数理情報学９ (数理統計学)　1．確率論の復習

解説

確率論の基礎事項について復習する．

1　1次元の確率変数

確率変数とはある事象の「起こりやすさ」の程度を表す確率を伴っている変数である．以下，離散値をとる場合と連続値をとる場合に分けて解説する．

1.1　離散値をとる確率変数

確率変数 X の取り得る値を $\{x_1, x_2, \ldots, x_n\}$ とする．X が x_i をとる確率を $P(X = x_i)$, $i = 1, \ldots n$ と書く．確率の定義から

$$P(X = x_i) \geq 0, \quad \sum_{i=1}^{n} P(X = x_i) = 1$$

を満たす．

例 **1** (偏りのあるサイコロ). 偏りのあるサイコロを振ったときに出る目を表す確率変数を X とする．X の値とそれに伴う確率を次のように記述する：

$$\begin{aligned} &P(X=1) = \frac{1}{2}, \quad P(X=2) = \frac{1}{4}, \quad P(X=3) = \frac{1}{16}, \\ &P(X=4) = \frac{1}{16}, \quad P(X=5) = \frac{1}{16}, \quad P(X=6) = \frac{1}{16}. \end{aligned}$$

□

期待値: 確率変数 X は $\{x_1, x_2, \ldots, x_n\}$ のうちのどれかの値を取る．その確率を $P(X = x_i)$, $i = 1, 2, \ldots, n$ とする．X の平均 $E[X]$ は以下のように定義される：

$$E[X] = \sum_{i=1}^{n} x_i P(X = x_i).$$

期待値とは，直感的には確率変数が取り得る値の真ん中という意味がある．

例 **2.** 例 *1* の確率変数の平均：

$$E[X] = 1 \times \frac{1}{2} + 2 \times \frac{1}{4} + 3 \times \frac{1}{16} + 4 \times \frac{1}{16} + 5 \times \frac{1}{16} + 6 \times \frac{1}{16} = \frac{17}{8} = 2.125$$

□

分散： 確率変数 X の分散 $V[X]$ は以下のように定義される．

$$V[X] = \sum_{i=1}^{n} \bigl(x_i - E[X]\bigr)^2 P(X = x_i)$$

分散とは，確率変数の値が期待値のまわりにどの程度ばらつくかを測るための指標である．

例 **3.** 例 *1* の確率変数の分散の計算：

$$V[X] = (1 - 2.125)^2 \times \frac{1}{2} + \cdots + (6 - 2.125)^2 \times \frac{1}{16} = \frac{151}{64} = 2.359375$$

□

1.2 連続値をとる確率変数

確率変数 X の取り得る値の集合が実数全体 (または実数上の適当な区間) の場合を考える．

$$
\begin{gathered}
\text{"}X \text{ が，確率密度関数 } f \text{ をもつ分布にしたがう"} \\
\Longleftrightarrow \text{任意の } A \subset \mathbb{R} \text{ について } P(X \in A) = \int_A f(x)dx
\end{gathered}
$$

ここで $f(x)$ は任意の $x \in \mathbb{R}$ に対して $f(x) \geq 0$ であり $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$ を満たす関数．$f(x)$ のことを確率変数 X の **(確率)** 密度関数とよぶ．

区間 A として $A = [x, x+dx]$ (dx は微少量) とすると

$$
P(X \in [x, x+dx]) \approx f(x)dx
$$

となる．X の期待値 $E[X]$ と分散 $V[X]$ は離散の場合と同じように定義される (ただし和が積分に変わる)：

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

例 **4.** 確率変数 X は区間 $[-1, 2]$ 上の一様分布にしたがうとする．一様分布の密度関数は以下のように書ける：

$$
f(x) = \begin{cases} 1/3 & x \in [-1, 2] \\ 0 & x \notin [-1, 2] \end{cases}
$$

X が区間 $[0, 1]$ に値をとる確率は

$$
P(X \in [0, 1]) = \int_0^1 f(x)dx = \frac{1}{3}
$$

となる．また X の期待値と分散は次のようになる：

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} xf(x)dx = \int_{-1}^{2} x\frac{1}{3}dx = \frac{1}{2} \\
V[X] &= \int_{-\infty}^{\infty} \left(x - \frac{1}{2}\right)^2 f(x)dx = \int_{-1}^{2} \left(x - \frac{1}{2}\right)^2 \frac{1}{3}dx = \frac{3}{4}
\end{aligned}
$$

