

数理統計：第1回講義 (4月9日(水))

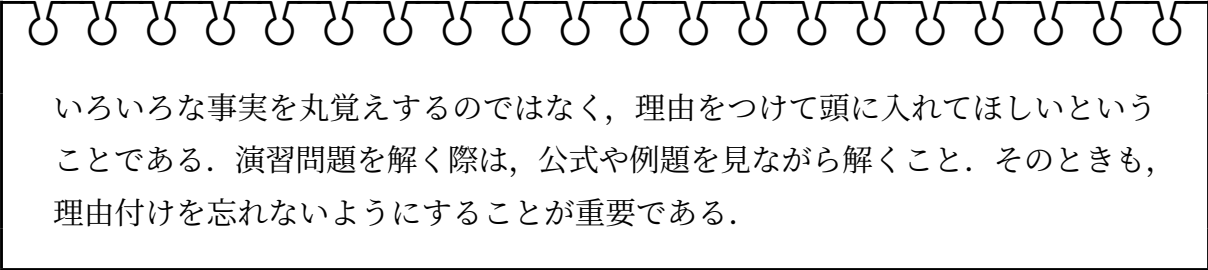
このノートを読んで、演習問題を解くことで、2025年度前期(水曜1限)「数理統計」(担当：松本)の第1回講義とする。教科書

「確率・統計の基礎」(松本裕行著, 学術図書)

を, 合わせて参照してほしい。

講義の主な目的は, 「推定」と「検定(統計的仮説検定)」の初歩を学習することである。それぞれ考え方は一つで, 必要な数学の準備をして推定なら推定の議論をすれば結論が得られる。このことを, 自分でノートを作って実感してほしい。

私から特にお願いしたいのは,



いろいろな事実を丸覚えするのではなく, 理由をつけて頭に入れてほしいということである。演習問題を解く際は, 公式や例題を見ながら解くこと。そのときも, 理由付けを忘れないようにすることが重要である。

演習問題, 問題は必ず, 教科書, このノートを見ながらで全くかまわないので, 自分で解くこと。そして, 必要な数学の準備と推定, 検定の議論を, 書くことを忘れないように。

第0章 確率・統計の復習

推定, 検定の学習を進めるためには,

- (i) 正規分布の基本的な性質
- (ii) ド・モワブル-ラプラスの定理 (二項分布の正規分布による近似)

が不可欠である.

ここでは, 2年次の「確率・統計」で学習したこれらのことを復習し, 確認のためレポートを課します.

1 連続型確率変数, 連続分布

一様分布のように, 確率変数 X の取り得る値の集合が実数全体またはその部分区間であるとき, X を**連続型確率変数**という. ここでは, $c \leq X \leq d$ の確率が, $\mathbf{R}$ 上の非負関数 f を用いて

$$P(c \leq X \leq d) = \int_c^d f(x)dx$$

によって積分で与えられる場合を考える. f を**確率密度**または**確率密度関数**と呼ぶ. 確率変数が値を取らないような x においては $f(x) = 0$ として確率密度 f の定義域を明示しないのが通例である. 確率密度から定まる $\mathbf{R}$ 上の値の散らばりを**連続分布**という.

定義 1. (1) 確率変数 X を確率密度 f をもつ連続型確率変数のとき,

$$\int_{-\infty}^{\infty} xf(x)dx$$

を X の**期待値**または**平均**と呼び, $E[X]$ または m と表す.

(2) $(X - m)^2$ の平均

$$E[(X - m)^2] = \int_{-\infty}^{\infty} (x - m)^2 f(x)dx$$

を X の**分散**と呼び, $V[X]$ または σ^2 ($\sigma > 0$) と表す. $\sigma = \sqrt{V[X]}$ を X の**標準偏差**と呼ぶ.

注意 1.1. (1) 分散について,

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

が成り立つ. これは,

$$\begin{aligned} V[X] &= \int_{-\infty}^{\infty} (x^2 - 2mx + m^2)f(x)dx \\ &= \int_{-\infty}^{\infty} x^2 f(x)dx - 2m \int_{-\infty}^{\infty} x f(x)dx + m^2 \int_{-\infty}^{\infty} f(x)dx \\ &= E[X^2] - 2m \times m + m^2 \end{aligned}$$

から分かる.

(2) α, β を定数とすると, $E[\alpha X + \beta] = \alpha E[X] + \beta$, $V[\alpha X + \beta] = \alpha^2 V[X]$ が成り立つ.

平均に対する主張は意味を考えれば明らかであろう. 分散に対する主張は, 次のようにすれば分かる:

$$\begin{aligned} V[\alpha X + \beta] &= E\left[\left\{(\alpha X + \beta) - (\alpha m + \beta)\right\}^2\right] = E[\alpha^2(X - m)^2] \\ &= \int_{-\infty}^{\infty} \alpha^2(x - m)^2 f(x)dx = \alpha^2 \int_{-\infty}^{\infty} (x - m)^2 f(x)dx = \alpha^2 V[X]. \end{aligned}$$

2 正規分布

定義 2. 実数 m, σ ($\sigma > 0$) に対して, 確率変数 X の確率密度 f が

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

によって与えられるとき, X は**正規分布** $N(m, \sigma)$ に従うという. とくに, $N(0, 1)$ を**標準正規分布**と呼ぶ.

