_1, \dots, X_n$ のおおのの期待値がすべて m ならば $E[\bar{X}_n] = m$. すなわち標本平均の期待値も m .

(2) 確率変数 $X_1, \dots, X_n$ が独立ですべて平均 m , 分散 σ^2 ならば $\bar{X}_n$ の分散は $\frac{\sigma^2}{n}$.

定理 59 (大数の法則). $X_1, X_2, \dots$ をすべて同じ平均 m , 分散 σ^2 をもつ独立確率変数とする. $S_n = X_1 + \dots + X_n$, $\bar{X}_n = \frac{S_n}{n}$ とする. 任意の正数 ε について

$$\lim_{n \rightarrow \infty} P(|\bar{X}_n - m| > \varepsilon) = 0.$$

すなわち経験的な平均 $\bar{X}_n$ が期待値 m に収束していくことがわかる. 例えば

(1) サイコロ投げで $X_i = i$ 回目の出目 とすると X_i は独立ですべて同じ分布に従う確率変数となる. 平均 (期待値) は $m = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \dots + \frac{1}{6} \times 6 = 3.5$ だから $\frac{X_1 + \dots + X_n}{n}$ はランダムな量だが n が大きくなると 3.5 に近づいていくと考えられる.

(2) 無限に投げ続ける硬貨投げで $X_i = i$ 回目の硬貨投げで表が出たら 1, 裏が出たら 0 という確率変数を考える. $\bar{X}_n = \frac{S_n}{n}$ は n 回の硬貨投げで表の出た比率である. 公平な硬貨投げであれば $\bar{X}_n$ は $\frac{1}{2}$ に近づいていくことがわかる.

定理 59 は Chebyshev(チェビシェフ) の不等式を用いて得られる式

$$P(|\bar{X}_n - m| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2}$$

から従う。

5.2 中心極限定理

正規分布が重要な分布だと述べたがそれは、独立確率変数の多数の和を考えると正規分布で近似できると考えられるからである。これは経験的に知られていた事であるが、数学の定理としては以下のように述べられる。

$X_1, X_2, \dots, X_n, \dots$ を同じ期待値 m , 分散 σ^2 をもつ独立な確率変数とする。 $S_n = X_1 + X_2 + \dots + X_n$ とおき

$$T_n = \frac{\sqrt{n} \cdot (\bar{X}_n - m)}{\sigma} = \frac{S_n - nm}{\sqrt{n\sigma^2}}$$

と定める。大数の法則によれば $\frac{\bar{X}_n - m}{\sigma}$ は 0 に近づいて行くことがわかるがこれを $\sqrt{n}$ 倍して

$$E[T_n] = 0, \quad V[T_n] = 1$$

のように正規化されているのである。したがって $n \rightarrow \infty$ でも T_n に関しては、なんらかの意味のある量が残ると期待できる。この極限が標準正規分布である。

定理 60 (中心極限定理). $X_1, X_2, \dots, X_n, \dots$ を同じ分布に従う期待値 m , 分散 σ^2 をもつ独立な確率変数とする。 $S_n = X_1 + X_2 + \dots + X_n$ とおく。 S_n は二項分布 $B(n, p)$ に従う。すべての $a < b$ となる実数について

$$\lim_{n \rightarrow \infty} P(a \leq T_n \leq b) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx.$$

注意 61. 大数の法則は標本平均と期待値の差の値が 0 に収束することを述べている。しかし、中心極限定理は $\frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ の値が収束すると言っているのではなく確率が収束すると述べていることに注意してほしい。

X_i ($i = 1, 2, \dots$) が $P(X_i = 1) = p, P(X_i = 0) = 1 - p$ となる独立確率変数とすると $E[X_i] = p, V[X_i] = p(1 - p)$ だから

$$\lim_{n \rightarrow \infty} P\left(a \leq \frac{S_n - np}{\sqrt{np(1-p)}} \leq b\right) = \int_a^b \frac{1}{\sqrt{2\pi}} e^{-\frac{x^2}{2}} dx$$

となる。これを de Moivre-Laplace(ド・モアヴル-ラプラス)の定理と言う。ド・モアヴル-ラプラスの定理は Stirling(スターリング)の公式

$$\lim_{n \rightarrow \infty} \frac{n!}{\sqrt{2\pi n} n^n e^{-n}} = 1$$

などを使って証明できるが一般の中心極限定理の証明は別の道具を使う必要がある。 $S_n = np + \sqrt{np(1-p)}T_n$ と書けるから二項分布 $B(n, p)$ は n が大きい時正規分布 $N(np, np(1-p))$ で近似できることになる。それを使って次の問題を考えよう。

問題 硬貨を 100 回投げたとき、表の出た回数を S とする $P(45 \leq S \leq 55)$ の確率を求めよ。

解 (1) $n = 100, p = 1/2$ だから S は正規分布 $N(50, 25)$ に従う確率変数 $\tilde{S}$ で近似できると考えられる。半目(半整数)の補正をして

$$P(45 \leq S \leq 55) \doteq P(44.5 \leq \tilde{S} \leq 55.5)$$

$\tilde{S}$ の正規化 $T = \frac{\tilde{S}-50}{\sqrt{25}}$ は標準正規分布に従う. よって 139 ページの表を用いて

$$\begin{aligned} P(44.5 \leq \tilde{S} \leq 55.5) &= P\left(\frac{44.5-50}{5} \leq T \leq \frac{55.5-50}{5}\right) \\ &\doteq P(-1.1 \leq T \leq 1.1) = 2 \times 0.364 = 0.728. \end{aligned}$$

半目の補正をしないと $P(-1 \leq T \leq 1) = 0.68$.

(2)

$$P(45 \leq S \leq 55) = \sum_{k=45}^{55} {}_{100}C_k \left(\frac{1}{2}\right)^{100}$$

計算機を使って近似計算するとこれは 0.7281 になる. ゆえに半目の補正をした方が正確. これは n があまり大きくない時に顕著である.