□

1.3 期待値と分散の性質

確率変数 X を 1 次変換した確率変数 $aX + b$ を考える．ここで a, b は定数とする．$aX + b$ の平均と分散は次のようになる：

$$
E[aX + b] = aE[X] + b \tag{1}
$$

$$
V[aX + b] = a^2 V[X] \tag{2}
$$

X が連続的な確率変数のときに確認する：

$$
\begin{aligned}
E[aX + b] &= \int (ax + b)f(x)dx = a\int xf(x)dx + b = aE[X] + b, \\
V[aX + b] &= \int (ax + b - aE[X] - b)^2 f(x)dx = \int (ax - aE[X])^2 f(x)dx \\
&= a^2 \int (x - E[X])^2 f(x)dx = a^2 V[X].
\end{aligned}
$$

期待値や分散を計算するとき，積率母関数が便利なことが多い．確率変数 X の積率母関数 $\phi_X(t)$ は

$$\phi_X(t) \overset{\text{定義}}{=} E[e^{tX}]$$

で与えられる．両辺を t で微分して (期待値と微分の入れ替えは出来ると仮定すると)

$$\frac{d^k}{dt^k}\phi_X(t) = E[X^k e^{tX}]$$

となる．したがって

$$\left.\frac{d^k}{dt^k}\phi_X(t)\right|_{t=0} = E[X^k]$$

となる．積率母関数から分散を次のように計算できる：

$$V[X] = E[X^2] - E[X]^2 = \left.\frac{d^2}{dt^2}\phi_X(t)\right|_{t=0} - \left(\left.\frac{d}{dt}\phi_X(t)\right|_{t=0}\right)^2.$$

確率変数 X の期待値を μ，分散を σ^2 とするとき次の不等式が成り立つ．

チェビシェフの不等式

$$P(|X-\mu| \geq \varepsilon) \leq \frac{\sigma^2}{\varepsilon^2}$$

チェビシェフの不等式を用いて，確率変数 X が期待値から離れた値をとる確率を定量的に評価できる．

Proof.

$$\sigma^2 = \int_{\mathbb{R}} (x-\mu)^2 f(x)dx \geq \int_{|x-\mu|\geq\varepsilon} (x-\mu)^2 f(x)dx \geq \int_{|x-\mu|\geq\varepsilon} \varepsilon^2 f(x)dx = \varepsilon^2 P(|X-\mu| \geq \varepsilon)$$

□

1.4 分布関数

確率変数 X が取り得る値の範囲は実数とする．離散の場合には, 確率変数が取る値の集合 $\{x_1, \ldots, x_n\}$ は実数の有限部分集合とする．このとき分布関数 $F(x)$ は

$$F(x) \overset{\text{定義}}{=} P(X \leq x)$$

で与えられる．分布関数は次の性質を満たす：

(1) $a \leq b \Longrightarrow F(a) \leq F(b)$. (2) $0 \leq F(x) \leq 1$. (3) $F(x)$ は右連続.

とくに 連続な確率変数の場合 には

$$F(x) = \int_{-\infty}^{x} f(z)dz$$

となるので，両辺を x で微分すると

$$\frac{d}{dx}F(x) = f(x)$$

となる．つまり分布関数の微分は密度関数に等しい．

1.5 確率変数の変換と密度関数の変換

確率変数 X の密度関数を $f(x)$ とする．関数 h で確率変数 X を $Y = h(X)$ と変換したとき Y の密度関数 $g(y)$ はどう書けるか，という問題について解説する．分布関数と密度関数の関係から $P(Y \le y)$ を y で微分すれば Y の密度関数が得られる．

関数 h が単調増加関数のときには

$$P(Y \le y) = P(h(X) \le y) = P(X \le h^{-1}(y)) = \int_{-\infty}^{h^{-1}(y)} f(x)dx.$$

となるので，これを y で微分して

$$g(y) = f(h^{-1}(y))(h^{-1})'(y), \quad (h^{-1})' \text{ は } h^{-1} \text{ の微分}$$

