

統計学入門 第5章 確率変数

張 梁

2007年3月16日

目 次

5 確率変数	2
5.1 確率変数 (random variable) と確率分布 (probability distribution)	2
5.1.1 離散型 (discrete type) の確率変数の確率分布	2
5.1.2 連続型 (continuous type) の確率変数の確率分布	3
5.1.3 累積分布関数 (cumulative distribution function, CDF)	5
5.2 確率変数の期待値と分散	6
5.2.1 期待値 (expectation)	6
5.2.2 分散と標準偏差	7
5.2.3 標準化	8
5.3 モーメント (moment) と モーメント母関数 (moment generating function)	9
5.3.1 歪度 (歪度係数) (skewness)	9
5.3.2 尖度 (kurtosis)	9
5.3.3 モーメント	9
5.3.4 モーメント母関数	10
5.4 チェビシェフの不等式 (Chebyshev's inequality)	11
5.5 確率変数の変換	11

5 確率変数

5.1 確率変数 (random variable) と確率分布 (probability distribution)

確率変数 ある特性に注目して試行を行うとき，その特性を表す数値が偶然的法則にしたがって種々の値をとり，とる各値に対してそれぞれ確率が与えられている変数を確率変数と呼び，大文字の X で表す。

さいころを振って出る目 X は確率変数である。

正しいさいころなら

$$P(X = 1) = \frac{1}{6}, P(X = 2) = \frac{1}{6}, \dots, P(X = 6) = \frac{1}{6} \quad (1)$$

であるが、一般には、

$$P(X = 1) = p_1, P(X = 2) = p_2, \dots, P(X = 6) = p_6 \quad (2)$$

である。ここで，

$$p_1 \geq 0, p_2 \geq 0, \dots, p_6 \geq 0, p_1 + p_2 + \dots + p_6 = 1 \quad (3)$$

確率分布 X のとる各値に対して各確率を対応させたもの，すなわち確率変数に対して定められた確率を確率分布という。

5.1.1 離散型 (discrete type) の確率変数の確率分布

離散型の確率分布 (probability distribution of discrete type) 一般に可算集合 $\{x_1, x_2, \dots\}$ の中の値をとる確率変数 X は離散型といわれ，それぞれの値の確率

$$P(X = x_k) = f(x_k) \quad (k = 1, 2, \dots) \quad (4)$$

を X の確率分布という。 f は (3) に対応して

$$f(x_k) \geq 0, \quad (k = 1, 2, \dots) \quad \text{かつ} \quad \sum_{k=1}^{\infty} f(x_k) = 1 \quad (5)$$

の条件を満たす。この f を離散型の確率分布という。確率分布は確率の重みの分布の様子を表している。二項分布，ポアソン分布，超幾何分布などはその例である。

$f(x_k)$ とか， x_k を x と記して $f(x)$ と書くことも多い。そのとき確率分布の条件は，すべての x に対して $f(x) \geq 0$ ，かつ $\sum_x f(x) = 1$ となる。

乱数 (random number, random digits) 0 から 9 までの数字で, 起きる確率が等しいことが理論的に保証されているとき, 出る数字を乱数という。これを表にしたものを乱数表という。乱数は以下の二つの性質を満足ものでなければならない。

- (i) 等確率性 (0 から 9 までの数字が出現する確率は等しくすべて 0.1)
- (ii) 無規則性 (i 番目に出た数字と j 番目に出た数字の間には相関がない)

5.1.2 連続型 (continuous type) の確率変数の確率分布

確率変数 X のとる値が関数 $f(x)$ によって

$$P(a \leq X \leq b) = \int_a^b f(x) dx \quad (6)$$

と表された場合, X を連続型の確率変数といい, X は連続型の確率分布をもつという。ただし, (3) に対応して

$$\text{すべての } x \text{ に対し, } f(x) \geq 0, \quad \text{かつ} \quad \int_{-\infty}^{\infty} f(x) dx = 1 \quad (7)$$

であるものとし, 関数 $f(x)$ を X の確率密度関数 (probability density function) 或いは単に密度関数という。確率変数 X が $f(x)$ 確率分布をもつとき, X は $f(x)$ に「従う」という。

ここで a と b とが極めて近いと考えると

$$P(x \leq X \leq x + \Delta x) \doteq f(x) \cdot \Delta x \quad (\Delta x \text{ は小}) \quad (8)$$

である。

また, $a = b$ とおけば,

$$P(X = a) = 0 \quad (9)$$

となるから、連続型の確率分布においては、一点の確率は 0 となる。したがって,

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

となり, 区間の端点を含めても含めなくても同じである。

連続型確率分布の例: 数学的に扱いやすい連続型の確率分布として,

指数分布 (exponential distribution) $\lambda > 0$ として

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & (x \geq 0) \\ 0 & (x < 0) \end{cases} \quad (10)$$

とすると, $f(x) \geq 0$ で

$$\int_{-\infty}^{\infty} f(x) dx = \int_0^{\infty} \lambda e^{-\lambda x} dx = 1 \quad (11)$$

