1 確率

1.1 素朴な意味での確率について

まず素朴な確率の概念を復習しよう．この素朴な意味での確率の定義は Laplace(=ラプラス) という人によります．高校数学の数学 A で出てくる確率はこの考え方に基づいています．

定義 1.1. (i) あるランダムな現象があり，　その結果起こり得る結果が 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．この起こり得る個々の結果を ω_i $(1 \leq i \leq n)$ と表して根元事象と言う．すべての起こり得る場合全体の集合 $\{\omega_1, \ldots, \omega_n\}$ を Ω と書き, 標本空間という．

(ii) Ω の部分集合を事象と言う．Ω 自身は全事象と言う．ある事象 A の要素の数 (場合の数) が r 通りであるとき, 事象 A の起こる確率 $P(A)$ を $P(A) = \dfrac{r}{n}$ とする．

例 1.2. (1) どの目の出るのも同様に確からしいサイコロを 1 回投げる．このような操作を**試行**と言う．この試行の結果は出た目を記録して $\Omega = \{1, 2, 3, 4, 5, 6\}$ となる．各 i $(1 \leq i \leq 6)$ が根元事象である．偶数が出るという事象 $A = \{2, 4, 6\}$ で確率 $P(A)$ は

$$P(A) = \frac{3}{6} = \frac{1}{2}$$

などとなる．

(2) どの目の出るのも同様に確からしいサイコロを独立に 2 回続けて投げる．このように互いに影響を受けることなく行う試行を**独立試行**と言う．起こり得る結果全体は (1 回目に出た目が i, 2 回目に出た目が j のとき (i, j) と書くと)

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

のようになり, この 36 通りの各要素が根元事象．少なくとも 1 回は 1 の目が出ると言う事象 A を考える．

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

である．$\sharp A = 36 - 5^2 = 11$ なので ($\sharp A$ 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 $P(A) = 11/36$.

また, 確率は次の加法性を持つことに注意しよう．

確率の加法性

(i) 事象 A, 事象 B に対して和集合 $A \cup B$ を A と B の和事象と言う．$A \cap B$ を A と B の積事象と言う．

(ii) 事象 A, B に対して, $P(A \cup B) = P(A) + P(B) - P(A \cap B)$. 特に事象 A, B が互いに排反すなわち, $A \cap B \neq \emptyset$ ならば $P(A \cup B) = P(A) + P(B)$.

この性質は長さ、面積や体積の次の性質 (あるいは有限個の要素からなる集合の要素の数と言っても良い) と同様な性質であり, 確率は面積や体積の親戚と言ってよいものです．

面積の加法性

(i) 平面内の集合 A に対して，　その面積を $|A|$ で表すことにする．

(ii) 平面内の集合 A, B に対して $|A \cup B| = |A| + |B| - |A \cap B|$. 特に $A \cap B = \emptyset$ を満たせば $|A \cup B| = P(A) + P(B)$.

ラプラス流の確率の定義を述べましたが, この流儀でとらえられるランダムな現象には限りがあります．そこで，　上記の確率の加法性に着目してもっと抽象的に確率空間を定義する必要があります。これが Kolmogorov(=コルモゴロフ) により定義された確率空間です．

1.2 現代的な確率の定義

定義 1.3. 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) $A_1, A_2, \ldots, A_i, \ldots$ 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 A の実数 $P(A)$ が定まり, 次をみたす.
(a) $0 \le P(A) \le 1$
(b) $P(\Omega) = 1$
(c) [可算加法性] $A_1, A_2, \ldots, A_i, \ldots$ が $\mathcal{F}$ の要素で $i \neq j$ のとき $A_i \cap A_j = \emptyset$(排反) ならば

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

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. また,$\mathcal{F}$ の元 (Ω の部分集合) は事象, 事象 $A \in \mathcal{F}$ に対して $P(A)$ は A の確率と言います.

例 1.4. (1) 1 回のサイコロ投げの確率空間

$$\begin{aligned} \Omega &= \{1, 2, 3, 4, 5, 6\}, \\ \mathcal{F} &= \Omega \text{ の部分集合全体} \\ P(A) &= \frac{A \text{ の要素の個数}}{6} \end{aligned}$$

(2) 独立に 2 回続けてサイコロ投げを行う場合の確率空間

$$\begin{aligned} \Omega &= \{(i, j) \mid 1 \le i \le 6, 1 \le j \le 6\} \\ \mathcal{F} &= \Omega \text{ の部分集合全体}, \\ P(A) &= \frac{A \text{ の要素の個数}}{36} \end{aligned}$$

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 $\Omega = \{\omega_1, \ldots, \omega_n\}$, $\mathcal{F}$ を Ω の部分集合全体, 各根元事象 ω_i の確率を p_i とし、$P(A) = \sum_{\{i \mid \omega_i \in A\}} p_i$ と定義すると $(\Omega, \mathcal{F}, P)$ は確率空間である. ここで $p_i = 1/n$ ならば根元事象がどれも同様に確からしく起こる場合に相当するが, 必ずしも p_i が全部同じである必要は無い. 例えばいかさまサイコロならばある目だけが特に出やすい場合もあるかもしれない.
(3) 無限回のサイコロ投げ

有限回だけサイコロを振る場合や根元事象の数が有限個のとき, (1), (2) で見たようにラプラス流の確率で間に合う (根元事象の確率がすべて等しい場合も考えるというふうに一般化していますが). 何回も独立にサイコロ投げを続けることを考える. その試行の結果として、1〜6 の数字の無限列が現れる. この無限列一

つ一つが根元事象とみなせる．すなわち Ω は $\Omega=\{(a_1,a_2,\ldots,a_n,\ldots)\mid a_i=1,\ldots,6\}$. $\mathcal{F}$ 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．説明すると長くなるので、省略するがこのような無限回の試行を考えるとラプラス流の確率の定義では収まらず、Kolmogorov 流の確率空間の定義を採用しなければならないのである．
(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 $P(A)=\dfrac{A\text{ の長さ}}{b-a}$ 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ーグ可測集合と言います.) この例を見れば, 確率と長さの概念の親和性がわかります．

上記の確率空間のいずれもなんらかのランダムな現象や試行があり、その結果得られる数値一つ一つが根元事象を、数値全体が標本空間になっていることを注意しておきます．このランダムな数値が確率変数,
ランダムな数値がどのように分布しているかを表すのが確率分布になります．確率変数の数学的定義は後にして, まず、確率分布の定義を述べよう．

2　確率分布

定義 2.1. (1) 標本空間 Ω が $\mathbb{R},[a,b],\{x_1,x_2,\ldots,x_n\}$ などのように $\mathbb{R}$ の部分集合になっている場合の確率空間を考える．このときの確率 P を**確率分布**, **確率法則** (あるいは短く**分布**, **法則**) という．すなわち，　$\mathbb{R}$ の部分集合 A に対してその確率 $P(A)$ が定まっているものを言う．
(2) $\mathbb{R}$ や $[a,b]$ のように連続的に変わり得る値を取るときの確率分布を**連続型確率分布**という．
(3) 全事象が飛び飛びの値 $\{a_1,a_2,\ldots\}$ (無限個あってもよい) からなる確率分布を**離散型確率分布**という．
(4) 確率分布 P に対して

$$F(z)=P\left((-\infty,z]\right)(=(-\infty,z]\text{ という集合の確率})$$

と定めると $F(z)$ は $\mathbb{R}$ 上の関数になる．これを確率分布 P の**分布関数**と言う．

確率分布の情報はすべてこの分布関数にあると考えられるので、確率分布と分布関数は等価な概念と言えます．離散型確率分布, 連続型確率分布の例をあげよう．後でまた大事な例をあげることにします．