を得る．関数 h が単調減少関数のときにも同様に計算できる．まとめると

変数変換の公式

関数 h は微分可能な単調関数とする．

$$X \sim f(x),\ Y = h(X) \ \text{ のとき } Y \sim g(y) = f(h^{-1}(y)) \left|(h^{-1})'(y)\right|, \quad (h^{-1})' \text{ は } h^{-1} \text{ の微分}$$

例 **5.** $X \sim N(0,1)$ のとき $Y = X^3$ の密度関数を求める．$f(x) = \frac{1}{\sqrt{2\pi}}e^{-x^2/2}$，$h^{-1}(y) = y^{1/3}$ なので，

$$g(y) = \frac{1}{\sqrt{2\pi}}e^{-y^{2/3}/2} \cdot \frac{y^{-2/3}}{3}$$

となる．□

関数 h が単調関数でないときにも $g(y)$ が計算できる場合がある．

例 **6.** $X \sim N(0,1)$ のとき $Y = X^2$ の密度関数を求める．

$$P(Y \le y) = P(X^2 \le y) = \begin{cases} P(-\sqrt{y} \le X \le \sqrt{y}) & y \ge 0 \\ 0 & y < 0 \end{cases} = \begin{cases} \displaystyle\int_{-\sqrt{y}}^{\sqrt{y}} \phi(x; 0, 1)dx & y \ge 0 \\ 0 & y < 0 \end{cases}$$

これを y で微分して

$$g(y) = \begin{cases} \dfrac{1}{\sqrt{2\pi y}}e^{-y/2} & y \ge 0 \\ 0 & y < 0 \end{cases}$$

となる．□

2 多次元の確率変数

複数の確率変数 $X = (X_1, X_2, \ldots, X_k)$ を同時に考える．以下では簡単のため 2 次元確率変数について説明する．

2.1 多次元確率変数の密度関数

例 **7.** 2 次元確率変数を $X = (X_1, X_2)$ とする．例えば

X_1 : 名古屋の降雨量

X_2 : 東京の降雨量

とすると，X_1 と X_2 は完全に一致はしないが，なにか関係がありそうである．多次元の確率変数を考えるときには，相互の関係に関心がある場合が多い．□

確率変数が連続値をとる場合を解説する (離散値のときも同様).

同時密度関数： 確率変数 $X = (X_1, X_2)$ の同時密度関数 $f(x_1, x_2)$ によって，X に関する確率は以下のように書ける：

$$\begin{aligned} &P\big[X_1 \text{ が } A_1 \text{ に値をとり，かつ } X_2 \text{ が } A_2 \text{ に値をとる}\big] \\ = \ & P(X_1 \in A_1, X_2 \in A_2) \ = \ \int_{A_1 \times A_2} f(x_1, x_2) dx_1 dx_2. \end{aligned}$$

同時密度関数は次の性質を満たす：

$$f(x_1, x_2) \geq 0, \quad \int_{\mathbb{R}^2} f(x_1, x_2) dx_1 dx_2 = 1.$$

周辺密度関数： 確率変数 $X = (X_1, X_2)$ の同時密度関数が $f(x_1, x_2)$ で与えられるとする．このとき X_1 にだけ注目すると，X_1 の密度関数は以下の $f_1(x_1)$ で与えらえれる．

$$f_1(x_1) = \int_{\mathbb{R}} f(x_1, x_2) dx_2$$

$f_2(x_2)$ も同様に定義される．X_1 に関する確率を考えると

$$P(X_1 \in A_1) = P(X_1 \in A_1, X_2 \in \mathbb{R}) = \int_{A_1 \times \mathbb{R}} f(x_1, x_2) dx_1 dx_2 = \int_{A_1} f_1(x_1) dx_1$$

が成り立つ．

条件付き密度関数： $X = (X_1, X_2)$ の同時密度関数 $f(x_1, x_2)$ に対して，$X_1 \in [x_1, x_1 + dx_1]$ という条件のもとで $X_2 \in [x_2, x_2 + dx_2]$ となる確率は，おおよそ

$$\frac{f(x_1, x_2) dx_1 dx_2}{f_1(x_1) dx_1} = \frac{f(x_1, x_2)}{f_1(x_1)} dx_2$$