6. 標本分布

6.1 標本調査、母集団

定義 62. (1) 母集団

「日本人全員の身長」, 「学力テストの受験者全員の得点」などの数値の集合 (数値をまとめた多次元のベクトルを考えることもある) を母集団と言う. これらの数はある確率分布に従って分布していると考え. 母集団が従っている分布を母分布と言う. 正規分布に従っている母集団を正規母集団と言う.

(2) 母平均、母分散

母分布の平均、分散のこと. 母分布を特徴付けるこれらの量を母数と言う.

(3) 標本調査

母集団からある一定の操作で数値を選ぶことを標本調査、選ばれた値の集合を標本と言う. 標本は無作為に選ばなければならない. 無作為に選ばれた標本

$$X_1, X_2, \dots, X_n, \dots$$

は母集団の分布に従う独立な確率変数であると考えられる. 実際に標本調査を行って得られるデータはこの確率変数の実現値と考えられる. 限られた数の標本から母集団を特徴付ける母数を推定するのが統計学の役割である.

(3) 統計量

標本 $X_1, X_2, \dots$ から計算される以下のような量を統計量と言う.

(a) $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ 標本平均と言う.

(b) $s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ 標本分散と言う.

(c) $u_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ 不偏分散と言う.

標本 $\{X_1, \dots, X_n, \dots\}$ が無作為抽出されたものならば同じ分布に従う独立確率変数になるので、大数の法則に従う. 従って分布の平均を m , σ^2 とすれば

$$\begin{aligned} \lim_{n \rightarrow \infty} P(|\bar{X}_n - m| \geq \varepsilon) &= 0 & (\clubsuit) \\ \lim_{n \rightarrow \infty} P\left(\left|\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2 - \sigma^2\right| \geq \varepsilon\right) &= 0 & (\diamond) \end{aligned}$$

となるので、標本平均、標本分散は平均 m 、分散 σ^2 を近似する統計量である。上の $(\diamond)$ は $\bar{X}_n$ の所が m になっていれば大数の法則 (定理 59) から直ちに従うがそうでないので、証明には少し計算が必要である。

注意 63. $\bar{X}_n, s_n^2$ は平均、分散を近似する量という意味があるが、不偏分散 u_n^2 はどんな意味があるのだろうか？もちろん $u_n^2 = \frac{n}{n-1} s_n^2$ だから n が大きい時は分散を近似しているのは同じですが。 u_n^2 は

$$E[u_n^2] = \sigma^2$$

をみたすという意味でよい統計量 (不偏統計量という) なのです。

6.2 χ^2 分布, t 分布

母集団の分布の分散がわかっていないとき無作為抽出された標本

$$X_1, X_2, \dots, X_n, \dots$$

から得られる標本分散 $s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$, 不偏分散 $u_n^2 = \frac{1}{n-1} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ で分散の代用になることがある。そのさい、 s_n^2, u_n^2 がどんな分布に従っているか知る必要がある。ここでは、母集団分布が正規分布に従う場合にこの問題を考える。その最初のステップとして次を述べる。

定理 64. $Z_1, \dots, Z_n$ を標準正規分布に従う独立確率変数とする。 $\chi_n^2 = \sum_{i=1}^n Z_i^2$ と定めると χ_n^2 の分布は密度関数

$$f_n(x) = \begin{cases} 0 & x \leq 0 \text{ のとき} \\ \frac{1}{2^{\frac{n}{2}} \Gamma(\frac{n}{2})} x^{\frac{n-2}{2}} e^{-\frac{x}{2}} & x > 0 \text{ のとき} \end{cases}$$

を持つ。

ただし

$$\Gamma(s) = \int_0^\infty x^{s-1} e^{-x} dx \quad (s > 0)$$

である (ガンマ関数)。 $\Gamma(n) = (n-1)!$ ($n = 1, 2, \dots$) のように整数値の値は簡単にわかる。

定理 64 は任意の有界な関数 φ について

$$E[\varphi(\sum_{i=1}^n Z_i^2)] = \int_{\mathbb{R}} \varphi(x) f_n(x) dx$$

が成立すること (これは変数変換をして示される) と命題 35 から得られる。例えば $n = 2$ のときは次のような計算になる。

$$\begin{aligned} E[\varphi(Z_1^2 + Z_2^2)] &= \int_{\mathbb{R}^2} \varphi(x^2 + y^2) \frac{e^{-\frac{x^2+y^2}{2}}}{2\pi} dx dy \quad (x = r \cos \theta, y = r \sin \theta \text{ と極座標に変換}) \\ &= \int_{r=0}^\infty \int_{\theta=0}^{2\pi} \varphi(r^2) \frac{e^{-\frac{r^2}{2}}}{2\pi} r dr d\theta \\ &= \int_0^\infty \varphi(r^2) e^{-\frac{r^2}{2}} r dr \quad (r^2 = t \text{ と変換}) \\ &= \int_0^\infty \varphi(t) \frac{e^{-\frac{t}{2}}}{2} dt \end{aligned}$$

したがって密度関数は $t > 0$ で $f_2(t) = \frac{e^{-t/2}}{2}$, $t \leq 0$ で $f_2(t) = 0$ となる。 $\mathbb{R}^3$ でも同様な極座標への変換公式を学んだ人もいると思う。同様に計算して見て下さい。

定義 65. 定理 64 に現れて来ている χ_n^2 の従う分布を自由度 n の χ^2 分布と言う.

自由度 n の χ^2 分布の平均は n である.

系 66. $X_1, \dots, X_n, \dots$ が正規分布 $N(m, \sigma^2)$ に従う独立確率変数のとき確率変数

$$\sum_{i=1}^n \left(\frac{X_i - m}{\sigma} \right)^2$$

は自由度 n の χ^2 分布に従う.

これらの結果を使うと次がわかる.

定理 67. $X_1, \dots, X_n$ を正規分布 $N(m, \sigma^2)$ に従う独立確率変数とする.

$$\frac{ns_n^2}{\sigma^2} = \frac{(n-1)u_n^2}{\sigma^2} = \sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2$$

は自由度 $n-1$ の χ^2 分布に従う.