例 2.2 (離散型確率分布の例)**.** (1) $P(\{i\})=\frac{1}{6}$ $(1\le i\le 6)$ で定まる $\mathbb{R}$ 上の確率は離散型の確率分布を定める．もっと定義に即して述べれば $A\subset\mathbb{R}$ に対して

$$P(A)=A\text{ に含まれる数 }1,\ldots,6\text{ の確率を全て足したもの}$$

で定義される確率である．この確率分布はどの目も同様に確からしく出るサイコロを 1 回投げたときの出目の確率分布である．
(2) ポアソン分布 $\lambda>0$ とする．

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

で定まる離散型確率分布をパラメータ λ のポアソン分布という．($\sum_{k=0}^{\infty}e^{-\lambda}\frac{\lambda^k}{k!}=1$ に注意しよう)．ポアソン (=Poisson) は人の名前である．ポアソン分布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偶然現象の回数の従う分布である．例えば,

(i) ある軍隊で馬に蹴られて死亡した 1 年間の兵士の数

(ii) 日本全国で行われる宝くじが多数行われているとする。このとき、仙台市で 1 年間で宝くじで一等が出た件数

(iii) ある店に 1 日で来店する客の数

(iv) ある電話機に 1 日でかかって来る電話の件数

などの数の分布はポアソン分布で近似できると考えられている．

連続型確率分布の典型的な例は確率密度関数を持つものがほとんどである．密度関数を持たない連続型確率分布ももちろん存在するが、初等的な段階ではあまり考える必要は無い．

定義 2.3. $\mathbb{R}$ 上の確率分布 P が確率密度関数 (単に密度関数とも言う) $f(x)$ をもつとは、次の時に言う：

(i) すべての x について $f(x) \geq 0$

(ii) $\displaystyle\int_{\mathbb{R}} f(x)dx = 1$

(iii) A を $\mathbb{R}$ の部分集合とする．確率 $P(A)$ が

$$P(A) = \int_A f(x)dx$$

で与えられる．

ただし $\int_A f(x)dx$ で f の集合 A での積分を表す．

確率分布 P が密度関数 $f(x)$ を持つ時, 分布関数 $F(z)$ と $f(x)$ の関係は

$$F(z) = P\left((-\infty, z]\right) = \int_{-\infty}^{z} f(x)dx$$

となる．したがって

$$F'(x) = f(x).$$

つまり,

定理 2.4. 確率分布 P が密度関数 $f(x)$ を持つ時, 分布関数との関係は

$$F'(x) = f(x).$$

例 2.5 (連続型確率分布の例)**.**

(1) 正規分布

m を実数値, $\sigma > 0$ とする．確率密度関数が

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

で与えられる連続型確率分布を正規分布と言い, 記号 $N(m, \sigma^2)$ と表す．確率分布 $N(0,1)$(平均 0, 分散 1 の正規分布) を標準正規分布という．

$$\int_{\mathbb{R}} \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-m)^2}{2\sigma^2}} = 1$$

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．後で明らかになるが, m, σ^2 はそれぞれ正規分布 $N(m,\sigma^2)$ の平均, 分散と一致する．数理統計ではこの正規分布が最重要な確率分布である．

(i) 日本人の身長の分布

(ii) 全国模試の試験の点数の分布

などは正規分布で近似できると考えられる．正規分布は負の値を取る確率もあるし、いくらでも大きな値を取る可能性もある．上記のデータは決して負にはならないし、テストの点は 100 点を越えないし、人間の身長が 4m にはならないであろう。あくまでも近似できるという事である．標準正規分布の分布関数は

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

で与えられるが, 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 $F(z)$ の近似値もわかることになる．

(2) 一様分布

実数 a, b $(a < b)$ を取る．$A \subset \mathbb{R}$ に対して

$$P(A) = \frac{(A \cap [a,b]) \text{ の長さ}}{b-a}$$

で定まる確率分布を区間 $[a,b]$ の一様分布と言う．例えば, $a < c < d < b$ のとき $P([c,d]) = \frac{d-c}{b-a}$ である．$[a,b]$ 上の一様分布は密度関数

$$f(x) = \frac{1_{[a,b]}(x)}{b-a}$$

をもつ確率分布と言える．ただし

$$1_{[a,b]}(x) = \begin{cases} 1 & a \leq x \leq b \text{ のとき} \\ 0 & x < a \text{ または } x > b \text{ のとき} \end{cases}$$

この確率分布の時は,$[a,b]$ 以外の数を取る確率は 0 である．$b - a = 1$ のときは $P(A)$ は A の長さと同じである．

(3) 指数分布

$\lambda > 0$ とする．密度関数

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

を持つ確率分布をパラメータ λ の指数分布と言う．$\int_{\mathbb{R}} \lambda e^{-\lambda x}dx = 1$ であることに注意しよう．この確率分布のもとでは負の数を取る確率は 0 である．指数分布は先に定義したポアソン分布と密接な関係がある．例えば,

(i) ある店に 1 日で来店する客の数
(ii) ある電話機に 1 日でかかって来る電話の件数

はポアソン分布に従うと述べたが, 客の来る時間間隔や電話のかかってくる時間の間隔は指数分布に従うと考えられる.

例題 1. 一様分布, 指数分布の分布関数を求めよ.

以上で色々なランダムな数 (例えば, ある店に 1 日で来店する客の数, 客が来る時間の間隔) の分布について述べてきた. このランダムな数というものを数学的に定式化したものが**確率変数**である.

3 確率変数

3.1 確率変数の定義

定義 3.1. (1) $(\Omega, \mathcal{F}, P)$ 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[1]
(2) 確率変数 X に対しては X が a 以上 b 以下になる確率が

$$P(\{\omega \in \Omega \mid a \leq X(\omega) \leq b\})$$

などのように定まる. より一般に $\mathbb{R}$ の部分集合 A に対して X が A の値を取る確率

$$P(\{\omega \in \Omega \mid X(\omega) \in A\})$$

も定義できる. $P(\{\omega \in \Omega \mid a \leq X(\omega) \leq b\})$, $P(\{\omega \in \Omega \mid X(\omega) \in A\})$ をそれぞれ $P(a \leq X \leq b)$, $P(X \in A)$ などと簡単に書くことが多い.

例 3.2. (1) サイコロを独立に 2 回続けて振る場合の確率空間上の確率変数
このとき標本空間は $\Omega = \{(i,j) \mid 1 \leq i \leq 6, 1 \leq j \leq 6\}$ 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.

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

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.

$$P(X_1 \leq i) = \text{サイコロを 1 回振って } i \text{ 以下の目が出る確率} = \frac{i}{6} \qquad (i = 1,2,3,4,5,6)$$

のように確率が与えられる. $Y(\omega) = X_1(\omega) + X_2(\omega), Z(\omega) = X_1(\omega)X_2(\omega), W(\omega) = \max(X_1(\omega), X_2(\omega))$ も確率変数である. これらの例では例えば X_1 は $\{1, \ldots, 6\}$, $X_1 + X_2$ は $\{2, \ldots, 12\}$ 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標本空間 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.

[1] 正確には任意の区間 $[a,b]$ について $\{\omega \mid a \leq X(\omega) \leq b\} \in \mathcal{F}$ となる関数のこと (**可測関数**とも言います) を言いますが, 今は気にする必要はありません. なぜこの条件を課すかというとこの条件がなければ集合 $\{\omega \mid a \leq X(\omega) \leq b\}$ の確率が計れないからです.

3.2 確率変数の分布

確率分布とは**標本空間 Ω が実数の部分集合であるときの確率**でした．確率変数に対してその確率分布を定義することができます．

定義 3.3. (1) X を確率空間 $(\Omega, \mathcal{F}, P)$ 上の確率変数とする．$A \subset \mathbb{R}$ に対して確率 $P_X(A)$ を

$$P_X(A) = P(X \in A)$$

のように定めれば P_X は確率分布である．この確率分布 P_X を**確率変数 X の確率分布, 確率法則** (短く**分布, 法則**) と言う．
(2) 確率変数 X に対して分布関数 F_X を

$$F_X(x) = P(X \leq x)$$

と定義する[2].

