2020.5.24(日)

情報統計力学 学習6

5月26日(火) 午後1時ー2時半

大数の法則の解説の続きと、中心極限定理とその理解に必要な解説。

1　大数の法則（続き）

前回、以下の**大数の弱法則**を証明した。

1.1　大数の弱法則

$\{X_i\}$ を互いに独立で、同じ分布を持つ確率変数とする。これを独立同分布を持つという。(i. i. d, independent and identically distributed)。また、$E((X_i)^2) = E((X_1)^2) < \infty$ とする。このとき、$S_n = X_1 + X_2 + \cdots + X_n$ とすると、任意の $\epsilon > 0$ に対して以下が成り立つ。

$$\lim_{n\to\infty} P(|\frac{S_n}{n} - E(X_1)| \geq \epsilon) = 0. \tag{1}$$

このような収束の仕方は、以下で定義する**確率収束**である。

定義 確率収束
確率変数列 $Y_1, Y_2, \cdots$ と確率変数 X について、任意の $\epsilon > 0$ に対して以下が成り立つとき、$\{Y_n\}$ は X に**確率収束する**という。

$$\lim_{n\to\infty} P(|Y_n - X| \geq \epsilon) = 0 \tag{2}$$

つまり、大数の弱法則は、$\frac{S_n}{n}$ が、$E(X_1)$ に確率収束することを言っている。

実は、もっと強いことが言える。

1.2　大数の強法則

$\{X_i\}$ を互いに独立で、同じ分布を持つ確率変数とする。また、$E((X_1)^4) < \infty$ とする。このとき、$S_n = X_1 + X_2 + \cdots + X_n$ とすると、以下が成り立つ。

$$P(\lim_{n\to\infty} \frac{S_n}{n} = E(X_1)) = 1. \tag{3}$$

証明は、テクニカルな予備知識を要するので省略する。以下に注意。
前回、シュワルツの不等式より、$E[|X_1|] \leq \sqrt{E[(X_1)^2]}$ を示した。更に、シュワルツの不等式で、$Y = 1, g_1(X) = X^2, g_2(Y) = 1$ とすると、$E[(X_1)^2)] \leq \sqrt{E[(X_1)^4]}$ となる。従って、

$$E[|X_1|] \leq \sqrt{E[(X_1)^2]} \leq \left(E[(X_1)^4]\right)^{1/4} \tag{4}$$

であるから、$E[(X_1)^4]$ が有限なので、$E[|X_1|], E[(X_1)^2]$ も有限となる。

これは、観測した平均値が真の平均値に収束することが確率 1 で起こることを意味している。つまり、ほとんど全ての観測において、観測した平均値が真の平均値に収束する。この収束は、**概収束**と呼ばれる。

定義 概収束
確率変数列 $Y_1, Y_2, \cdots$ と確率変数 X について以下が成り立つとき、$\{Y_n\}$ は X に**概収束する**といい、$Y_n \to X$ a.s. と書く。(a.s. は、allmost surely の略)。

$$P(\lim_{n\to\infty} Y_n = X) = 1. \tag{5}$$

つまり、大数の強法則は、$\frac{S_n}{n}$ が、$E(X_1)$ に概収束することを言っている。
実は、大数の強法則、弱法則ともに、$E[|X_1|] < \infty$ の仮定のみで成立する。