となる．したがって $X_1 \in [x_1, x_1 + dx_1]$ という条件のもとでの X_2 の密度関数 $f(x_2|x_1)$ は

$$f(x_2|x_1) = \frac{f(x_1, x_2)}{f_1(x_1)} = \frac{f(x_1, x_2)}{\displaystyle\int_{\mathbb{R}} f(x_1, x_2) dx_2}$$

となる．定義から

$$\int_{\mathbb{R}} f(x_2|x_1) dx_2 = 1$$

が成り立つ．条件付き確率の意味を説明しよう．確率密度関数が $f(x_1, x_2)$ であるような確率分布から，たくさんの観測値

$$(x_1(1), x_2(1)), (x_1(2), x_2(2)), \ldots, (x_1(m), x_2(m))$$

が得られているとする．このうち $x_1(i) \in [x_1, x_1 + dx_1]$ を満たすデータ $(x_1(i), x_2(i))$ だけを集めてきたとき，その中で $x_2(i) \in [x_2, x_2 + dx_2]$ となるものの割合が，おおよそ $f(x_2|x_1) dx_2$ で与えられる．

2.2　共分散と相関係数

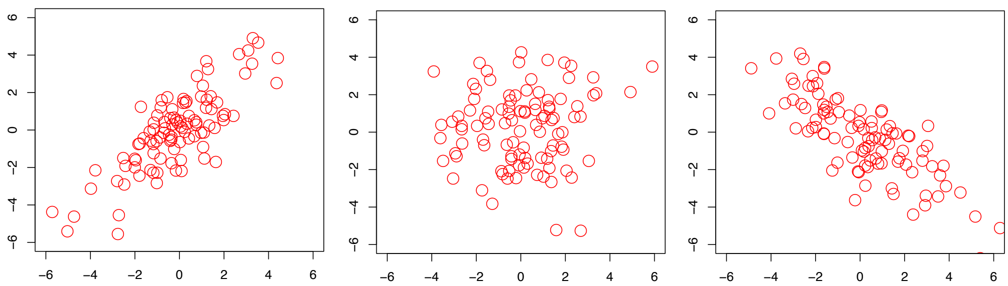
2 次元確率変数 (X, Y) の密度関数を $f(x, y)$ とする．2 つ確率変数 X, Y の関連の強さを測る共分散 $\mathrm{Cov}(X, Y)$ は

$$\mathrm{Cov}(X, Y) \overset{\text{定義}}{=} E[(X - E[X])(Y - E[Y])] = \int_{\mathbb{R}^2} (x - E[X])(y - E[Y]) f(x, y) dxdy$$

で定義される．ここで期待値 $E[X]$ は

$$E[X] = \int_{\mathbb{R}^2} x f(x, y) dxdy = \int_{\mathbb{R}} x f_X(x) dx, \quad (f_X(x) \text{ は } X \text{ の周辺密度関数})$$

で与えられる ($E[Y]$ も同様).



$\mathrm{Cov}(X, Y) > 0$　　$\mathrm{Cov}(X, Y) \cong 0$　　$\mathrm{Cov}(X, Y) < 0$

共分散について以下の公式が成り立つ：任意の確率変数 X, Y, Z と定数 a, b について

$$\begin{gathered} \mathrm{Cov}(X, X) = V[X], \\ \mathrm{Cov}(X + Y, Z) = \mathrm{Cov}(X, Z) + \mathrm{Cov}(Y, Z), \\ \mathrm{Cov}(aX, bY) = ab \times \mathrm{Cov}(X, Y). \end{gathered}$$

確率変数 X, Y の共分散を，それぞれの分散で規準化した値である相関係数は

$$\rho(X, Y) \overset{\text{定義}}{=} \frac{\mathrm{Cov}[X, Y]}{\sqrt{V[X]}\sqrt{V[Y]}}$$

で定義される．不等式 $-1 \leq \rho(X, Y) \leq 1$ が成り立つ．また次の等式が成り立つ：

$$\text{定数 } a, b, c, d \ (a, c > 0) \text{ に対して } \rho(aX + b, cY + d) = \rho(X, Y).$$

2.3　確率変数の独立性

2 つの確率変数 X, Y が互いに影響を及ぼし合うことはないとき，X と Y は独立という．たとえばサイコロを 2 回振るとき，1 回目の値 X と 2 回目の値 Y は互いに影響しないと考えてよいので独立と言える．独立性をきちんと定義すると以下のようになる．