例えば確率変数 X の確率分布が P のとき, **確率分布 P に従う確率変数** X, **確率変数 X の法則は** P などと言ったりします．

3.3 確率変数の期待値・平均

確率変数の期待値、平均を離散型、連続型の場合に分けて定義する．

定義 3.4. (1) 離散型のとき
確率変数 X の取り得る値が $\{a_i\}_{i=1}^N$ で確率が $P(X = a_i) = p_i$ であるとする．
ただし,$\{a_i\}$ はすべて互いに異なる数であるとし $N = \infty$ の場合も許すことにする．
X の**平均 (期待値)** を

$$E[X] = \sum_{i=1}^{N} a_i p_i \tag{3.1}$$

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

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx \tag{3.2}$$

と定義する．

期待値・平均というのは文字通り**確率変数が平均的にどんな値を取るか**を表しているものです．

注意 3.5. (1) 上記の確率変数の期待値の定義式をよく見ると**その確率分布にのみ依存している**ことがわかる．したがって，　この期待値, 平均を**確率分布 P_X の平均, 期待値**のようにも言う．離散型と連続型で分けて定義したが, 上記の定義のいずれも確率で重みをつけて和を取っているという意味で同様な定義であることに注意しよう．

[2]確率変数 X の確率分布 P_X の定義から P_X の分布関数と確率変数 X の分布関数 F_X は同じになります．

(2) $\Omega = \{\omega_1, \ldots, \omega_n\}$ のように根元事象が有限個であるとき確率変数 X に対して期待値は

$$E[X] = \sum_{i=1}^{n} X(\omega_i) P(\{\omega_i\}) \tag{3.3}$$

のようにも書けます. ただし $P(\{\omega_i\})$ は根元事象 ω_i の確率です. 式 (3.1) と (3.3) はほとんど同じに見えますが, 少し違っています. (3.3) を X の取る値でまとめると (3.1) になります.
(3) 実は密度関数をもつ連続型確率変数の期待値の定義は離散型の場合の定義から自然に導かれる. これを説明しよう. X の分布が密度関数 $f(x)$ を持つ連続型の確率変数とする. N を大きな自然数とする. 整数 k に対して $A_k = \left\{\omega \in \Omega \mid \frac{k}{N} \le X(\omega) < \frac{k+1}{N}\right\}$ と事象を定める. $\cup_k A_k = \Omega$, $A_k \cap A_l = \emptyset$ $(k \ne l)$ となるのでどの事象 A_k が起こるかで完全に場合分けされる. そこで, A_k が起こった時 $\frac{k}{N}$ という値を取る確率変数を $X_N(\omega)$ とすると定義から

$$|X_N(\omega) - X(\omega)| \le \frac{1}{N}$$

従って N が大きい時 X_N の期待値と X の期待値はほぼ等しいと考えられる. X_N の期待値は

$$\begin{aligned}
E[X_N] &= \sum_{k=-\infty}^{\infty} \frac{k}{N} P\left(X_N = \frac{k}{N}\right) \\
&= \sum_{k=-\infty}^{\infty} k\frac{k}{N} P\left(\frac{k}{N} \le X < \frac{k+1}{N}\right) \\
&= \sum_{k=-\infty}^{\infty} \frac{k}{N} \int_{k/N}^{(k+1)/N} f(x)dx \\
&\fallingdotseq \sum_{k=-\infty}^{\infty} \frac{k}{N} f\left(\frac{k}{N}\right) \frac{1}{N} \\
&\fallingdotseq \int_{\mathbb{R}} x f(x) dx.
\end{aligned}$$

従って $E[X] = \int_{\mathbb{R}} x f(x) dx$ とするのが妥当であるとわかる.

X を確率変数とする. g を $\mathbb{R}$ 上の関数とすると $g(X)$ も確率変数になります. $g(X)$ の期待値について次が成立します. この定理は期待値の計算で非常によく使われる定理です. (1) は定義から比較的簡単にわかりますが, (2) の方は少なからず議論がいるので, 証明は述べませんが, 平均という意味から直感的には理解できるでしょう.

定理 3.6. X を確率変数とし, g を $\mathbb{R}$ 上の関数とする.
(1) X が定義 3.4 (1) の離散型のとき

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

(2) X が定義 3.4 (2) のように密度関数をもつとき

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

証明. (1) を根元事象が有限個つまり $\Omega = \{\omega_1, \ldots, \omega_n\}$ のときは注意 3.5 (2) から

$$E[g(X)] = \sum_{i=1}^{n} g(X(\omega_i)) P(\{\omega_i\}). \tag{3.4}$$

$A_i = \{\omega \mid X(\omega) = a_i\}$ とおくと $\cup_{i=1}^N A_i = \Omega$, $A_i \cap A_j = \emptyset$ $(i \neq j)$ であり $p_i = P(A_i)$. したがって

$$\begin{aligned}
\sum_{i=1}^n g(X(\omega_i))P(\{\omega_i\}) &= \sum_{i=1}^N \sum_{\omega \in A_i} g(X(\omega))P(\{\omega_i\}) \\
&= \sum_{i=1}^N g(a_i) \sum_{\omega \in A_i} P(\{\omega\}) \\
&= \sum_{i=1}^N g(a_i)P(A_i) = \sum_{i=1}^N g(a_i)p_i. \qquad (3.5)
\end{aligned}$$

□

定理 3.7 (期待値の線形性)**.** X, Y を確率変数, a, b を定数とするとき

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

とくに $g : \mathbb{R} \to \mathbb{R}$ に対して $E[g(X) + h(X)] = E[g(X)] + E[h(X)]$.

証明. $\Omega = \{\omega_1, \ldots, \omega_n\}$ のように根元事象が有限個の場合に証明する. 注意 3.5 (2) より

$$\begin{aligned}
E[aX + bY] &= \sum_{i=1}^n (aX(\omega_i) + bY(\omega_i))P(\{\omega_i\}) \\
&= \sum_{i=1}^n aX(\omega_i)P(\{\omega_i\}) \sum_{i=1}^n +bY(\omega_i))P(\{\omega_i\}) \\
&= aE[X] + bE[Y]. \qquad (3.6)
\end{aligned}$$

□

3.4 分散・標準偏差

前の節で確率変数の期待値, 分布の期待値というものを導入しました. これは確率分布にとって大事な量ですが, これだけでは確率変数がこの平均の近くに集中しているのか, あるいはものすごく散らばっているのかなどということは平均だけを見ていてもわかりません. そこで, **確率変数の値がどのぐらい平均の近くに集中しているかあるいは散らばっているかを表す尺度**が必要になります. その代表例が確率分布の**分散**です.

定義 3.8. 確率変数 X に対して**分散・標準偏差**を

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

と定義する. 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.

定理 3.6 によれば

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

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

となります.

注意 3.9. 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、分散・標準偏差は確定しますが,

(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ります. 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ですが, この分布の平均値は定義できません.

$E[g(X) + h(X)] = E[g(X)] + E[h(X)]$ という式を用いると次の定理が得られます.

定理 3.10. (1) $V[X] = E[X^2] - E[X]^2$. (2) $V[aX + b] = a^2 V[X]$.

分散を用いると確率変数 X が平均値 m から離れた値を取る確率を評価できます. 次の評価は大数の法則を証明するときに使われます.

定理 3.11 (Chebyshev(=チェビシェフ) の不等式)**.** X の分散を σ^2, 期待値を m とすると

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

4 独立性

4.1 事象の独立性

すでにサイコロ投げの時に「サイコロを振る」という試行の独立性の概念が出てきました. この独立な試行の結果は独立な事象になります. 例えば,

- 1 回目のサイコロ投げの試行で 1 が出るという事象 A
- 2 回目のサイコロ投げの試行で 5 が出るという事象 B

は独立な事象になります.
赤玉 3 個, 白玉 2 個が入っている袋から玉を無作為に取り出してはもとに戻すという復元抽出を 2 回行ったときに

- 1 回目の試行で赤玉が取り出されるという事象 C
- 2 回目の試行で白玉が取り出されるという事象 D

もやはり独立となります. この**独立な事象**というものを数学的に定式化して実際に上の事象が独立であることを示して見よう.