注意 68. この定理に関して注意を述べる. $Y_i = \frac{X_i - \bar{X}_n}{\sigma}$ とおく. 定理 55 を用いると Y_i は平均 0 分散 $\frac{n-1}{n}$ の正規分布に従う. したがって $E[\sum_{i=1}^n Y_i^2] = \sum_{i=1}^n E[Y_i^2] = n-1$ となり自由度 $n-1$ の χ^2 分布の平均と一致する. $\sum_{i=1}^n Y_i = 0$ からわかるように $Y_1, \dots, Y_n$ は独立ではなく、自由度が 1 減った χ^2 分布が現れることになる. もう少し $n-1$ になる理由を述べる. $Z_i = \frac{X_i - m}{\sigma}$ とおくと $\{Z_i\}_{i=1}^n$ は標準正規分布に従う独立確率変数である. 線形代数の知識を使うと

$$\begin{aligned} \sum_{j=1}^n a_{ij}^2 &= 1 & (1 \leq i \leq n-1) \\ \sum_{j=1}^n a_{ij}a_{kj} &= 0 & (k \neq i) \end{aligned}$$

をみたす実数 $\{a_{ij}\}_{1 \leq i \leq n-1, 1 \leq j \leq n}$ の組みをうまく選び新たな確率変数

$$U_i = \sum_{j=1}^n a_{ij}Z_j \quad (1 \leq i \leq n-1)$$

を導入すると

$$\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2 = \sum_{i=1}^{n-1} U_i^2$$

と書き直せることがわかる. 定理 56 によれば $U_1, \dots, U_{n-1}$ は標準正規分布に従う独立確率変数だから $\sum_{i=1}^n \left(\frac{X_i - \bar{X}_n}{\sigma} \right)^2$ は自由度 $n-1$ の χ^2 分布に従うとわかる.

ある母集団の分布が正規分布 $N(m, \sigma^2)$ に従っている場合、無作為抽出して得られる標本 $X_1, \dots, X_n$ に対して統計量

$$T_n = \frac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$$

は標準正規分布に従う。しかしこの統計量を考えるには分散がわかっていなければならない。分散がわからないときを考えよう。大数の法則によれば n が大きい時、標本分散 $s_n^2 = \frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2$ は分散を近似するから T_n の代わりとして

$$\tilde{T}_n = \frac{\sqrt{n}(\bar{X}_n - m)}{s_n} = \frac{S_n - nm}{\sqrt{\sum_{i=1}^n (X_i - \bar{X}_n)^2}}$$

を考えても n が大きい時はやはり標準正規分布で近似できることがわかる。しかし、 n がそんなに大きく無いときは、 s_n^2 は分散に近いわけではないから標準正規分布で近似できるとは言えない。 n が大きくないときは $\tilde{T}_n$ の分布がなんなのかははっきりさせねばならない。実際、ギネスビールの技術者 William Gosset はビール醸造に関する実験を多数繰り返して行うことができないすなわち、大量の標本を得られないということから、 n が大きくない状況 (小標本の問題) に直面した。その結果 $\tilde{T}_n$ の分布を研究することになり、 t 分布と呼ばれる分布の発見につながった。実際は s_n^2 を不偏分散 u_n^2 で置き換えたものが t 分布に従うのである。

定義 69. X を標準正規分布に従う確率変数、 Y を自由度 n の χ^2 分布に従う確率変数で X と Y は独立とする。確率変数

$$t(n) = \frac{X}{\sqrt{Y/n}}$$

の分布は密度関数

$$f_n(x) = \frac{1}{n^{1/2} B(n/2, 1/2)} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}}$$

を持つ。 $t(n)$ の従う分布を自由度 n の t 分布と言う。

ここに

$$B(p, q) = \int_0^1 x^{p-1} (1-x)^{q-1} dx \quad p > 0, q > 0$$

はオイラーのベータ関数と呼ばれる関数。

証明. $n = 2$ の場合に証明してみよう。 $x = \frac{1}{2} + t$ ($-\frac{1}{2} \leq t \leq \frac{1}{2}$), $t = \frac{1}{2} \sin \theta$ ($-\pi/2 \leq \theta \leq \pi/2$) と変換すると $B(1/2, 1/2) = \int_{-1/2}^{1/2} \frac{4}{\sqrt{1-4t^2}} dt = \int_{-\pi/2}^{\pi/2} 2d\theta = 2\pi$ だから $f_2(x) = \frac{1}{2^{3/2}\pi} \left(1 + \frac{x^2}{2}\right)^{-3/2}$. φ を有界な関数として

$$E \left[\varphi \left(\frac{\sqrt{2}X}{Y} \right) \right] = \int_{-\infty}^{\infty} \varphi(t) f_2(t) dt$$

を示せばよい。

$$\begin{aligned} E \left[\varphi \left(\frac{\sqrt{2}X}{Y} \right) \right] &= \iint_{\{-\infty < x < \infty, y > 0\}} \varphi(\sqrt{2}x/y) \frac{e^{-\frac{x^2}{2}}}{\sqrt{2\pi}} \frac{e^{-\frac{y}{2}}}{2} dx dy \\ &= \iint_{\{-\infty < t < \infty, y > 0\}} \varphi(t) \frac{e^{-\frac{y(1+t^2)}{2}}}{4\sqrt{\pi}} \sqrt{y} dt dy \quad (x = t\sqrt{\frac{y}{2}} \text{ と変換}). \\ &= \int_{-\infty}^{\infty} \varphi(t) \frac{1}{4\sqrt{\pi}} \left(\int_0^{\infty} e^{-\frac{y(1+t^2)}{2}} \sqrt{y} dy \right) dt, \end{aligned}$$

$$\int_0^{\infty} e^{-\frac{y(1+t^2)}{2}} \sqrt{y} dy = \int_0^{\infty} \exp \left(-\frac{r^2(1+\frac{t^2}{2})}{2} \right) 2r^2 dr = \frac{2}{(1+\frac{t^2}{2})^{3/2}} \int_0^{\infty} e^{-\frac{r^2}{2}} r^2 dr = \frac{\sqrt{2\pi}}{(1+\frac{t^2}{2})^{3/2}},$$

だから示された。 □

$\tilde{T}_n$ の分布との関係は次のようになる.