と規格化されているから, 確率分布の性質を満たす。定積分すると

$$\begin{aligned} P(a \leq X \leq b) &= \int_a^b \lambda e^{-\lambda x} dx \\ &= - \int_{-\lambda a}^{-\lambda b} e^t dt \quad (t = -\lambda x) \\ &= -[e^t]_{-\lambda a}^{-\lambda b} \\ &= e^{-\lambda a} - e^{-\lambda b} \end{aligned} \quad (12)$$

ということがわかる。

電球が偶発的に切れるまでの寿命 X などは, 一般的に待ち時間といわれる。災害や故障がまったく偶然的要因に支配されているときは, 待ち時間は指数分布に従う。

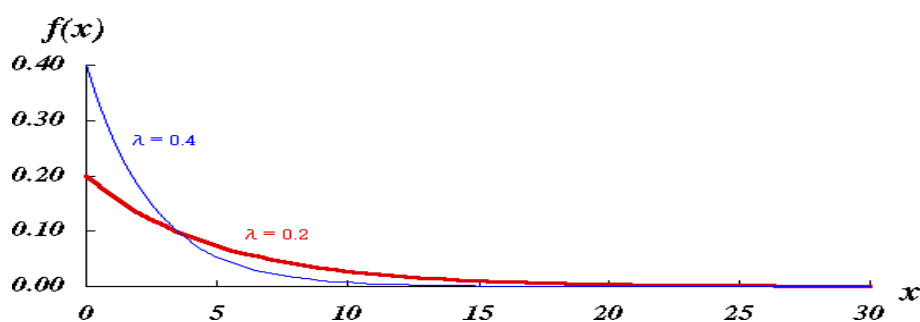


図 1: 指数分布密度関数

一様分布 (uniform distribution)(連続型) ある区間 $[a, b]$ 内の全ての値を同等にとる分布を一様分布という。確率密度関数は,

$$f(x) = \begin{cases} \frac{1}{b-a} & (a \leq x \leq b) \\ 0 & (\text{その他}) \end{cases} \quad (13)$$

である。

標準的には区間 $[0,1]$ 上の一様分布を考えることが多い。

$$f(x) = \begin{cases} 1 & (0 \leq x \leq 1) \\ 0 & (x < 0, x > 1) \end{cases} \quad (14)$$

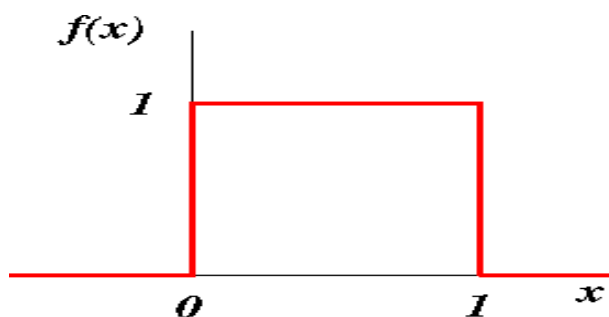


図 2: 区間 $[0,1]$ の一様分布密度関数

区間 $[0,1]$ の任意の値を等しくとる一様乱数 (uniform random number) はこの確率分布に従う。

5.1.3 累積分布関数 (cumulative distribution function, CDF)

確率変数 X に対して, x を実数とするととき x 以下の確率

$$F(x) = P(X \leq x) \quad (15)$$

を X の累積分布関数と呼ぶ。

連続型の $F(x)$ は密度関数 f の定積分

$$F(x) = \int_{-\infty}^x f(u) du \quad (16)$$

となる。 $F(x)$ から

$$F'(x) = f(x) \quad (17)$$

として, 元の密度関数 $f(x)$ が復元される。したがって, $F(x)$ でも $f(x)$ でも確率分布を表すことができる。累積分布関数は、次の性質をもつ。

- (1) (非負性) $0 \leq F(x) \leq 1$
- (2) (単調性) $F(x_1) \leq F(x_2) \quad (x_1 < x_2)$
- (3) (右連続性) 各点 x で $\varepsilon \downarrow 0$ のとき $F(x + \varepsilon) \rightarrow F(x)$
- (4) (極限) $\lim_{x \rightarrow -\infty} F(x) = 0 \quad \lim_{x \rightarrow \infty} F(x) = 1$

指数分布の累積分布関数

$$F(x) = \begin{cases} \int_{-\infty}^x f(u)du = \int_{-\infty}^x 0 \cdot du = 0, & (x < 0) \\ \int_{-\infty}^x f(u)du = \int_0^x \lambda e^{-\lambda u} du = 1 - e^{-\lambda x}, & (x \geq 0) \end{cases} \quad (18)$$

一様分布の累積分布関数

$$F(x) = \begin{cases} 0 & (x < a) \\ \frac{x-a}{b-a} & (a \leq x \leq b) \\ 1 & (x > b) \end{cases} \quad (19)$$