定義 4.1. 事象 A, B が独立であるとは

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

をみたすときに言う.

上記の事象 A, B は適当な確率空間 $(\Omega, \mathcal{F}, P)$ があって A, B は標本空間 Ω の部分集合 (もっと正確には $\mathcal{F}$ の元) であるから $A \cap B \subset \Omega$ となり確率 $P(A \cap B)$ が定まることに注意しよう．

例 4.2. 上記の玉を取り出す復元抽出のときの確率空間が何かを考えよう．試行の結果は 1 回目,2 回目の試行で出る玉で決まる．赤玉の一つ一つを区別して R_1, R_2, R_3, 白玉を区別して W_1, W_2 としよう (赤を R, 白を W で表現する)．根元事象をすべて並べて標本空間は

$$\Omega = \{(R_*, R_*), (R_*, W_*), (W_*, R_*), (W_*, W_*) \mid * \text{ は } R \text{ のときは } 1, 2,\ W \text{ のときは } 1, 2, 3 \text{ の値を取る } \}$$

で与えられる．またこの $5 \times 5 = 25$ 個の根元事象どれもがすべて同様に確からしく起こる．さて

$$\begin{aligned} C &= \{(R_1, *), (R_2, *) \mid * \text{ は 5 個のうちどの玉でもよい } \} & (4.1) \\ D &= \{(*, W_1), (*, W_2), (*, W_3) \mid * \text{ は 5 個のうちどの玉でもよい } \} & (4.2) \\ C \cap D &= \{(R_i, W_j) \mid i = 1, 2, j = 1, 2, 3\} & (4.3) \end{aligned}$$

なので,

$$\begin{aligned} P(C) &= \frac{\sharp C}{25} = \frac{10}{25} & (4.4) \\ P(D) &= \frac{\sharp D}{25} = \frac{15}{25} & (4.5) \\ P(C \cap D) &= \frac{\sharp (C \cap D)}{25} = \frac{2 \times 3}{25} = \frac{10}{25} \times \frac{15}{25} = P(C)P(D) & (4.6) \end{aligned}$$

となり確かに独立である．

演習問題をあげる．

例題 2. 独立に 2 回サイコロを投げる試行において 1 回目に 1 の目が出る事象 A, 2 回目に 5 の目が出る事象 B が独立であることを確認せよ．確率空間が何で，　同様に確からしい根元事象は何かに注意して確認すること．

例題 3. 赤玉 3 個, 白玉 2 個が入っている袋から玉を無作為に取り出し，　取り出した玉をもとに戻さず再度玉を無作為に取り出すという非復元抽出を行うとき, 1 回目の試行で赤玉が取り出されるという事象 C と 2 回目の試行で白玉が取り出されるという事象 D は独立ではない. 同様に確からしい根元事象が何か明らかにしてこれを定義に従って示せ.

例題 4. 事象 A, B が独立ということとその余事象 A^c, B^c が独立ということが同値であることを示せ.

ヒント：確率の加法性を用いる.

高校数学の数学 C で学んだように条件付き確率を用いて事象の独立性を言い換えることができます.

定義 4.3. $(\Omega, \mathcal{F}, P)$ を確率空間とする．正の確率の事象 B を取る．

$$P(A|B) = \frac{P(A \cap B)}{P(B)}, \quad A \in \mathcal{F}$$

と定めると $P(\cdot \mid B)$ は B を標本空間とする確率になる．この確率 $P(\cdot \mid B)$ を事象 B で条件つけられた条件付き確率と言う．

条件付き確率を用いると事象の独立性は次のように言い換えられる．

命題 4.4. $P(B) > 0, P(A) > 0$ とする．事象 A, B が独立であるための必要十分条件は

$$P(A|B) = P(A), \quad P(B|A) = P(B)$$

となることである．すなわち A **と** B **が独立というのは** B **で条件つけても** A **の起こる確率は変わらないし，** A **で条件つけても** B **の起こる確率は変わらないということと同値である**.

最後に 3 個以上の場合に事象の独立を定義しよう．

定義 4.5. 事象 $A_1, \dots, A_n$ が独立とは任意の $\{i_1, \dots, i_k\} \subset \{1, \dots, n\}$ に対して

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

例えば事象 A, B, C が独立とは

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

がすべて成立するときに言います．注意すべきは

$$A \text{ と } B \text{ が独立},\ A \text{ と } C \text{ が独立},\ B \text{ と } C \text{ が独立}$$

だとしても A, B, C が独立とは限らないということです． 3つ以上の事象の独立性は思ったほど単純では無いということです. 次に確率変数の独立性を定義しましょう．

4.2 確率変数の独立性

事象の独立性を定義しましたがこれに基づいて確率変数の独立性を定義しよう．例えば，サイコロを 2 回振る独立試行で 1 回目の目の数を X_1, 2 回目の目の数を X_2 として, 事象を

$$\begin{aligned} A &= \{X_1 = i\} = \{1 \text{ 回目に } i \text{ が出るという事象 }\} && (4.7) \\ B &= \{X_2 = j\} = \{2 \text{ 回目に } j \text{ が出るという事象 }\} && (4.8) \end{aligned}$$

と定義すると A, B は独立な事象なので,

$$P(A \cap B) = P(A)P(B). \tag{4.9}$$

(4.9) を書き直すと

$$P\left(X_1 = i, X_2 = j\right) = P\left(X_1 = i\right) P\left(X_2 = j\right).$$

この事と一般に確率変数は連続的に値を取り得ることを考慮して一般的に二つの確率変数の独立性を定義しよう．

定義 4.6. 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 $a < b, c < d$ について

$$P\left(X \in [a, b], Y \in [c, d]\right) = P(X \in [a, b])P(Y \in [c, d])$$

が成立するときに言う．

注意 4.7. X, Y が独立とはすべての $a < b, c < d$ に対して事象 $\{\omega \mid X(\omega) \in [a, b]\}$ と事象 $\{\omega \mid Y(\omega) \in [c, d]\}$ とが独立になるということと同値です.

例 4.8. 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 $\Omega = \{(i,j) \mid 1 \le i \le 6, 1 \le j \le 6\}$ 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．

$$
\begin{aligned}
X_1(\omega) &= i \ (\omega = (i,j) \text{ のとき}) \\
X_2(\omega) &= j \ (\omega = (i,j) \text{ のとき})
\end{aligned}
$$

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 $X_1{-}X_2$ は独立では無い．なぜなら $P\left(X_1 + X_2 = 5, X_1 - X_2 = 0\right) = 0$ だが $P(X_1{+}X_2 = 5) > 0$, $P(X_1 - X_2 = 0) > 0$ だから．

例題 5. サイコロを続けて 2 回振る独立試行を考える．1 回目の目の数を X_1, 2 回目の目の数を X_2 とし $Y = X_1 + X_2, Z = X_1X_2, W = \max(X_1, X_2)$ と新たな確率変数を定める．Y, Z, W の確率分布と平均を求めよ．

二つの確率変数の独立性を定義しましたが, もっと一般に n 個の確率変数 $X_1, \ldots, X_n$ の独立性も定義できます．

定義 4.9 (n 個の確率変数の独立性)**.** 確率変数 $X_1, \ldots, X_n$ が独立とは任意の区間 $I_1, \ldots, I_n$ ($[a_i, b_i]$ のような区間) に対して