定理 70. $X_1, \dots, X_n$ を $N(m, \sigma^2)$ に従う独立確率変数とする. 標本平均 $\bar{X}_n$, 不偏分散 u_n^2 , 標本分散 s_n^2 により定まる統計量

$$\hat{T}_n = \frac{\sqrt{n}(\bar{X}_n - m)}{u_n} = \frac{\sqrt{n-1}(\bar{X}_n - m)}{s_n}$$

は自由度 $n-1$ の t -分布に従う.

証明.

$$\hat{T}_n = \frac{\sqrt{n} \frac{\bar{X}_n - m}{\sigma}}{\sqrt{\frac{u_n^2}{\sigma^2}}}$$

と書き直せる.

- (i) $\sqrt{n} \frac{\bar{X}_n - m}{\sigma}$ は標準正規分布に従う (定理 55)
- (ii) $\frac{(n-1)u_n^2}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従う (定理 67)

はすでに示した. したがって $\sqrt{n} \frac{\bar{X}_n - m}{\sigma}$ と $\frac{(n-1)u_n^2}{\sigma^2}$ が独立であることを示せばよい. これは

$\text{Cov}(\bar{X}_n, X_i - \bar{X}_n) = 0$ なので定理 56 により $\bar{X}_n$ と $X_i - \bar{X}_n$ が独立であることからわかる.

□

注意 71. (1) 定理 56 を用いて $\bar{X}_n, X_i - \bar{X}_n$ の独立性を示そう. $E[\bar{X}_n] = m$, $E[X_i - \bar{X}_n] = 0$ なので

$$\begin{aligned} \text{Cov}(\bar{X}_n, X_i - \bar{X}_n) &= E[(\bar{X}_n - m)(X_i - \bar{X}_n)] \\ &= E \left[\left\{ \frac{1}{n} \sum_{j=1}^n (X_j - m) \right\} \left\{ (X_i - m) - \frac{1}{n} \sum_{k=1}^n (X_k - m) \right\} \right] \\ &= \frac{1}{n} E \left[\sum_{j=1}^n (X_j - m)(X_i - m) \right] - \frac{1}{n^2} E \left[\sum_{1 \leq j \leq n, 1 \leq k \leq n} (X_j - m)(X_k - m) \right] \\ &= \frac{1}{n} \sigma^2 - \frac{1}{n^2} \cdot n \sigma^2 = 0. \end{aligned}$$

したがって独立である. しかし、本当はこれだけでは不十分である. というのは

定理 $Y_1, \dots, Y_n$ を独立な確率変数とする. この確率変数を二つに分けて関数と合成した確率変数 $Z_1 = f(Y_1, \dots, Y_p)$, $Z_2 = g(Y_{p+1}, \dots, Y_n)$ を考える. Z_1 と Z_2 は独立である

は正しい. しかし

(*) Y_1 と Z , Y_2 と Z , ..., Y_n と Z は独立とする. このとき $\sum_{i=1}^n Y_i$ と Z は独立である.

ということは一般的には成り立たない. 事象の独立性でも同じようなことがあったことを思い出してほしい. $(Y_1, \dots, Y_n, Z)$ の同時分布が正規分布であるときは, (*) は正しい命題である. 従って $\bar{X}_n$ と $\sum_{i=1}^n (X_i - \bar{X}_n)^2$ が独立と言ってよいのである.

(2) 大数の法則により標本分散 s_n^2 は $n \rightarrow \infty$ で母分散 σ^2 に近づいていく. したがって $u_n^2 = \frac{n}{n-1} s_n^2$ も σ^2 に近づいていく. よって $\hat{T}_n$ の密度関数 $f_{n-1}(x)$ は標準正規分布の密度関数 $\frac{e^{-x^2/2}}{\sqrt{2\pi}}$ に近づくと予想される. この予想は正しい.

$$\lim_{n \rightarrow \infty} \sqrt{n} \int_0^1 x^{-1/2} (1-x)^{\frac{n}{2}-1} dx = \sqrt{2\pi} \quad (\clubsuit)$$

$$\lim_{n \rightarrow \infty} \left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} = e^{-\frac{x^2}{2}}. \quad (\spadesuit)$$

から示される. $(\spadesuit)$ は e^x の定義からわかる. $(\clubsuit)$ は $x = t^2/n$ と変換してみれば証明できる. 詳細は考えて見て下さい.

7. 区間推定

以下の推定について説明する。

- (1) 正規母集団の母平均の推定 (母分散が既知の場合)
- (2) 正規母集団の母平均の推定 (母分散が未知の場合)
- (3) 母比率の推定

- (1) 正規母集団の母平均の推定 (母分散が既知の場合)

次の教科書に載っている例題を考えよう。

例題 1 あるメーカーの自動車のガソリン 1 リットルあたりの走行距離は標準偏差 $\sigma = 0.50\text{km}$ の正規分布にしたがうという。10 台をランダムに選んで 1 リットルあたりの走行距離を調べたところ、次の結果を得た：

17.5, 18.0, 18.3, 17.7, 18.5, 18.0, 18.6, 17.2, 18.7, 18.2 (単位 km)

母平均に対する**信頼度** 95 %、99 %の**信頼区間**を求めよ。

解 標本調査の結果標本平均 $\bar{x}_{10}$ は

$$\bar{x}_{10} = \frac{17.5 + 18.0 + 18.3 + 17.7 + 18.5 + 18.0 + 18.6 + 17.2 + 18.7 + 18.2}{10} = 18.07 \approx 18.1$$

である。自動車の 1 リットルあたりの走行距離の母平均を m とする。 X_n を平均 m , 標準偏差 0.50 の正規分布に従う独立確率変数とし $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ と定める。

$$\frac{\bar{X}_{10} - m}{0.5} \sqrt{10}$$

は標準正規分布に従う。標準正規分布にしたがう確率変数 T について

$$P(|T| \leq 1.96) = 0.95, \quad P(|T| \leq 2.58) = 0.99$$

だから

$$P\left(\left|\frac{\bar{X}_{10} - m}{0.5} \sqrt{10}\right| \leq 1.96\right) = 0.95.$$