確率変数 X, Y が離散値をとる場合：

$$X \text{ と } Y \text{ が独立} \overset{\text{def}}{\Longleftrightarrow} P(X = x_i, Y = y_j) = P(X = x_i)P(Y = y_i) \text{ が任意の } x_i, y_j \text{ で成立.}$$

確率変数 X, Y が連続値をとる場合：　(X, Y) の同時密度関数を $f(x, y)$ とする．また X, Y の周辺密度関数をそれぞれ $f_X(x), f_Y(y)$ とする．

$$X \text{ と } Y \text{ が独立} \overset{\text{def}}{\Longleftrightarrow} f(x, y) = f_X(x) f_Y(y) \text{ が任意の } x, y \text{ で成立.}$$

複数の確率変数 $X_1, X_2, \ldots, X_k$ の独立性も同様に定義される：

$$
\begin{aligned}
& X_1, X_2, \ldots, X_k \text{ はすべて独立} \\
\overset{\text{def}}{\Longleftrightarrow}\ & f(x_1, \ldots, x_k) = f_1(x_1) \cdots f_k(x_k) \\
\Longleftrightarrow\ & \text{任意の集合 } A_1, \ldots, A_k \text{ に対して} \\
& P(X_1 \in A_1, \ldots, X_k \in A_k) = P(X_1 \in A_1) \times \cdots \times P(X_k \in A_k)
\end{aligned}
$$

ここで f_i は X_i の周辺分布．

確率変数の独立性について重要事項をまとめておく．

重要事項

- X と Y が独立ならば

$$
\begin{aligned}
E[XY] &= E[X]E[Y] \\
V[X+Y] &= V[X] + V[Y]
\end{aligned}
$$

- X と Y との独立性とは関係なく，いつでも

$$
E[X+Y] = E[X] + E[Y]
$$

3 正規分布

統計学において重要な正規分布について解説する．連続的な値をとる確率変数 X の密度関数が

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

と書けるとき

「X は期待値が μ で分散が σ^2 の正規分布にしたがう」

と言い

$$
X \sim N(\mu, \sigma^2)
$$

と書く．とくに $N(0,1)$ の分布を標準正規分布という．$\phi(x; \mu, \sigma^2)$ の関数形を図 1 に示す．期待値と分散を計算すると

$$
\begin{aligned}
E[X] &= \int_{-\infty}^{\infty} x\phi(x; \mu, \sigma^2)dx = \mu, \\
V[X] &= \int_{-\infty}^{\infty} (x-\mu)^2 \phi(x; \mu, \sigma^2)dx = \sigma^2
\end{aligned}
$$

となっていることが確認できる．X の取る値が $\mathbb{R}$ の部分集合 A に含まれる確率は

$$
P(X \in A) = \int_A \phi(x; \mu, \sigma^2)dx
$$

で与えられる．$X \sim N(\mu, \sigma^2)$ のとき，X の値がどのあたりに散らばるのかを計算すると以下のようになる：

$$
\begin{aligned}
& P(\mu - \sigma \le X \le \mu + \sigma) \doteqdot 0.6827, \quad P(\mu - 2\sigma \le X \le \mu + 2\sigma) \doteqdot 0.9545, \\
& P(\mu - 3\sigma \le X \le \mu + 3\sigma) \doteqdot 0.9973
\end{aligned}
$$

したがって，X の実現値は $[\mu - 3\sigma, \mu + 3\sigma]$ の範囲にほぼ入ってくる．

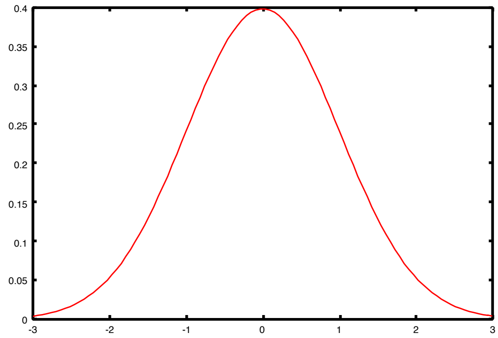


Figure 1: 正規分布の確率密度関数のプロット．$\mu = 0, \sigma^2 = 1$ としている．

4 確率変数の収束

確率変数の収束について解説する．とくに「大数の法則」と「中心極限定理」が重要である．確率変数の収束に関するこれらの定理から，観測データ数が多いときの統計量の振舞いを調べることができる．