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

となる時に言う．すなわち, 任意の区間 I_i に対して事象

$$\{X_1 \in I_1\}, \ldots, \{X_n \in I_n\}$$

が独立のときに言う．

5 重要な確率分布

すでに確率分布として

正規分布，ポアソン分布，一様分布，指数分布

などの重要な例をあげました．ここではこれらの平均, 分散を求めよう．さらにこれら以外に重要な確率分布も導入する．

5.1 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, $B(n, p)$ と表します．この成功の回数はある独立な確率変数の和として表されます．$X_i\ (1 \le i \le n)$ を $P(X_i = 1) = p$, $P(X_i = 0) = 1 - p$ となる独立な確率変数とします (すなわち成功したら 1, 失敗したら 0 となる).

$$S_n = X_1 + \cdots + X_n \tag{5.1}$$

と定義すると S_n の分布が二項分布である. 成功が k 回という事象は $\{S_n = k\}$ のことですが, i_1 回,...,i_k 回目の試行で成功し, 他の試行では失敗している事象を $A_{i_1,\dots,i_k}$ と表すと独立の定義から

$$P(A_{i_1,\dots,i_k}) = P\left(X_i = 1\ (i = i_1,\cdots,i_k), X_j = 0\ (j \neq i_1,\cdots,i_k)\right) = p^k(1-p)^{n-k}.$$

$i_1,\dots,i_k$ が動く時 $A_{i_1,\dots,i_k}$ は排反事象で事象 $\{S_n = k\}$ のすべてをつくすので、確率の加法性から

$$\begin{aligned} P\left(S_n = k\right) &= \sum_{i_1,\cdots,i_k} P(A_{i_1,\dots,i_K}) \\ &= {}_n\mathrm{C}_k p^k(1-p)^{n-k}. \end{aligned} \tag{5.2}$$

まとめると $\{0,1,\dots,n\}$ のいずれかの値を取る離散型確率変数 X で

$$P(X = k) = {}_n\mathrm{C}_k p^k q^{n-k} \qquad 0 \le k \le n$$

となる確率変数の分布が二項分布 $B(n,p)$ です. 式 (5.1) および期待値の線形性を用いて S_n の平均を求めることができます. まず

$$E[X_i] \quad = \quad 1 \cdot p + 0 \cdot (1-p) = p. \tag{5.3}$$

従って

$$E[S_n] \quad = \quad E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = np. \tag{5.4}$$

まとめると

定理 5.1. X が二項分布 $B(n,p)$ に従うとする.
(1) $E[X] = np$, (2) $V[X] = npq$,

分散については注意 6.2 (3) で証明を与える.

(2) ポアソン分布

ポアソン分布の定義およびどのようなランダムな現象がポアソン分布に従うかはすでに説明した. 定義は以下の通りであった.

定義 5.2. $\lambda > 0$ 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

$$P\left(\{k\}\right) = e^{-\lambda}\frac{\lambda^k}{k!} \qquad k = 0,1,\dots$$

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.

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 $n\to\infty$ としてみよう．

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

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

定理 5.3. 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と $E[X]=V[X]=\lambda$. 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

$$P(X=k)=(1-p)^k p, \qquad k=0,1,2,\ldots$$

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 5.4. X がパラメータ p の幾何分布に従うとする．このとき $E[X]=\frac{1-p}{p}$, $V[X]=\frac{1-p}{p^2}$. すべての n について $E[X^n]<\infty$.

平均は次のように計算できる．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．

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

5.2 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

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

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

密度関数

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

をもつ確率分布をパラメータ λ の指数分布と言った．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 $t=\frac{k}{n}$ $k=0,1,\ldots$ 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 $n\to\infty$ 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．$P(T=k/n)=\left(1-\frac{\lambda}{n}\right)^k\frac{\lambda}{n}$ (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．

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

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．

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 $p_k={}_{k_t+1}\mathrm{C}_k\left(\frac{\lambda}{n}\right)^k\left(1-\frac{\lambda}{n}\right)^{k_t+1-k}$. これは二項分布 $B(k_t+1,p)$ $(p=\lambda/n)$ に従う．$(k_t+1)\frac{\lambda}{n}\fallingdotseq\lambda t$ だから $n\to\infty$(したがって $k_t\to\infty$ となることに注意) で $\lim_{n\to\infty}p_k=\frac{(\lambda t)^k}{k!}e^{-\lambda t}$. 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 5.5. X がパラメータ λ の指数分布に従うとき，$E[X]=\frac{1}{\lambda}$, $V[X]=\frac{1}{\lambda^2}$, $E[X^n]=\frac{n!}{\lambda^n}$.

(3) 正規分布 $m\in\mathbb{R},\sigma>0$ とする．確率密度関数

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

をもつ連続型の $\mathbb{R}$ 上の確率分布を正規分布といい、$N(m,\sigma^2)$ と表した．

命題 5.6. 確率変数 X が正規分布 $N(m,\sigma^2)$ に従うとする．
(1) 確率変数 X が正規分布 $N(m,\sigma^2)$ に従うとする．このとき $E[X]=m$, $V[X]=\sigma^2$.
(2) $p(\neq 0),q$ 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．

(2) 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).

Proof. (1)

$$\begin{aligned}
E[X] &= \int_{\mathbb{R}} x\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{(x-m)^2}{2\sigma^2}\right)dx\\
&= \int_{\mathbb{R}}(\sigma y+m)\frac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\frac{y^2}{2}\right)\sigma dy (y=\tfrac{x-m}{\sigma}\text{ と変換した})\\
&= \int_{\mathbb{R}}(\sigma y+m)\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy=m\quad (\textstyle\int_{\mathbb{R}}\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{y^2}{2}\right)dy=1\text{ を用いた}).
\end{aligned}$$

$$\begin{aligned}
V[X] &= E[(X-E[X])^2] \\
&= E[(X-m)^2] \\
&= \int_{\mathbb{R}} (x-m)^2 \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx \\
&= \int_{\mathbb{R}} (\sigma y)^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = \sigma^2, \qquad (5.5)
\end{aligned}$$

ここで

$$\int_{\mathbb{R}} y^2 \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{y^2}{2}\right) dy = 1 \qquad (5.6)$$

を用いた. (5.6) は 1 回部分積分をすれば示される (1 年の微積分の講義でやっているはず!).

(2) $p>0$ の場合に示す. $p<0$ の場合も同様に示せる.

$$\begin{aligned}
P(pX+q \le z) &= P\left(X \le \frac{z-q}{p}\right) \\
&= \int_{-\infty}^{\frac{z-q}{p}} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-m)^2}{2\sigma^2}\right) dx. \qquad (5.7)
\end{aligned}$$

$x=\frac{u-q}{p}$ と変数変換すると

$$\begin{aligned}
P(pX+q \le z) &= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(\frac{u-q}{p}-m)^2}{2\sigma^2}\right) \frac{1}{p} du \\
&= \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi\sigma^2 p^2}} \exp\left(-\frac{(u-q-mp)^2}{2\sigma^2 p^2}\right) du. \qquad (5.8)
\end{aligned}$$

(5.8) の右辺の積分の密度関数が $pX+q$ の確率密度関数である. 従って $pX+q$ は平均は $mp+q$, 分散 $p^2\sigma^2$ の正規分布に従うとわかる. □

今証明した命題から $\frac{X-m}{\sigma}$ は標準正規分布に従うことがわかります. 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 $F(x)=P(T\le x)$ がわかれば X の分布もわかることになります. というのは

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

となるから.

6 多次元の確率分布

実数に値を取るランダムな量 (確率変数) に対して確率分布が定まりました. しかし, このランダムな量というのは数の組のように 2 次元,3 次元, 一般的に n 次元的な広がりがあれば多次元の確率分布を定めることになります. まず簡単な 2 次元の確率分布について説明しましょう.

6.1　2 次元の確率分布

定義 6.1. (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) $\mathbb{X} = (X, Y)$ について $A \subset \mathbb{R}^2$ の確率を

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

で定めることができる．この $\mathbb{R}^2$ 上の確率を $\mathbb{X} = (X, Y)$ 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．

X, Y が離散型、連続型の場合にこの定義 6.1 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．(X, Y) は $\{(a_i, b_j)\}_{1 \leq i \leq M, 1 \leq j \leq N}$ のいずれかの点の値を取る．

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

とする．$\sum_{1 \leq i \leq M, 1 \leq j \leq N} p_{ij} = 1$ であり (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 6.2 (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．

2 次元の確率分布があるとき, その周辺分布は一意的に決まりますが, 周辺分布が同じでももとの同時分布が同じとは限りません. 各自例を考えて見て下さい.