$\left|\frac{\bar{X}_{10} - m}{0.5} \sqrt{10}\right| \leq 1.96$ は

$$\bar{X}_{10} - \frac{0.5}{\sqrt{10}} \times 1.96 \leq m \leq \bar{X}_{10} + \frac{0.5}{\sqrt{10}} \times 1.96$$

と同値である。 $\bar{X}_{10}$ を標本調査の結果の標本平均 $\bar{x}_{10}$ に置き換えた区間

$$\left[\bar{x}_{10} - \frac{0.5}{\sqrt{10}} \times 1.96, \quad \bar{x}_{10} + \frac{0.5}{\sqrt{10}} \times 1.96\right]$$

が信頼度 95 %の信頼区間である。具体的には

$$\left[18.1 - \frac{0.5}{\sqrt{10}} \times 1.96, 18.1 + \frac{0.5}{\sqrt{10}} \times 1.96\right] \approx [17.8, 18.4].$$

□

注意 72. (1) 信頼度 99 % の信頼区間は

$$\left[18.1 - \frac{0.5}{\sqrt{10}} \times 2.58, 18.1 + \frac{0.5}{\sqrt{10}} \times 2.58 \right]$$

となる. 信頼度をあげると区間の幅が広くなることに注意. また, 信頼区間は標本調査の結果に依存することにも注意すること.

(2) T を標準正規分布に従う確率変数とする. $0 < \alpha \leq 1$ に対して

$$P(|T| \geq z(\alpha)) = \alpha$$

となる $z(\alpha)$ を標準正規分布の両側 100α % 点という. 例えば

$$\text{両側 5 \% 点 } z(0.05) = 1.96, \quad \text{両側 1 \% 点 } z(0.01) = 2.58$$

である.

一般に母分布が $N(m, \sigma^2)$ の正規母集団から無作為抽出した標本 $X_1, \dots, X_n$ の正規化

$$T_n = \frac{\bar{X}_n - m}{\sigma} \sqrt{n}$$

は $N(0, 1)$ に従うから

$$P\left(-z(\alpha) \leq \frac{\bar{X}_n - m}{\sigma} \sqrt{n} \leq z(\alpha)\right) = 1 - \alpha.$$

従って

$$\bar{X}_n - \frac{\sigma}{\sqrt{n}} z(\alpha) \leq m \leq \bar{X}_n + \frac{\sigma}{\sqrt{n}} z(\alpha)$$

となる確率が $1 - \alpha$.

n 個の標本調査を行った結果, 標本平均 $\bar{x}_n$ を得た時, m が区間

$$\left[\bar{x}_n - \frac{\sigma}{\sqrt{n}} z(\alpha), \bar{x}_n + \frac{\sigma}{\sqrt{n}} z(\alpha) \right] \quad (\star 1)$$

に入るのは α が小さいなら非常に確からしいであろう. したがって標本を用いて定まる上記の $(\star 1)$ を信頼度 $100(1 - \alpha)$ % の信頼区間というのである.

(2) 正規母集団の母平均の推定 (母分散が未知の場合)

次の問題を考えよう.

例題 2 工場に新しい機械を入れてボルトを作った. 製品の中から 20 個を選んでその長さを測ったところ平均が 2.52cm, 標本標準偏差が 0.11 であったという. ボルトの長さは正規分布に従うとして, 母平均の信頼度 95 %, 99 % の信頼区間を求めよ.

解 この問題では新しい機械を使っているため, 過去に蓄積された情報が無いため, 母分散は未知である. ボルトの長さが正規分布 $N(m, \sigma^2)$ に従うとする. Gosset の定理 (講義資料の定理 70) によれば $X_1, \dots, X_n$ を $N(m, \sigma^2)$ に従う独立確率変数とすると

$$\hat{T}_n = \frac{\bar{X}_n - m}{s_n} \times \sqrt{n-1}$$

は自由度 $n - 1$ の t 分布に従う. ただし $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ は標本平均, $s_n = \sqrt{\frac{1}{n} \sum_{i=1}^n (X_i - \bar{X}_n)^2}$ は標本標準偏差である. 本問題では $n = 20$ の場合にあたる. 自由度 19 の t 分布について教科書巻末の t 分布表 (141 ページ) によれば

$$P(|t| \leq 2.09) = 0.95, \quad P(|t| \leq 2.86) = 0.99.$$

すなわち

$$P\left(\left|\frac{\bar{X}_{20}-m}{s_{20}}\sqrt{19}\right|\leq 2.09\right)=0.95.$$

したがって

$$\bar{X}_{20}-\frac{s_{20}}{\sqrt{19}}\times 2.09\leq m\leq \bar{X}_{20}+\frac{s_{20}}{\sqrt{19}}\times 2.09$$

となる確率は 0.95 である. ここで、 $\bar{X}_{20}, s_{20}$ を標本調査の結果の $\bar{x}_{20}=2.52, \bar{s}_{20}=0.11$ に変えて得られる区間

$$\left[2.52-\frac{0.11}{\sqrt{19}}\times 2.09, 2.52+\frac{0.11}{\sqrt{19}}\times 2.09\right]=: [2.45, 2.57].$$

が信頼度 95 % の信頼区間である.

□

注意 73. (1) 分散 σ^2 がわからなくても、 n が十分大きければ中心極限定理と大数の法則により、確率変数

$$\frac{\bar{X}_n-m}{s_n}\sqrt{n-1}$$

の分布は標準正規分布で近似できるので、 t 分布を使う必要は無い. しかし、 $n=20$ は大きくないため、 t 分布を使う必要があるのである.