4.1 確率収束と大数の法則

まず確率収束について説明する．

確率収束： $Z_n, n = 1, 2, 3, \ldots$ を確率変数列として $c \in \mathbb{R}$ を定数とする．以下の条件

$$\forall \varepsilon > 0 \quad \lim_{n\to\infty} P\left(|Z_n - c| > \varepsilon\right) = 0$$

が成り立つとき「Z_n は c に確率収束する」といい，

$$Z_n \xrightarrow{p} c \quad (n \to \infty)$$

と書く．

例 **8.** 確率変数列 Z_n を

$$Z_n = 1 + \frac{1}{n}X, \qquad X \sim N(0, 1)$$

とする．

$$P(|Z_n - 1| > \varepsilon) = P(|X| > n\varepsilon) = 1 - \int_{-n\varepsilon}^{n\varepsilon} \phi(x; 0, 1)dx \longrightarrow 0 \quad (n \to \infty).$$

したがって $Z_n \xrightarrow{p} 1$ である． □

次に大数の法則を説明する．確率変数 $X_1, X_2, \ldots, X_n$ はすべて独立で，同一の分布にしたがうとする．その期待値は $E[X_i] = \mu$，分散は $V[X_i] = \sigma^2$ とする．このとき

$$\text{標本平均: } \bar{X}_n = \frac{1}{n}(X_1 + \cdots + X_n)$$

の挙動について考える．まず $\bar{X}_n$ の期待値と分散を計算すると

$$E[\bar{X}_n] = \frac{1}{n}E[X_1 + \cdots + X_n] = \frac{1}{n}\{E[X_1] + \cdots + E[X_n]\} = \frac{1}{n} \cdot n\mu = \mu$$
$$V[\bar{X}_n] = \frac{1}{n^2}V[X_1 + \cdots + X_n] = \frac{1}{n^2}\big(V[X_1] + \cdots + V[X_n]\big) = \frac{1}{n^2} \cdot n\sigma^2 = \frac{1}{n}\sigma^2 \quad \text{(独立性を使う)}$$

となる．ここで $\bar{X}_n$ にチェビシェフの不等式を適用すると，任意の $\varepsilon > 0$ に対して

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

が成り立つ．したがって次の関係が成り立つ．

大数の法則

$X_1, X_2, \ldots, X_n$ が独立で同一の分布にしたがい，期待値 μ と分散が存在するとき

$$\forall\varepsilon > 0 \text{ に対して} \quad \lim_{n\to\infty} P(|\bar{X}_n - \mu| > \varepsilon) = 0 \quad \text{すなわち } \bar{X}_n \xrightarrow{p} \mu$$

が成り立つ．

大数の法則が意味することは「観測数が増えるにしたがって，標本平均が期待値から ε 以上ズレる確率は 0 に近づく」ということである．観測数が多ければ標本平均は期待値に非常に近い値をとるので，標本平均によって期待値を推定することができる．

4.2 法則収束と中心極限定理

正規分布は統計学によく出てくる重要な分布であるが，その主な理由は中心極限定理にある．まず法則収束について説明する．

法則収束： $F_n(x)$ を X_n の分布関数として，$F(x)$ を X の分布関数とする．このとき $F(x)$ の任意の連続点 x において

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

が成り立つとき「X_n は X(または F) に法則収束する」といい，

$$X_n \xrightarrow{d} X \quad \text{もしくは} \quad X_n \xrightarrow{d} F$$

と書く．分布収束とも言う．代表的な分布の場合には，F の代わりに分布を表す記号を使って

$$X_n \xrightarrow{d} N(0,1)$$

などと書く． □

中心極限定理を説明する．確率変数を $X_1, \ldots, X_n$ は独立に同一の分布にしたがうとする．期待値を $E[X_i] = \mu$，分散を $V[X_i] = \sigma^2$ とする．このとき確率変数 Y_n を

$$Y_n = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \frac{(X_i - \mu)}{\sigma} = \sqrt{n}\ \frac{\bar{X}_n - \mu}{\sigma}$$