命題 2.1. 確率変数 X が正規分布 $N(m, \sigma)$ に従うとき, $E[X] = m, V[X] = \sigma^2$ が成り立つ.

命題 2.2. 確率変数 X が正規分布 $N(m, \sigma)$ に従うとき, 実数 p, q に対して $Y = pX + q$ は $N(pm + q, p^2\sigma^2)$ に従う.

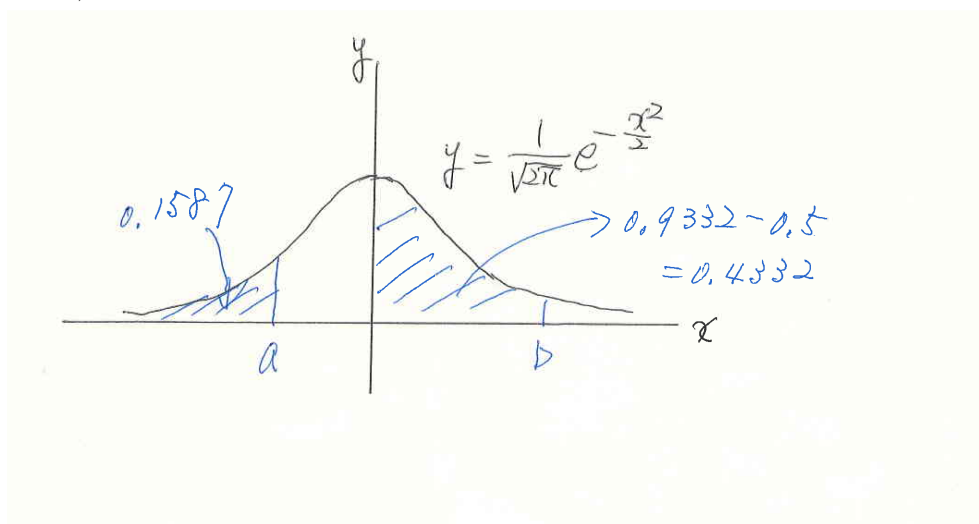
とくに $T = \frac{X - m}{\sigma}$ は標準正規分布に従う. T を X の**正規化**または**標準化**と呼ぶ.

例題 2.3. 確率変数 X が正規分布 $N(m, \sigma)$ に従うし, X の正規化を T とする.

(1) $P(T \leq a) = 0.1587, P(T \leq b) = 0.9332$ となる a, b を正規分布表から求めよ.

(2) $P(X \leq 6) = 0.1587, P(T \leq 16) = 0.9332$ であるとき, X の平均と分散を求めよ.

解答. . (1) $a < 0, b > 0$ であることを理解し, 次を念頭に正規分布表を見ると, $a = -1, b = 1.5$ を得る.



(2) 条件を

$$P(X \leq 6) = P\left(\frac{X - m}{\sigma} \leq \frac{6 - m}{\sigma}\right) = 0.1587,$$

$$P(X \leq 16) = P\left(\frac{X - m}{\sigma} \leq \frac{16 - m}{\sigma}\right) = 0.9332$$

と書き直すと, (1) より m, σ に対する連立方程式

$$\frac{6 - m}{\sigma} = -1, \quad \frac{16 - m}{\sigma} = 1.5$$

が得られる. したがって, $m = 10, \sigma = 4$ となり, 平均は 10, 分散は $4^2 = 16$ となる. $\square$

命題 2.4. 確率変数 X が正規分布 $N(m, \sigma)$ に従うとき, 積率母関数は $M_X(t) = e^{mt + \frac{1}{2}\sigma^2 t^2}$ である.

証明. 指数を平方完成することにより, 積分

$$\int_{-\infty}^{\infty} e^{tx} \frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{(x-m)^2}{2\sigma^2}} dx$$

を計算しても結論を得る.

ここでは, 正規化をして考える. $T = \frac{X - m}{\sigma}$ は $N(0, 1)$ に従うので, その積率母関数は

$$M_T(t) = \int_{-\infty}^{\infty} e^{tu} \frac{1}{\sqrt{2\pi}} e^{-\frac{u^2}{2}} du = \int_{-\infty}^{\infty} \frac{1}{\sqrt{2\pi}} e^{-\frac{(u-t)^2}{2}} du \times e^{\frac{t^2}{2}} = e^{\frac{t^2}{2}}$$

となる. よって,

$$M_X(t) = E[e^{t(m+\sigma T)}] = e^{tm} E[e^{(\sigma t)T}] = e^{tm} e^{\frac{(\sigma t)^2}{2}}$$

となり, 結論を得る. $\square$

統計においては、それぞれが正規分布 $N(m, \sigma^2)$ に従う独立確率変数列 $X_1, X_2, \dots, X_n$ から話が始まることが多い。その和が重要である。