(2) t を自由度 n の t 分布にしたがう確率変数とする. $0<\alpha<1$ に対して

$$P(|t|\geq t_n(\alpha))=\alpha$$

となる $t_n(\alpha)$ を自由度 n の t 分布の両側 100α % 点という. 例えば $n=19$ (本問題の場合) のとき

$$\text{両側 } 5\% \text{ 点 } t_{19}(0.05)=2.09, \quad \text{両側 } 1\% \text{ 点 } t_{19}(0.01)=2.86$$

である. 一般に母分布が $N(m, \sigma^2)$ の正規母集団から無作為抽出した標本 $X_1, \dots, X_n$ について

$$\hat{T}_n=\frac{\bar{X}_n-m}{s_n}\sqrt{n-1}$$

は自由度 $n-1$ の t 分布に従うから

$$P\left(-t_{n-1}(\alpha)\leq \frac{\bar{X}_n-m}{s_n}\sqrt{n-1}\leq t_{n-1}(\alpha)\right)=1-\alpha.$$

従って

$$\bar{X}_n-\frac{s_n}{\sqrt{n-1}}t_{n-1}(\alpha)\leq m\leq \bar{X}_n+\frac{s_n}{\sqrt{n-1}}t_{n-1}(\alpha)$$

となる確率が $1-\alpha$.

したがって α が小さい時、 n 個の標本調査を行った結果、標本平均 $\bar{x}_n$ 、標本標準偏差 $\bar{s}_n$ を得た時、 m が区間

$$\left[\bar{x}_n-\frac{\bar{s}_n}{\sqrt{n-1}}\times t_{n-1}(\alpha), \bar{x}_n+\frac{\bar{s}_n}{\sqrt{n-1}}\times t_{n-1}(\alpha)\right] \quad (\star 2)$$

に入るのは非常に確からしいであろう. したがって標本を用いて定まる上記の $(\star 2)$ を信頼度 $100(1-\alpha)$ % の信頼区間というのである.

(3) 母比率の推定

母集団の各要素がある特性 A に属するか属さないという状況で A に属するという比率 (割合) p を区間推定する問題を考える. この比率 p を母比率という. 例えば

(i) ある地方の各家庭であるテレビ番組を見た比率 (視聴率調査)

(ii) 日本の有権者で民主党を支持する人の比率

などの推定である.

例題 3 ある地方であるテレビ番組を視聴したかどうか 200 人の人に調査したところ、視聴した割合は 23.5 %であった. この地方における視聴率の信頼度 95 %, 99 %の信頼区間を求めよ.

解 視聴率を p とする.

$$P(X_i = 1) = p, \quad P(X_i = 0) = 1 - p$$

となる独立確率変数 $X_1, \dots, X_n$ の和 $S_n = X_1 + \dots + X_n$ が無作為に n 人の視聴者を選んだときの視聴していた人の人数を表す確率変数である. $V[X_i] = p(1-p)$ だから n が大きい時、中心極限定理により

$$T_n = \frac{\bar{X}_n - p}{\sqrt{p(1-p)}} \sqrt{n}$$

は $N(0, 1)$ で近似できる. ($\bar{X}_n = \frac{S_n}{n}$ である). $n = 200$ の場合、標準正規分布の両側 5 %点 1.96 を用いると

$$P\left(\left|\frac{\bar{X}_{200} - p}{\sqrt{p(1-p)}} \times \sqrt{200}\right| \leq 1.96\right) = 0.95.$$

従って

$$\bar{X}_{200} - \frac{1.96\sqrt{p(1-p)}}{\sqrt{200}} \leq p \leq \bar{X}_{200} + \frac{1.96\sqrt{p(1-p)}}{\sqrt{200}}$$

となる確率は 0.95 である. ここで $\bar{X}_{200}$ を標本調査の結果の 0.235 に変えて信頼度 95 %の信頼区間を求めろ。。。となりそうだが、この区間の表示には推定したい未知の p が含まれてしまっている. ここで次の二つの立場で考えよう.

(i) 信頼区間を大きく取る立場

p は確かに不明なので、最悪の場合を考える.

$$p(1-p) = p - p^2 = -\left(p - \frac{1}{2}\right)^2 + \frac{1}{4} \leq \frac{1}{4}$$

だから $p(1-p)$ を $\frac{1}{4}$ に取っておけば十分と考えられる. そこで

$$\left[\bar{x}_{200} - \frac{1.96\sqrt{\frac{1}{4}}}{\sqrt{200}}, \bar{x}_{200} + \frac{1.96\sqrt{\frac{1}{4}}}{\sqrt{200}} \right]$$

を信頼度 95 %の信頼区間とする.

(ii) p を近似する値として標本調査で得られた値を用いる立場

視聴率調査の結果の $\bar{x}_{200} = 0.235$ で置き換えて

$$\left[0.235 - \frac{1.96 \times \sqrt{0.235 \times (1-0.235)}}{\sqrt{200}}, 0.235 + \frac{1.96 \times \sqrt{0.235 \times (1-0.235)}}{\sqrt{200}} \right]$$

を信頼度 95 %の信頼区間とする.

□

さらに例題をあげる。

例題 4 全国模試の数学のテストで無作為に 400 人の学生の点数を調べたところ、平均点が 65 点であった。得点の分布の標準偏差が 10 であることが知られている場合、数学の試験の点の母平均を信頼度 95 % で区間推定せよ。

ヒント：試験の得点が正規分布に従うかこの問題では明らかではないが 400 人という標本は十分大きいとしてよいので、正規分布で近似することを考える。無作為抽出して得られる n 個の得点の確率変数を $X_1, \dots, X_n$ とする。本問題では $n = 400$ だが十分大きいので、 X_i の分布が正規分布かどうか明らかではないが正規化した

$$\frac{\bar{X}_{400} - m}{10} \times \sqrt{400}$$

は標準正規分布で近似できると考えられる。

例題 5 (1) 全国模試の数学のテストで無作為に 30 人の学生の点数を調べたところ、平均点が 65 点、標本標準偏差が 10 であった。得点の分布は正規分布に従うとし、数学の試験の点の母平均 m を信頼度 95 % で区間推定せよ。

(2) 全国模試の数学のテストで無作為に 300 人の学生の点数を調べたところ、平均点が 65 点、標本標準偏差が 10 であった。数学の試験の点の母平均 m を信頼度 95 % で区間推定せよ。