(2) X, Y の同時分布が密度関数を持つ連続型のとき

(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

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

となる時に言う．命題 6.2 に対応して次の命題が成立する．

命題 6.3. (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

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

で与えられる．

離散型確率変数の時は，独立性を次のように言い換えることができる．

定理 6.4. X, Y を離散型の確率変数で $P(X = a_i) = p_i \ (1 \leq i \leq M)$, $P(Y = b_j) = q_j \ (1 \leq j \leq N)$,

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

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．

命題 6.5. 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 6.6. (1) X, Y を独立な確率変数とすると $E[XY] = E[X]E[Y]$.
(2) X, Y の共分散を $\mathrm{Cov}(X,Y) = E[(X-E[X])(Y-E[Y])]$ と定義する. X, Y が独立ならば $\mathrm{Cov}(X,Y) = 0$.
(3) X, Y を独立な確率変数とすると $V[\alpha X + \beta Y] = \alpha^2 V[X] + \beta^2 V[Y]$.

証明. (1) X, Y が離散型確率変数のときに証明する. $\{x_1, \ldots, x_n\}$, $\{y_1, \ldots, y_m\}$ をそれぞれ X, Y の取る値とする. $E_i = \{\omega \mid X(\omega) = x_i\}$, $F_j = \{\omega \mid X(\omega) = y_j\}$ とおくと

$$X(\omega) = \sum_{i=1}^{n} x_i 1_{E_i}(\omega), \quad Y(\omega) = \sum_{j=1}^{m} y_j 1_{F_j}(\omega),$$

ここで 1_A は $1_A(\omega) = 1 \ \ (\omega \in A$ のとき$)$, $1_A(\omega) = 0 \ \ (\omega \in A^c$ のとき$)$ となる A の定義関数である. ゆえに

$$\begin{aligned}
E[XY] &= E\left[\left(\sum_{i=1}^{n} x_i 1_{E_i}\right)\left(\sum_{j=1}^{m} y_j 1_{F_j}(\omega)\right)\right] \\
&= \sum_{1 \le i \le n, 1 \le j \le m} x_i y_j E[1_{E_i} 1_{F_j}] \\
&= \sum_{1 \le i \le n, 1 \le j \le m} x_i y_j P(E_i \cap F_j) \qquad (6.1) \\
&= \sum_{1 \le i \le n, 1 \le j \le m} x_i y_j P(E_i) P(F_j) \qquad (6.2) \\
&= \left(\sum_{1 \le i \le n} x_i P(E_i)\right)\left(\sum_{1 \le j \le m} y_j P(F_j)\right) = E[X]E[Y].
\end{aligned}$$

(6.1),(6.2) で

$$E[1_{E_i} 1_{F_j}] = E[1_{E_i \cap F_j}] = P(E_i \cap F_j),$$

$$P(E_i \cap F_j) = P(\{X = x_i, Y = y_j\}) = P(X = x_i) P(Y = y_j) = P(E_i) P(F_j)$$

となることを用いた. (2) は (1) の結果から従う. (3) を示す.

$$\begin{aligned}
V[\alpha X + \beta Y] &= E\left[\left\{(\alpha X + \beta Y) - E[\alpha X + \beta Y]\right\}^2\right] \\
&= E\left[\left\{\alpha(X - E[X]) + \beta(Y - E[Y])\right\}^2\right] \\
&= \alpha^2 E[(X - E[X])^2] + \beta^2 E[(Y - E[Y])^2] + 2\alpha\beta E[(X - E[X])(Y - E[Y])] \\
&= \alpha^2 V[X] + \beta^2 V[Y] \qquad (6.3)
\end{aligned}$$

□

注意 6.7. (1) X, Y が独立の時 $E[XY] = E[X]E[Y]$ となるということは非常によく使われる大事な性質です. 注意してほしいのは二つの確率変数 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ということです.
(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 $E[\phi(X)\psi(Y)] = E[\phi(X)]E[\psi(Y)]$ となるということは同値です.

(3) $X_1, \ldots, X_n$ が独立の時

$$V\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n V[X_i].$$

(4) 定理 6.6 を用いて二項分布の分散を求めてみよう．まず X_i を $P(X_i = 1) = p, P(X_i = 0) = 1 - p$ となる独立確率変数とすると $S_n = X_1 + \cdots + X_n$ の分布が二項分布 $B(n, p)$ です．

$$\begin{aligned} E[X_i] &= 1 \cdot p + 0 \cdot (1-p) = p && (6.4) \\ E[(X_i - p)^2] &= (1-p)^2 p + (-p)^2(1-p) = p(1-p)\{1-p+p\} = p(1-p). && (6.5) \end{aligned}$$

従って

$$\begin{aligned} E[S_n] &= E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n E[X_i] = np && (6.6) \\ E[(S_n - E[S_n])^2] &= E\left[\left(\sum_{i=1}^n (X_i - E[X_i])\right)^2\right] = E\left[\sum_{i=1}^n (X_i - E[X_i])^2\right] = np(1-p). && (6.7) \end{aligned}$$

6.2 $\mathbb{R}^n$ 上の確率分布

定義 6.8. (1) $X_1, \ldots, X_n$ を同じ確率空間で定義された確率変数とする．n **次元確率変数** $\mathbb{X} = (X_1, \ldots, X_n)$ に対して $\mathbb{R}^n$ 上の**確率分布**を

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

のように定めることができる．$P_{\mathbb{X}}$ を $\mathbb{X}$ の**確率分布,(確率) 法則**と言う．
(2) n 次元確率変数 $\mathbb{X} = (X_1, \ldots, X_n)$ の分布が**密度関数** $f(x_1, \ldots, x_n)$ を持つとは, 任意の部分集合 $A \subset \mathbb{R}^n$ に対して

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

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の場合と同様に定義する．

n 次元確率変数 $\mathbb{X}$ に対して**平均 (ベクトル)**，**共分散行列**を定義する．**共分散行列**は**分散**の多次元版にあたるものです．

定義 6.9. n 次元確率変数 $\mathbb{X} = (X_1, \ldots, X_n)$ に対して n 次元ベクトル

$$m = (E[X_1], \ldots, E[X_n]) \tag{6.8}$$

を $\mathbb{X}$ の平均 (ベクトル) という．また (i, j) 成分が

$$E[(X_i - E[X_i])(X_j - E[X_j])]$$

である n 次正方行列を $\mathbb{X}$ の共分散行列と言う．

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．多次元正規分布について説明しよう．

定義 6.10. 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 $f(x_1, \ldots, x_n)$ が

$$f(x_1, \ldots, x_n) \quad = \quad \frac{1}{\sqrt{2\pi}^n \sqrt{\det A}} \exp\left(-\frac{1}{2} \sum_{1 \le i \le j \le n} (A^{-1})_{ij}(x_i - m_i)(x_j - m_j) \right)$$

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．$(A^{-1})_{ij}$ は A の逆行列 A^{-1} の (i, j) 成分を表す．これは 1 次元の正規分布の拡張である．

$f(x_1, \ldots, x_n)$ が

$$\int_{\mathbb{R}^n} f(x_1, \ldots, x_n) dx_1 \cdots dx_n = 1$$

をみたすことは積分の変数変換の公式から証明されることを注意しておく．

m, A は次のような意味がある．

命題 6.11. n 次元確率変数 $\mathbb{X}$ が多次元正規分布 $N(m, A)$ に従うとき X の平均ベクトルは m, 共分散行列は A である．