まず、標準正規分布から始める。

命題 2.5. $T_1, T_2, \dots, T_n$ を、それぞれ標準正規分布 $N(0, 1)$ に従う独立確率変数列とするとき、 $T = T_1 + \dots + T_n$ は正規分布 $N(0, n)$ に従い、 $\bar{T}_n = \frac{T_1 + \dots + T_n}{n}$ は $N(0, \frac{1}{n})$ に従う。

証明. T については、積率母関数を計算する：

$$\begin{aligned} E[e^{tT}] &= E[e^{t(T_1 + \dots + T_n)}] = E[e^{tT_1} \times \dots \times e^{tT_n}] \\ &= E[e^{tT_1}] \times \dots \times E[e^{tT_n}] = (e^{\frac{t^2}{2}})^n = e^{\frac{nt^2}{2}}. \end{aligned}$$

$\bar{T}_n$ については、命題 2.2 を用いればよい。 □

T の分散が n であることが重要である。直感的には、 $T_1, \dots, T_n$ は正の値と負の値が入り交じるのが普通で、これらが相殺して散らばりが小さくなるということである。または、 T の平均が 0 であることは容易に分かるので、分散は T^2 の平均と等しく、これは、 T_i と T_j の独立性から

$$\begin{aligned} V[X] &= E[T^2] = E[(T_1 + \dots + T_n)^2] = E\left[\sum_{i=1}^n T_i^2 + 2 \sum_{i < j} T_i T_j\right] \\ &= \sum_{i=1}^n E[T_i^2] + 2 \sum_{i < j} E[T_i T_j] = n + 2 \sum_{i < j} E[T_i] E[T_j] = n \end{aligned}$$

となることから分かる。

定理 2.6. $X_1, X_2, \dots, X_n$ を、それぞれ標準正規分布 $N(m, \sigma^2)$ に従う独立確率変数列とするとき、 $S_n = X_1 + \dots + X_n$ は正規分布 $N(nm, n\sigma^2)$ に従い、 $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ は $N(m, \frac{\sigma^2}{n})$ に従う。

証明は、 $T_i = \frac{X_i - m}{\sigma}$ が標準正規分布に従うことから、命題 2.2 を用いればよい。

3 ド・モワブル-ラプラスの定理

$X_1, X_2, \dots, X_n$ を同じ確率分布に従う独立確率変数列とすると、それぞれの分布は正規分布でなくても、 $\bar{X}_n = \frac{X_1 + \dots + X_n}{n}$ を正規化した $\frac{\bar{X}_n - m}{\sqrt{\frac{\sigma^2}{n}}}$ の確率分布が $n \rightarrow \infty$ のとき標準正規分布 $N(0, 1)$ に収束する。これが**中心極限定理**である。

同じ試行を独立に繰り返すとき、ある確率 p の事象 A を考えて

$$X_i = \begin{cases} 1, & i \text{ 回目に } A \text{ が起きる,} \\ 0, & i \text{ 回目に } A \text{ が起きない} \end{cases}$$

とおくと、 $S_n = X_1 + \dots + X_n$ は n 回中 A の起きた回数を表し、 S_n は二項分布 $B(n, p)$ に従う。

テレビ番組の視聴率調査や世論調査を考えるとときには、次が基本的な道具となる。

ド・モワブル-ラプラスの定理

定理 3.1. S_n を二項分布 $B(n, p)$ に従う確率変数列とすると、 $T_n = \frac{S_n - nm}{\sqrt{npq}}$ の確率分布は、 $n \rightarrow \infty$ のとき、標準正規分布 $N(0, 1)$ に収束する。とくに、次が成り立つ：

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

例 3.2. サイコロを 50 回投げるとき、1 の出る回数を S_{50} とする。 $6 \leq S_{50} \leq 10$ の確率を、(i) 二項分布表を使って (ii) ド・モワブル-ラプラスの定理を使って求める。

(i) 二項分布に従うことを用いる場合は、次のように和を計算すればよい：

$$\begin{aligned} P(6 \leq S_{50} \leq 10) &= P(S_{50} = 6) + P(S_{50} = 7) + \dots + P(S_{50} = 10) \\ &\doteq 0.112 + 0.140 + 0.151 + 0.141 + 0.116 \doteq 0.66. \end{aligned}$$

(ii) ド・モワブル-ラプラスの定理を用いる場合は、**半目の補正**を行って、

$$P(6 \leq S_{50} \leq 10) = P(5.5 \leq S_{50} \leq 10.5) = P\left(\frac{5.5 - 50 \cdot \frac{1}{6}}{\sqrt{50 \cdot \frac{1}{6} \cdot \frac{5}{6}}} \leq T_n \leq \frac{10.5 - 50 \cdot \frac{1}{6}}{\sqrt{50 \cdot \frac{1}{6} \cdot \frac{5}{6}}}\right)$$

と変形する。すると、正規分布表から次を得る：

$$P(6 \leq S_{50} \leq 10) \doteq P(-1.07 \leq T_n \leq 0.82) \doteq 0.36 + 0.29 = 0.65.$$