問題
確率変数列 $Y_1, Y_2, \cdots$ が確率変数 X に概収束するなら、確率収束することを示せ。

次に、確率変数列の和の分布が、ある特定の条件の下で正規分布に収束することを示そう。これは中心極限定理と呼ばれる。そのために、いくつか準備を行う。

2 中心極限定理

2.1 特性関数

確率変数 X に対して、次の t の関数を X の**特性関数**という。

$$\varphi(t) \equiv \varphi_X(t) = E(e^{itX}),\ (-\infty < t < \infty) \tag{6}$$

i は虚数単位。
離散型確率変数の場合。確率関数を $p_X(x)$ とする。

$$\varphi_X(t) = \sum_{k=1}^{\infty} e^{itx_k} p_X(x_k). \tag{7}$$

連続型確率変数の場合。確率密度関数を $f_X(x)$ とする。

$$\varphi_X(t) = \int_{-\infty}^{\infty} e^{itx} f_X(x) dx. \tag{8}$$

これは、関数 $f_X(x)$ のフーリエ変換である。

以下のように、特性関数は常に存在する。
離散型確率変数の場合。

$$|\varphi_X(t)| \le \sum_{k=1}^{\infty} |e^{itx_k} p_X(x_k)| = \sum_{k=1}^{\infty} p_X(x_k) = 1. \tag{9}$$

従って、特性関数は t の連続関数の項からなる級数であるが、それが $\mathbb{R}$ で一様絶対収束するから、特性関数は $\mathbb{R}$ で存在し連続となる。

連続型確率変数の場合。

$$|\varphi_X(t)| \le \int_{-\infty}^{\infty} |e^{itx} f_X(x)| dx = \int_{-\infty}^{\infty} f_X(x) dx = 1. \tag{10}$$

従って、積分で定義される特性関数は有限である。

主な性質。

1. $|\varphi_X(t)| \le 1,\ \ \varphi_X(0) = 1.$
2. $\varphi_X(-t) = \overline{\varphi_X(t)}.\ \ \overline{\varphi_X(t)}$ は $\varphi_X(t)$ の複素共役。
3. $\varphi_X(t)$ は t に関して連続。

以下を示せ。

問題　確率変数 X を考える。
(1) 定数 $a, b \in \mathbb{R}$ に対して、$Y = aX + b$ とすると、$\varphi_Y(t) = e^{itb}\varphi_X(at)$.
(2) $\varphi_X(-t) = \varphi_{-X}(t)$.

問題　互いに独立な確率変数 X, Y の特性関数をそれぞれ $\varphi_X(t), \varphi_Y(t)$ とするとき、確率変数 $Z = X + Y$ の特性関数は

$$\varphi_{X+Y}(t) = \varphi_X(t)\varphi_Y(t),\ (-\infty < t < \infty) \tag{11}$$

となる。

これは次のように一般化される。

問題　互いに独立な確率変数 $X_1, X_2, \cdots, X_n$ の特性関数を $\varphi_{X_1}(t), \varphi_{X_2}(t), \cdots, \varphi_{X_n}(t)$ とするとき、確率変数 $Z = X_1 + X_2 + \cdots + X_n$ の特性関数は

$$\varphi_{X_1+X_2+\cdots+X_n}(t) = \varphi_{X_1}(t)\varphi_{X_2}(t)\cdots\varphi_{X_n}(t),\ (-\infty < t < \infty) \tag{12}$$

となる。

確率変数のモーメント (積率)
確率変数 X の k 次モーメント (積率) とは、$E(X^k)$ のことである。特に、1 次モーメントは平均値 $E(X) = \mu$ で、重心という意味がある。$E[(X-a)^k]$ を a のまわりの k 次モーメントという。分散 $V(X) = E[(X-\mu)^2]$ は、平均 μ のまわりの 2 次モーメントである。

モーメントは、特性関数を用いて以下のように計算できる。

$$E(X^k) = (-i)^k \varphi_X^{(k)}(0),\ (k = 0, 1, 2, \cdots). \tag{13}$$

ここで、$\varphi_X^{(k)}(t)$ は、$\varphi_X(t)$ の t での k 階導関数である。
特に、平均、分散は、$E(X) = -i\varphi'_X(0), V(X) = -\varphi''_X(0) + \{\varphi'_X(0)\}^2$ となる。
証明