ヒント：無作為抽出して得られる n 個の得点の確率変数を $X_1, \dots, X_n$ とする。標本分散を s_n^2 とする。

(1) の場合は $n = 30$ が大きくないので

$$\frac{\bar{X}_{30} - m}{s_{30}} \times \sqrt{29}$$

は自由度 29 の t 分布に従うとして計算する。(2) の場合は $n = 300$ は十分大きいので

$$\frac{\bar{X}_{300} - m}{s_{300}} \times \sqrt{300}$$

は $N(0, 1)$ に従う確率変数であるとして計算する。

例題 6 ある選挙区で一人の候補者の支持率を信頼度 95 % の信頼区間の幅が 2 % 以下であるように推定するにはどの位の大きさの標本を抽出すればよいか？

ヒント：標本の数を n とし、標本調査の結果の支持率の割合を $\bar{x}_n$ とする。 $p(1-p) \leq \frac{1}{4}$ を用いて信頼区間を大きく取る立場をとると信頼度 95 % の信頼区間は

$$\left[\bar{x}_n - \frac{\frac{1}{2}}{\sqrt{n}} \times 1.96, \bar{x}_n + \frac{\frac{1}{2}}{\sqrt{n}} \times 1.96 \right]$$

したがって信頼区間の幅は $\frac{1.96}{\sqrt{n}}$ 。これが 0.02 より小さければよい。

8. 仮説検定

8.1 仮説検定の考えかた

仮説検定とは、ある母集団の母数に対する仮説を標本調査の結果を見て、認める (採択する) か認めないか (棄却する) を判断することを言う。

例 あるテレビ番組のある週の視聴率は 30 % であるという。次の週に 300 軒の家庭で調査したところ、99 軒の家庭で視聴されていたという。視聴率は変化したと言えるだろうか。

考え方 視聴率は 30 % のままであるという仮説 H (Hypothesis の頭文字の H) を考える。この仮説が正しいければ標本調査の結果は母平均の $300 \times 30 = 90$ 軒からあまり離れていないはずである。

- (i) 平均の 90 から 9 以上離れたデータが得られる確率が小さければ、まれな現象が起こったわけだから、仮説は間違いと推測し棄却する。
- (ii) 平均の 90 から 9 以上離れたデータが得られる確率が小さくなければ、仮説に反するデータが得られたわけではないので、仮説を棄却することをしない。(仮説 H を採択する、ただし積極的に採択するというわけではない)。

の方針で考える。仮説 H の下、300 軒のうち視聴した軒数 X は $B(300, 0.3)$ の二項分布に従う。300 個のデータは十分数が大きいので、中心極限定理により X は正規分布 $N(90, 300 \times 0.3 \times 0.7)$ に従う確率変数 $\tilde{X}$ で近似できる。半整数の補正をして

$$\begin{aligned} P(|X - 90| \geq 9) &\doteq P(|\tilde{X} - 90| \geq 8.5) \\ &= P\left(\left|\frac{\tilde{X} - 90}{\sqrt{300 \times 0.3 \times 0.7}}\right| \geq \frac{8.5}{\sqrt{300 \times 0.3 \times 0.7}}\right) \\ &= P(|T| \geq 1.07) \quad (T \text{ は標準正規分布にしたがう確率変数}) \\ &= 1 - 0.36 \times 2 = 0.28 \end{aligned}$$

確率 0.28 は小さいとは言えないので、仮説を棄却することは無い、ということになる。□

注意 74. 仮説は棄却されなかったが、この仮説が間違っている可能性はもちろんある。その場合、棄却しないという判断は誤りということになる。この間違いを**第二種の過誤**と言う。

上記の方針 (i), (ii) で確率が小さければと言ったがどのくらいの確率まで考えるかで棄却、採択が変わる。この確率を危険率、有意水準という。

仮説検定の流れを述べると次のようになる：

- (1) 帰無仮説 H を設定する。
- (2) 帰無仮説 H の確率分布に従う独立確率変数 $X_1, \dots, X_n$ の統計量 $T(X_1, \dots, X_n)$ の分布を決定する。
- (3) 有意水準 (危険率) α を設定し、

$$P(T(X_1, \dots, X_n) \in W_\alpha) = \alpha$$

となる危険率 α の棄却域 W_α を求める。

- (4) 実際に標本調査を行いその標本 $x_1, \dots, x_n$ を統計量に代入した値 $T(x_1, \dots, x_n)$ が W_α に属していれば仮説を棄却し、 W_α に属さなければ仮説を採択することにする。

なお、教科書では 300 軒の調査で危険率 5 パーセントの標本数に対する棄却域は

$$W_{0.05} = \{n \mid 0 \leq n \leq 74 \text{ または } n \geq 106\}$$

と計算されている。この問題では視聴率が増えたかどうかを検定したので、90 軒より極端に多いか、少ないかの両方の範囲が棄却域になる。つまり**両側検定**になっている。

また、もう少し正確に言うと帰無仮説に対してもう一つの仮説

対立仮説 H' : 視聴率 $\neq 30\%$

を立て、どちらを取るかを判断したことになる。

宣伝活動を行って視聴率が上がったかどうかを判定したいときもあるであろう。その場合は以下のように片側検定を行うことになる。

例題 1 あるテレビ番組のある週の視聴率は 30 % であった。さらに宣伝活動を行ったところ、次の週は、1000 軒で調査したところ、330 軒で視聴されていたと言う。視聴率は上がったと言えるか？危険率 5 % で検定せよ。

解 帰無仮説 H 、および対立仮説 H' は

H : 視聴率は 30 % である、

H' : 視聴率は 30 % より大きい

である。 X_i を $P(X_i = 1) = 0.3, P(X_i = 0) = 0.7$ となる独立確率変数とする。 $S = \sum_{i=1}^{1000} X_i$ と定める。 S は二項分布 $B(1000, 0.3)$ に従う確率変数である。 S の平均は 300 なので、 S が 300 よりどのくらい大きければ確率が 0.05 になるかを見ればよい。つまり

$$P(S \geq 300 + a) = 0.05$$

となる a を求めることになる。 S の分布は正規分布 $N(300, 1000 \times 0.3 \times 0.7)$ で近似できるので、正規分布 $N(300, 210)$ に従う確率変数 $\tilde{S}$ を考えると

$$\begin{aligned} P(S \geq 300 + a) &\approx P(\tilde{S} \geq 300 + a - 0.5) && \text{(半整数の補正)} \\ &= P\left(\frac{\tilde{S} - 300}{\sqrt{210}} \geq \frac{a - 0.5}{\sqrt{210}}\right) \\ &= P\left(T \geq \frac{a - 0.5}{\sqrt{210}}\right) && (\star 1) \end{aligned}$$