次の定理は定理 3.6 の多次元版です．

定理 6.12. (1) $\mathbb{X} = (X_1, \ldots, X_n)$ が離散型の確率変数で $\mathbb{X}$ の取る値 (ベクトル) を $\{a_1, \ldots, a_N\}$ とし，$P(\mathbb{X} = a_i) = p_i$ $(1 \le i \le N)$ 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) $\mathbb{X} = (X_1, \ldots, X_n)$ の結合分布が連続型の分布で確率密度関数 $f(x_1, \ldots, x_n)$ を持つとき，$\mathbb{R}^n$ 上の有界関数 φ に対して

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

確率変数 $X_1, \ldots, X_n$ のひとつひとつ X_i が密度関数をもち, $X_1, \ldots, X_n$ が独立のときは, $(X_1, \ldots, X_n)$ の結合分布は X_i たちの密度関数で決まってしまう．すなわち

命題 6.13. n 個の確率変数 $\{X_1, \ldots, X_n\}$ 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．$\{X_1, \ldots, X_n\}$ 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 $\mathbb{X} = (X_1, \ldots, X_n)$ の分布が密度関数 $f(x_1, \ldots, x_n) = f_1(x_1) \cdots f_n(x_n)$ を持つ事である．

多次元正規分布, 正規分布に従う独立確率変数に対して次の定理 6.14 が成立します．**区間推定**や**検定**で重要な t **分布**に関する定理 (定理 8.6) の証明で必要になる重要な定理です．t-分布はギネス社の研究者 **William Gosset** (1876–1937) により発見された分布で Gosset のペンネーム"Student" を冠して **Student の t-分布**とも言います．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、ペンネームを用いていたとのことです．この定理をきちんと証明しようとすると面倒なのと, 線形代数の知識を必要とするので, 初等的な数理統計の教科書ではあまり述べられません．重要な定理なので, ここでは参考のために述べておきます．

定理 6.14. (1) $\mathbb{X} = (X_1, \ldots, X_n)$ を多次元正規分布 $N(m, A)$ に従う n 次元確率変数とする．P を (k, n) 行列で階数が k, v を k 次元の縦ベクトルとする．

$$\mathbb{Y} = {}^t\left(P {}^t\mathbb{X} + v\right)$$

と k 次元確率変数を定義する (^{t} は行列の転置を表す). $\mathbb{Y}$ の法則も (k 次元) 正規分布である.
(2) n 次元確率変数 $\mathbb{X} = (X_1, \ldots, X_n)$ の分布が多次元正規分布に従うとする. $\{X_1, \ldots, X_n\}$ が独立であるための必要十分条件はすべての i, j $(i \neq j)$ について X_i, X_j が独立であることである.
(3) 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 $\mathrm{Cov}(X, Y) = 0$ である.

注意 6.15 (正規分布の特殊性について).
(1) $(X_1, \ldots, X_n)$ の法則が正規分布に従うとき定理 6.14 (2) はペアごとに確率変数 X_i, X_j が独立ならば $\{X_1, \ldots, X_n\}$ が独立と言っていますが, まったく一般の確率変数を考えるとこれは成立しません. 事象の独立性も 3 個以上の場合, ペアごとに事象が独立でも 3 個の事象としては独立とは限らないと述べたことを思い出して下さい。それと同様です. したがって正規分布に従う確率変数は非常に特殊な (あつかいやすい) 性質を持つことがわかります.
(2) 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します. このとき $X_1 + X_2$ と $X_1 - X_2$ は独立です. これは $\mathrm{Cov}(X_1 + X_2, X_1 - X_2) = E[(X_1 + X_2)(X_1 - X_2)] = 1 - 1 = 0$ と定理 6.14 (3) からわかります. 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ました. やはり正規分布というものが特殊なわけです.

系 6.16. 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.
(1) $(X_1, \ldots, X_n)$ の結合分布は n 次元正規分布 $N(m, A)$ である. ただし $m = (m_1, \ldots, m_n)$, A は (i, i) 成分が σ_i^2 となる対角行列. すなわち

$$P\left((X_1, \ldots, X_n) \in A\right) = \int_A \prod_{i=1}^n \frac{1}{\sqrt{2\pi\sigma_i^2}} \exp\left(-\frac{(x_i - m_i)^2}{2\sigma_i^2}\right) dx_1 \cdots dx_n.$$

(2) $\sum_i a_i^2 > 0, \sum_i b_i^2 > 0$ かつ n 次元ベクトル

$$(a_1, \ldots, a_n), \quad (b_1, \ldots, b_n)$$

が一次独立とする.

$$Y = \sum_{i=1}^n a_i X_i, \quad Z = \sum_{i=1}^n b_i X_i$$

と確率変数を定義する. このとき (Y, Z) の共分散行列は正定値かつ (Y, Z) の分布は 2 次元正規分布に従う.
(3) (2) の状況で Y と Z が独立になるための条件は $\sum_{i=1}^n a_i b_i \sigma_i^2 = 0$ となることである.

証明. 上の定理の (1) は命題 6.13 からわかる. (2) は定理 6.14 (1) から直ちに従う. (3) は定理 6.14 (3) から従う. なぜなら

$$\begin{aligned}
\mathrm{Cov}(Y, Z) &= E\left[\left(\sum_{i=1}^n a_i (X_i - E[X_i])\right) \cdot \sum_{j=1}^n b_j (X_j - E[X_j])\right] \\
&= \sum_{i,j=1}^n a_i b_j E\left[(X_i - E[X_i])(X_j - E[X_j])\right] \\
&= \sum_{i=1}^n a_i b_i E[(X_i - E[X_i])^2] = \sum_i a_i b_i \sigma_i^2 \qquad (6.9)
\end{aligned}$$

となるから. □

注意 6.17. 系 6.16 の証明からわかるように $\mathrm{Cov}(X_i, X_j) = 0$ $(i \neq j)$ ならば

$$\mathrm{Cov}(\sum_{i=1}^{n} a_i X_i, \sum_{i=1}^{n} b_i X_i) = \sum_{i=1}^{n} a_i b_i \mathrm{Cov}(X_i, X_i).$$

7 極限定理

7.1 大数の法則

確率変数 $X_1, X_2, \ldots, X_n$ に対して

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

と定め, **経験的な平均**、**標本平均**と呼ぶ.

定理 7.1. (1) 確率変数 $X_1, \ldots, X_n$ のおの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 $X_1, \ldots, X_n$ が独立ですべて平均 m, 分散 σ^2 ならば $\bar{X}_n$ の分散は $\frac{\sigma^2}{n}$.

この節では無限個の独立な確率変数

$$X_1, X_2, \ldots, X_n, \ldots$$

を考えます. 無限個の確率変数の独立性を定義しましょう.

定義 7.2. (1) 確率変数の無限列 $\{X_1, X_2, \ldots, X_n, \ldots\}$ が独立とは任意の n について $\{X_1, \ldots, X_n\}$ が独立の時に言う.
(2) 確率変数の無限列 $\{X_1, X_2, \ldots, X_n, \ldots\}$ 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.

定理 7.3 (大数の法則)**.** $X_1, X_2, \ldots$ 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. $S_n = X_1 + \cdots + X_n, \bar{X}_n = \frac{S_n}{n}$ とする. 任意の正数 ε について

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

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.

注意 7.4. (1) $\{X_i\}_{i=1}^{\infty}$ が i.i.d. で平均と分散が存在すれば定理の仮定はみたされる.
(2) 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 $m = \frac{1}{6} \times 1 + \frac{1}{6} \times 2 + \cdots + \frac{1}{6} \times 6 = 3.5$ だから $\frac{X_1 + \cdots + X_n}{n}$ 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.

(3) 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 7.3 は Chebyshev(チェビシェフ) の不等式を用いて得られる式

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

から従う.

7.2 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.