とおく．$E[Y_n] = 0,\ V[Y_n] = 1$ になっている．Y_n の分布は $n \to \infty$ の極限で正規分布に近づく：

中心極限定理

$$\forall y \in \mathbb{R} \text{ に対して } \lim_{n\to\infty} P(Y_n \leq y) = \int_{-\infty}^{y} \phi(x; 0, 1) dx.$$

すなわち $Y_n \xrightarrow{d} N(0,1) \quad (n \to \infty)$

証明は参考文献などを参照のこと．確率変数が (期待値と分散は存在するような) どのような分布にしたがっていても，Y_n という確率変数を考えれば常に正規分布に法則収束していく．この性質はさまざまな統計量の性質を調べるのに非常に役に立つ．

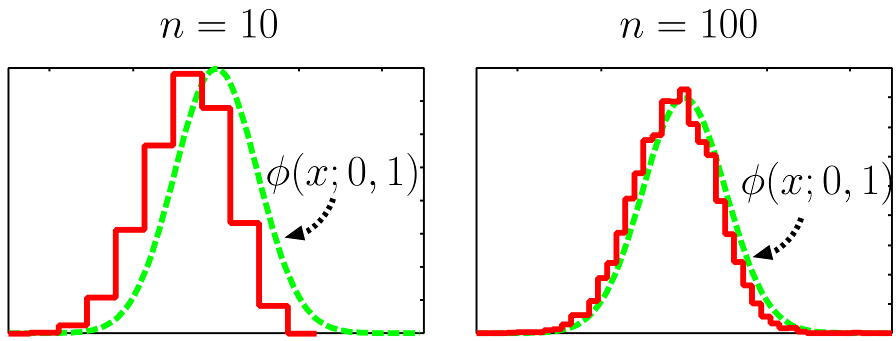


Figure 2: 中心極限定理を数値的に確かめる．左図：$n = 10$, 右図：$n = 100$.

例 **9.** 具体的に Y_n の振舞いを数値実験で眺めてみる．確率変数 $X_1 \ldots, X_n$ は 0 か 1 の値をとり，その確率は

$$P(X_i = 0) = 0.3, \quad P(X_i = 1) = 0.7$$

で与えられるとする．確率変数 X_i の期待値と分散はそれぞれ $E(X_i) = 0.7,\ \ V(X_i) = 0.21$ となる．このとき $n = 10$ の場合と $n = 100$ の場合に

$$Y_n = \sqrt{n}\frac{\bar{X}_n - \mu}{\sigma} = \sqrt{n}\frac{\bar{X}_n - 0.7}{\sqrt{0.21}}$$

がどのような確率分布にしたがうかを計算すると図 *4.2* のヒストグラムのようになる．n が大きくなるのにしたがって，標準正規分布の密度関数 $\phi(x;0,1)$ に近づいていく様子が分かる． □

4.3 スラツキーの公式

確率変数の収束に関するスラツキーの公式を以下に示す．

$$
\begin{array}{llcl}
(a) & X_n \xrightarrow{p} c\ (\text{定数}) & \Longleftrightarrow & X_n \xrightarrow{d} c\ (\text{定数}) \\
(b) & X_n \xrightarrow{p} c,\ \ f : c\ \text{で連続} & \Longrightarrow & f(X_n) \xrightarrow{p} f(c) \\
(c) & X_n \xrightarrow{d} X,\ \ f : \text{連続} & \Longrightarrow & f(X_n) \xrightarrow{d} f(X) \\
(d) & X_n \xrightarrow{d} X,\ Y_n \xrightarrow{p} c & \Longrightarrow & X_n + Y_n \xrightarrow{d} X + c,\ \ X_n Y_n \xrightarrow{d} cX
\end{array}
$$

通常の実数における収束と同じような関係 $(b), (c), (d)$ が成り立っている．これらの関係は，推定量の分布の漸近的な性質 (データ数が多いときの性質) を調べるときによく使われる．

演習問題

1. 確率変数 X は $0, 1, -1$ の値を以下の確率でとる：

$$P(X = -1) = p,\ \ P(X = 0) = p,\ \ P(X = 1) = 1 - 2p.$$

ただし $0 \le p \le 1/2$ とする．X の期待値と分散を求めよ．

2. 確率変数 X の期待値を μ，分散を σ^2 とする．

   (a) $2X + 1$ の期待値と分散を μ, σ^2 を用いて表せ．

   (b) $E[X(X - 1)]$ を μ, σ^2 を用いて表せ．

3. 確率変数 X の密度関数 $f(x)$ は，a を定数として

$$f(x) \ = \ \begin{cases} ax(1-x) & (0 \le x \le 1) \\ 0 & \text{その他} \end{cases}$$