ここで T は $N(0, 1)$ に従う確率変数である。 $(\star 1)$ の等号は正規分布の性質から従うものである。正規分布表より

$$P(T \geq 1.65) = 0.05$$

だから

$$\frac{a - 0.5}{\sqrt{210}} = 1.65$$

を解いて $a = 24.4$. 従って、標本平均に対する棄却域は $[324.4, \infty)$. 半整数の補正をしないと $a = 23.9$ で棄却域は $[323.9, \infty)$ である.

いずれの場合でも 330 という数字は棄却域に入っているので、帰無仮説は棄却され対立仮説 H' が採択されることになる. $\square$

注意 75. (1) 危険率を 1 % とすると $P(T \geq 2.33) = 0.01$ なので $\frac{a-0.5}{\sqrt{210}} \geq 2.33$ を解いて $a \geq 34.3$. ゆえに棄却域は $[334.3, \infty)$. したがって仮説は採択され視聴率はあがったとは言えない、となる。

(2) 危険率 5 % の検定では、帰無仮説を棄却したが、この判断が誤りである可能性はある。これを**第一種の過誤**と言う。

仮説検定の例題をさらに二つあげる。例題 2 と例題 4 は教科書に載っている例題です。

例題 2(母平均の検定)

あるメーカーが平均 1500 時間、標準偏差 30 の寿命をもつ蛍光灯を改良しようとした。試作品の中から 20 本選んで標本調査したところ、標本平均は 1517 時間の寿命であったという。分布は正規分布であり、標準偏差は変わらないものとし、危険率 1 パーセントで改良されたかどうか検定せよ。

解

H : 平均は 1500 である。

H' : 平均は 1500 より大きい。

の片側検定の問題である。 $N(1500, 30)$ に従う i.i.d. X_i ($i = 1, 2, \dots$) を考える。標本平均 $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ 対して統計量

$$T = \frac{\bar{X}_n - 1500}{30} \times \sqrt{n}$$

は標準正規分布に従う。正規分布表より

$$P(T \geq 2.33) = 0.01.$$

$$\frac{\bar{X}_{20} - 1500}{30} \times \sqrt{20} \geq 2.33$$

を $\bar{X}_{20}$ について解いて $\bar{X}_{20} \geq 1515.6$. 危険率 1 % の棄却域は 1515.6 時間以上となる。したがって仮説 H は棄却され、改良されたと判断する. $\square$

注意 76. 改良後も標準偏差 30 の正規分布に従っているとすると信頼度 99 パーセントの信頼区間は $P(|T| \geq 2.58) = 0.01$ を用いて

$$\left[1517 - \frac{30}{\sqrt{20}} \times 2.58, 1517 + \frac{30}{\sqrt{20}} \times 2.58 \right]$$

と求まる。

例題 3 総点が 1000 点である全国模試の結果、全国平均は 595 点、標準偏差 50 点であったと言う。A 高校の受験者のうち、30 人を選んで平均を計算したところ 610 点であったと言う。A 高校の受験者の成績は全国平均より高いと考えられるか？得点の分布は正規分布に従うとし、有意水準 5 パーセントで検定せよ。

ヒント：

帰無仮説 H: A 高校の模試の得点分布は $N(595, 50)$ に従う。

対立仮説 H': A 高校の模試の得点分布の平均は 595 点より高い。

の検定を行うことになる。帰無仮説 H の下、A 高校の受験者の試験結果から無作為抽出して得られる標本 $X_1, \dots, X_n$ に対して標本平均 $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ を考えると

$$\frac{\bar{X}_n - 595}{50} \times \sqrt{n}$$

は標準正規分布に従う。一方 $N(0, 1)$ に従う確率変数 T に対して

$$P(T \geq 1.65) = 0.05$$

である。従って標本調査の結果の値 $\bar{x}_{20}$ が

$$\frac{\bar{x}_{20} - 595}{50} \times \sqrt{30} \geq 1.65$$

を満たしていれば棄却される。

例題 4 (t 分布を使った母平均の検定) ある食品の包装には内容量 100 グラムと印してある。この食品 20 個について調べたところ平均が 98.5 グラム、標本標準偏差 3 グラムであったと言う。表示に誤りがあると言えるか? 危険率 5 パーセントで検定せよ。

解 この食品の内容量は正規分布 $N(m, \sigma^2)$ に従うと考える。ただし標準偏差 σ は未知である。

帰無仮説は

H: $m = 100$

である。20 個の標本 $X_1, \dots, X_{20}$ を抽出したとすると、標本平均、標本標準偏差

$$\bar{X}_{20} = \frac{X_1 + \dots + X_{20}}{20}, \quad s_{20} = \sqrt{\frac{1}{20} \sum_{i=1}^{20} (X_i - \bar{X}_{20})^2}$$

で定まる統計量

$$\frac{\bar{X}_{20} - 100}{s_{20}} \times \sqrt{19}$$

は自由度 19 の t 分布に従う。 t 分布表によると自由度 19 の t 分布に従う確率変数 t について

$$P(|t| \geq 2.09) = 0.05.$$

従って

$$P\left(\left|\frac{\bar{X}_{20} - 100}{s_{20}} \times \sqrt{19}\right| \geq 2.09\right) = 0.05.$$

統計量 $T(X_1, \dots, X_{20}) = \frac{\bar{X}_{20} - 100}{s_{20}} \times \sqrt{19}$ に標本調査の結果を代入すると

$$\frac{98.5 - 100}{3} \times \sqrt{19} = -2.18.$$

従って棄却域に入っているので棄却され、表示に誤りがあると判断される。

□