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

$$T_n = \frac{\sqrt{n}\cdot(\bar{X}_n - m)}{\sigma} = \frac{S_n - nm}{\sqrt{n\sigma^2}} \tag{7.1}$$

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 $\sqrt{n}$ 倍していて

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

のように正規化されているのである. したがって $n \to \infty$ 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 7.5 (中心極限定理)**.** $X_1, X_2, \ldots, X_n, \ldots$ 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. T_n を (7.1) のように正規化した確率変数とする. すべての $a < b$ となる実数について

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

注意 7.6. 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 $\dfrac{\sqrt{n}(\bar{X}_n - m)}{\sigma}$ 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 $(i = 1, 2, \ldots)$ 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 $E[X_i] = p, \mathrm{V}[X_i] = p(1-p)$ だから

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

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

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

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.

例題 6. 硬貨を 100 回投げたとき, 表の出た回数を S とする $P(45 \le S \le 55)$ 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

$$P\left(45 \le S \le 55\right) \fallingdotseq P\left(44.5 \le \tilde{S} \le 55.5\right)$$

$\tilde{S}$ 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

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

半目の補正をしないと $P(-1 \le T \le 1) = 0.68$.
(2)

$$P\left(45 \le S \le 55\right) = \sum_{k=45}^{55} {}_{100}\mathrm{C}_k \left(\frac{1}{2}\right)^{100}$$

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.

8 標本分布

8.1 標本調査、母集団

定義 8.1. (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る. **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**を**標本**と言う. 標本は無作為に選ばれなければならない. 無作為に選ばれた標本

$$X_1, X_2, \ldots, X_n, \ldots$$

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 *(*理想化された意味での*)* 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 $X_1, X_2, \ldots$ から計算される以下のような量を**統計量**と言う.

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

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

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

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

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 $(\diamondsuit)$ は $\bar{X}_n$ の所が m になっていれば大数の法則 (定理 7.3) 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 8.2. $\bar{X}_n$, 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．

8.2 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 $u_n^2=\frac{1}{n-1}\sum_{i=1}^{n}(X_i-\bar{X}_n)^2$ で分散の代用にすることがある．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 8.3. $Z_1,\dots,Z_n$ 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

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

を持つ．

ただし

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

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 8.3 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) 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．

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

この式で $\varphi(x)=1_{(-\infty,z]}(x)$ とすると

$$
P\left(Z_1^2+Z_2^2 \le z\right) = \int_0^z \frac{e^{-t/2}}{2} dt.
$$

したがって密度関数は $t>0$ で $f_2(t)=\dfrac{e^{-t/2}}{2}$, $t \le 0$ 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 8.4. 定理 8.3 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 8.5. $X_1,\ldots,X_n,\ldots$ 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 8.6. $X_1,\ldots,X_n$ 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 8.7. 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 6.14 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．$\sum_{i=1}^n Y_i=0$ からもわかるように $Y_1,\ldots,Y_n$ 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

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

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

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

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 6.14 によれば $U_1, \ldots, U_{n-1}$ は標準正規分布に従う独立確率変数だから $\displaystyle\sum_{i=1}^{n} \left(\frac{X_i - \bar{X}_n}{\sigma}\right)^2$ 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 $X_1, \ldots, X_n$ 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 8.8. 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

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

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.

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

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

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 8.9. $X_1,\ldots,X_n$ 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) $\sqrt{n}\dfrac{\bar{X}_n-m}{\sigma}$ は標準正規分布に従う (定理 6.14)

(ii) $\dfrac{(n-1)u_n^2}{\sigma^2}$ は自由度 $n-1$ の χ^2 分布に従う (定理 8.6)

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

$\mathrm{Cov}(\bar{X}_n,X_i-\bar{X}_n)=0$ なので定理 6.14 (3) 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 8.10. (1) 定理 6.14 (3) 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

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

したがって独立である．しかし、本当はこの議論だけだと不十分である．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

$(\star)$ 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、$\mathrm{Cov}(\sum_{i=1}^{n} Y_i, Z) = \sum_{i=1}^{n}\mathrm{Cov}(Y_i, Z) = 0$ となるので $(\star)$ は正しい命題である．本当はもう少し議論が必要であるがこのような議論で $\bar{X}_n$ 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 $n \to \infty$ 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 $\dfrac{e^{-x^2/2}}{\sqrt{2\pi}}$ に近づくと予想される．この予想は正しい．

$$
\begin{aligned}
\lim_{n\to\infty}\sqrt{n}\int_0^1 x^{-1/2}(1-x)^{\frac{n}{2}-1}dx &= \sqrt{2\pi} \qquad (\clubsuit) \\
\lim_{n\to\infty}\left(1 + \frac{x^2}{n}\right)^{-\frac{n+1}{2}} &= e^{-\frac{x^2}{2}}. \qquad (\spadesuit)
\end{aligned}
$$

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．

9 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 7. 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 $\sigma = 0.50$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

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

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

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

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

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

□

注意 9.1. (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に依存することも注意すること.

(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

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

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 $X_1, \ldots, X_n$ 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

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

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 8. 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 (定理 8.9) によれば $X_1, \ldots, X_n$ 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 $\bar{X}_n = \frac{X_1 + \cdots + X_n}{n}$ 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

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

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
$$\left[2.52 - \frac{0.11}{\sqrt{19}} \times 2.09,\ 2.52 + \frac{0.11}{\sqrt{19}} \times 2.09\right] \fallingdotseq [2.45, 2.57].$$
が信頼度 95 %の信頼区間である． □

注意 9.2. (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
$$P\left(|t| \geq t_n(\alpha)\right) = \alpha$$
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
$$\text{両側 5 \%点} \quad t_{19}(0.05) = 2.09, \qquad \text{両側 1 \%点} \quad t_{19}(0.01) = 2.86$$
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 $X_1, \ldots, X_n$ 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
$$\left[\bar{x}_n - \frac{\bar{s}_n}{\sqrt{n-1}} \times t_{n-1}(\alpha),\ \bar{x}_n + \frac{\bar{s}_n}{\sqrt{n-1}} \times t_{n-1}(\alpha)\right] \qquad (\star 2)$$
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 に属すか属さないという状況で A に属すという比率 (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 9. 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.

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

となる独立確率変数 $X_1, \ldots, X_n$ の和 $S_n = X_1 + \cdots + X_n$ 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 (分布は**二項分布** $B(n, p)$). $V[X_i] = p(1-p)$ 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 %の信頼区間を求める。。。。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 10. 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 $X_1, \ldots, X_n$ 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 11. (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 $X_1, \ldots, X_n$ 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 12. 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.

10 仮説検定

10.1 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 10.1. 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, **帰無仮説**という) を考える. この仮説が正し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

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

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 10.2. 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．逆に仮説が正しいのに仮説を棄却してしまうこともありえる．この間違いを**第一種の過誤**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 $X_1, \ldots, X_n$ の統計量 $T(X_1, \ldots, X_n)$ の分布を決定する.

(3) 有意水準 (危険率)α を設定し、

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

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 $x_1, \ldots, x_n$ を統計量に代入した値 $T(x_1, \ldots, x_n)$ 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

$$W_{0.05} = \{n \mid 0 \le n \le 74 \text{ または } n \ge 106\}$$

と計算されている．この問題では視聴率が変化し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っている．

宣伝活動を行って視聴率があ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 13. 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 よりどのぐ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

$$P\left(S \ge 300 + a\right) = 0.05$$

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

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

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

$$P\left(T \ge 1.65\right) = 0.05$$

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． □

注意 10.3. (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。これが**第一種の過誤**である.

仮説検定の例題を教科書からさらに二つあげる.

例題 14 (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 $(i = 1, 2, \ldots)$ を考える. 標本平均 $\bar{X}_n = \frac{X_1+\cdots+X_n}{n}$ 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

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

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

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. □

注意 10.4. 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 15. 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 $X_1, \ldots, X_n$ に対して標本平均 $\bar{X}_n = \frac{X_1+\cdots X_n}{n}$ 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 16 (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

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

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

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

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． $\square$