で定義されている．

(a) 定数 a を求めよ．

(b) X の期待値と分散を求めよ．

4. 確率変数 X の密度関数 $f(x)$ を

$$f(x) \ = \ \begin{cases} x/2 & (0 \le x \le 2) \\ 0 & \text{その他} \end{cases}$$

とする．

(a) X の分布関数 $F(z) = P(X \le z)$ を求めよ

(b) $Y = 2X^2 + 1$ の分布関数と確率密度関数を求めよ．

5. X を正規分布 $N(\mu, 1)$ にしたがう確率変数とする．

(a) $Y = e^X$ の密度関数を計算せよ．

(b) $Y = |X|$ の密度関数を計算せよ．

6. 確率変数 U が区間 $(0,1)$ 上の一様分布にしたがうとする．このとき $X = -2\log U$ の密度関数を求めよ．

7. X と Y を確率変数とするとき

$$V[X+Y] = V[X] + V[Y] + 2\ \mathrm{Cov}[X,Y]$$

となることを証明せよ．

8. 確率変数 X, Y は互いに独立として

$$E(X) = \mu_1, \quad V(X) = \sigma_1^2$$
$$E(Y) = \mu_2, \quad V(Y) = \sigma_2^2$$

とする．このとき $V(X-3Y)$ と $E(X^2Y^2)$ を，$\mu_1, \mu_2, \sigma_1^2, \sigma_2^2$ を用いて表せ．

9. 互いに独立な 3 つの確率変数 X, Y, Z の分散をそれぞれ σ_1^2, σ_2^2, σ_3^2 とする．

(a) $X+Y$ と $X-Y$ の共分散を求めよ．

(b) $X+Y$ と $X+Z$ の共分散を求めよ．

10. X, Y を確率変数とする．また a, b, c, d を定数として a, c は正の値とする．X と Y の相関係数を $\rho(X,Y)$ とするとき

$$\rho(aX+b, cY+d) = \rho(X,Y)$$

が成り立つことを示せ．

11. 区間 $[0,1]$ からランダムに 1 点とりその点を X とする．次に区間 $[X,1]$ からランダムに 1 点とりその点を Y とする．ここで「ランダム」とは一様分布にしたがうこととする．

(a) Y の期待値を求めよ．

(b) Y の分散を求めよ．

(c) $E[XY]$ を求めよ．

12. 確率変数 $X_1, X_2, \ldots, X_n$ が独立に期待値 0，分散 1 の正規分布にしたがうとする．

    (a) X_1^2 の積率母関数 $E[e^{tX_1^2}]$ を計算せよ．

    (b) 確率変数 Z を $Z = \sum_{i=1}^{n} X_i^2$ と定義する．Z の積率母関数 $E[e^{tZ}]$ を計算せよ．

13. 1 次元確率変数 X の分布関数を $F(x)$ $(x \in \mathbb{R})$ とする．また U を区間 $[0,1]$ 上の一様分布にしたがう確率変数とする．F^{-1} を F の逆関数とするとき，確率変数 $F^{-1}(U)$ の分布関数は F であることを示せ．

14. 確率変数 $X_1, \ldots, X_n$ は独立に一様分布 $U[0,c]$ にしたがうとする．一様分布 $U[0,c]$ $(c > 0)$ の密度関数 $f(x)$ は

$$f(x) = \begin{cases} 1/c & 0 \leq x \leq c \\ 0 & \text{その他} \end{cases}$$

    で与えられる．以下の問に答えよ．

    (a) $Y = \max_{1 \leq i \leq n} X_i$ とするとき，Y の分布関数を求めよ．

    (b) $Z = \min_{1 \leq i \leq n} X_i$ とするとき，Z の分布関数を求めよ．

15. X, Y をともに離散な確率変数とする．「$\mathrm{Cov}(X,Y) = 0$」は成り立つが「X, Y は独立ではない」というような例を構成せよ．実際に条件を満たしていることも示すこと．