$$\varphi'_X(t) = \frac{d}{dt}\varphi_X(t) = iE[Xe^{itX}]$$

で、$t = 0$ と置くと、$E(X) = -i\varphi'_X(0)$。微分を繰り返すと、そのつど、iX が掛けられるので、

$$\varphi_X^{(k)}(t) = i^k E[X^k e^{itX}].$$

$t = 0$ と置くことにより、$E(X^k) = (-i)^k \varphi_X^{(k)}(0)$。分散は、

$$V(X) = E(X^2) - E(X)^2 = -\varphi''_X(0) + \{\varphi'_X(0)\}^2.$$

以下に示すように、特性関数は、分布を特徴づける特性量の一つで、分布と１対１対応する。

2.1.1 レヴィ(P. Lévy) の反転公式

X は連続型確率変数でその特性関数を $\varphi_X(t)$ とする。このとき、以下が成り立つ。

$$P(a < X \le b) = \frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ita} - e^{-itb}}{it}\varphi_X(t)dt \tag{14}$$

但し、$a < b$ とする。

証明 密度関数を $f_X(x)$ とすると、

$$\begin{aligned}
&\frac{1}{2\pi}\int_{-T}^{T}\frac{e^{-ita} - e^{-itb}}{it}\varphi_X(t)dt = \frac{1}{2\pi}\int_{-T}^{T}dt\int_a^b e^{-itu}du\int_{-\infty}^{\infty}e^{itx}f_X(x)dx \\
&= \frac{1}{2\pi}\int_{-\infty}^{\infty}dxf_X(x)\int_a^b du\int_{-T}^{T}dte^{it(x-u)} = \frac{1}{2\pi}\int_{-\infty}^{\infty}dxf_X(x)\int_a^b du\frac{e^{iT(x-u)} - e^{-iT(x-u)}}{i(x-u)} \\
&= \frac{1}{\pi}\int_{-\infty}^{\infty}dxf_X(x)\int_a^b du\frac{\sin\Big(T(x-u)\Big)}{(x-u)} = \frac{1}{\pi}\int_{-\infty}^{\infty}dxf_X(x)\int_{T(x-b)}^{T(x-a)}dv\frac{\sin v}{v}
\end{aligned}$$

ここで、$v = T(x-u)$ とした。また、次式が成立。

$$\lim_{T\to\infty}\int_{T(x-b)}^{T(x-a)}dv\frac{\sin v}{v} = \begin{cases} 0 & (x < a \quad \text{または} \quad x > b) \\ \int_{-\infty}^{\infty}dv\frac{\sin v}{v} = \pi & (a < x < b). \end{cases}$$

従って、

$$\frac{1}{2\pi}\lim_{T\to\infty}\int_{-T}^{T}\frac{e^{-ita} - e^{-itb}}{it}\varphi_X(t)dt = \int_a^b dxf_X(x) = P(a < X \le b).$$

但し、積分や極限の順序は入れ替え可能であることが示せる。

この定理より、特性関数と確率は１対１に対応することが分かる。

次に、独立同分布の場合に、ある条件下である極限が正規分布になることを示す。
特に、２項分布について、ある極限が正規分布になることを示す。

２項分布

コイン投げで表か裏がでるように、１回の試行で２通りのいずれかであるものを考え、そろぞれの確率を $0 < p, q < 1$ で表す。勿論、$p + q = 1$。このような試行を**ベルヌーイ試行**とか**２項試行**という。ベルヌーイ試行を n 回繰り返して、そのうち表が出る回数を X とすると、X は離散型確率変数となり、その確率関数は、

$$p_X(x) = P(X = x) = {}_nC_x p^x q^{n-x}, \ (x = 0, 1, 2, \cdots, n) \tag{15}$$

となる。

確率変数 X の確率関数が上式で与えられているとき、この離散型分布を　**２項分布**といい、記号で、$B(n, p)$ と書く。確率変数 X が２項分布 $B(n, p)$ に従うことを、$X \sim B(n, p)$ と書く。特に $n = 1$ の２項分布 $B(1, p)$ を　**ベルヌーイ分布**といい、記号で、$\mathrm{Ber}(p)$ と表す。

問題

確率変数 X が２項分布 $B(n, p)$ に従うとき、以下を示せ。
(1) X の期待値は、$E(X) = np.$
(2) X の分散は、$V(X) = npq.$
(3) 特性関数は、$\varphi_X(t) = (e^{it}p + 1 - p)^n.$

ここで、確率変数の**分布収束**、あるいは、**法則収束**について説明する。

$C_b(\mathbb{R})$ を $\mathbb{R}$ 上で定義された有界な連続関数全体がつくるベクトル空間とする。確率変数列 $\{X_n\}(n \in \mathbb{N})$、および確率変数 X について、任意の $f \in C_b(\mathbb{R})$ に対して、

$$E(f(X_n)) \to E(f(X)) \ (n \to \infty)$$

が成り立つとき、確率変数 X は、X に**分布収束する**、あるいは、**法則収束する**といい、

$$X_n \xrightarrow{d} X \ (n \to \infty) \quad \text{あるいは} \ X_n \xrightarrow{\mathcal{L}} X \ (n \to \infty)$$

と表す。このとき、X_n, X の分布関数 $F_{X_n}(x), F_X(x)$ について、$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$ が、全ての連続点 x で成立することが示せる。なお、一般に x はベクトル。

2.1.2 リンデベルグ=レヴィ(Lindeberg-Lévy) の中心極限定理

$X_1, X_2, \cdots, X_n$ は独立同分布 (i.i.d.) の確率変数列で、$E(X_1) = \mu, V(X_1) = \sigma^2 < \infty$ を満たすとする。このとき、$n \to \infty$ の下で、

$$\sqrt{n}\left(\frac{1}{n}\sum_{k=1}^{n} X_k - \mu\right) \xrightarrow{d} Z \sim N(0, \sigma^2). \tag{16}$$

証明

$Y = X_k - \mu$ の特性関数を φ とし、$Z_n = (\sum_{k=1}^n X_k - n\mu)/\sqrt{n}$ の特性関数を φ_n とすると、$n \to \infty$ のとき、

$$\begin{aligned}
\varphi_n(t) &= E\left\{\exp\left(it\frac{\sum_{k=1}^n X_k - n\mu}{\sqrt{n}}\right)\right\} = \left(\varphi(\frac{t}{\sqrt{n}})\right)^n = \left(E(e^{i\frac{t}{\sqrt{n}}}(X_k - \mu))\right)^n \\
&= \left(E\left[1 + i\frac{t}{\sqrt{n}}(X_k - \mu) - \frac{t^2}{2n}(X_k - \mu)^2 + o(\frac{t^2}{n})\right]\right)^n \\
&= \left(1 - \frac{t^2}{2n}\sigma^2 + o(\frac{t^2}{n}))\right)^n \to \exp(-\frac{1}{2}t^2\sigma^2)
\end{aligned}$$

ここで、$o(x)$ は、$x \to 0$ のとき、0 になる量を表す。最後の式は、正規分布 $N(0, \sigma^2)$ の特性関数である。特性関数と分布は 1 対 1 対応するから、$Z \quad \sim N(0, \sigma^2)$ となる。

問題

正規分布 $N(0, \sigma^2)$ の特性関数が、$\exp(-\frac{1}{2}t^2\sigma^2)$ となることを示せ。

2.1.3 ド・モアブル=ラプラス (de Moivre-Laplace) の中心極限定理

$X_1, X_2, \cdots, X_n$ は独立同分布 (i.i.d.) の確率変数列で、パラメータ p のベルヌーイ分布に従っているとする。$S_n = X_1 + X_2 + \cdots + X_n$ に対して、

$$\frac{S_n - np}{\sqrt{np(1-p)}} \xrightarrow{d} Z \sim N(0, 1) \ (n \to \infty) \tag{17}$$

が成り立つ。

証明

$\mu = E(X_k) = p, \sigma^2 = V(X_k) = p(1-p)$ であるので、リンデベルグ=レヴィ (Lindeberg-Lévy) の中心極限定理の証明で、$Y = (X_k - \mu)/\sigma, Z_n = (\sum_{k=1}^{n} X_k - n\mu)/(\sqrt{n}\sigma)$ とおくことにより従う。