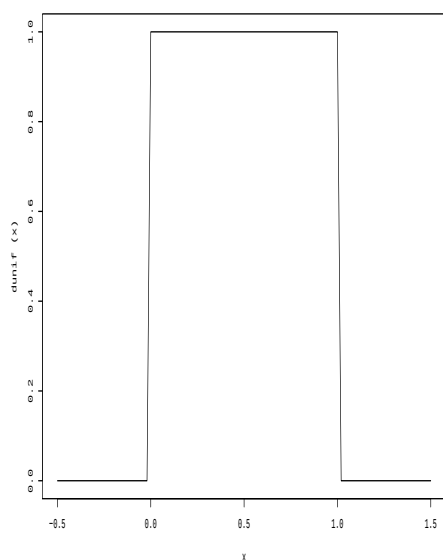


図 3: 一様分布の密度関数

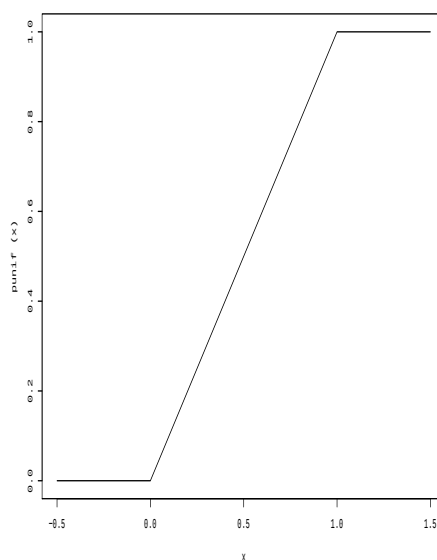


図 4: 一様分布の累積分布関数

5.2 確率変数の期待値と分散

5.2.1 期待値 (expectation)

期待値 確率変数 X に対して、それがとる値の重みつき平均を、確率変数の期待値といい、 $E(x)$ と書く。

$$E(X) = \sum_x x f(x) \quad (\text{離散型}) \quad (20)$$

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx \quad (\text{連続型}) \quad (21)$$

X の関数についても同様に期待値を定義する。

指数分布の期待値 式 (9) に従う確率変数 X に対し，期待値は

$$\begin{aligned} E(X) &= \int_{-\infty}^{\infty} x f(x) dx \\ &= \int_0^{\infty} x \cdot \lambda e^{-\lambda x} dx \\ &= \frac{1}{\lambda} \end{aligned} \quad (22)$$

一様分布の期待値 式 (13) に従う確率変数 X に対し，期待値は

$$E(X) = \int_{-\infty}^{\infty} x f(x) dx = \int_0^1 x \cdot 1 \cdot dx = \frac{1}{2} \quad (23)$$

期待値の演算 E について、次のような性質を持つ。

- (a) $E(c) = c$
- (b) $E(X + C) = E(x) + c$
- (c) $E(cX) = cE(X)$
- (d) $E(aX \pm bY) = aE(X) \pm bE(Y)$
- (e) $E(XY) = E(X)E(Y) + \sigma_{XY}$

5.2.2 分散と標準偏差

分散 期待値が同じでも，異なった確率分布はいくらでもある。期待値 $E(X)$ からのずれの量 $X - E(X)$ をばらつきを考える。期待値を $\mu = E(X)$ ，

$$E(X - \mu) = E(X) - \mu = \mu - \mu \equiv 0 \quad (24)$$

となり，正負打ち消しあってばらつきが 0 になってしまう。かわりに $(X - \mu)^2$ を考える，

$$V(X) = E\{(X - \mu)^2\} = \sigma^2 \quad (25)$$

と定義しこれを X の分散という。このとき必ず， $V(X) \geq 0$ であり， $V(X)$ の値の大きいほど X のばらつきは大きい。定義から，

$$V(X) = \sum_x (x - \mu)^2 f(x) \quad (\text{離散型}) \quad (26)$$

$$V(X) = \int_{-\infty}^{\infty} (x - \mu)^2 f(x) dx \quad (\text{連続型}) \quad (27)$$

であるが,

$$\begin{aligned}
 V(X) &= E(X^2) - 2\mu E(X) + \mu^2 \\
 &= E(X^2) - \mu^2 \\
 &= E(X^2) - (E(X))^2
 \end{aligned} \tag{28}$$

標準偏差 分散の平方根 $D(X) = \sqrt{V(X)} = \sigma$ とおき, X の標準偏差という。
分散の演算 V について、次のような性質を持つ。

- (a) $V(c) = 0$ $(= E(c^2) - (E(c))^2 = c^2 - c^2)$
- (b) $V(X + c) = V(X)$
 $(= E((X + c)^2) - E(X + c)^2 = (E(X^2) + 2cE(X) + c^2) - ((E(X))^2 + 2cE(X) + c^2))$
- (c) $V(cX) = c^2V(X)$ $(= E(c^2X^2) - (E(cX))^2 = c^2E(X^2) - c^2(E(X))^2)$
- (d) $V(aX \pm bY) = a^2V(X) + b^2V(Y) \pm 2ab\sigma_{XY}$ (証明略)

5.2.3 標準化

任意の X からその期待値を引いて, $E(X - E(X)) = 0$ となるが,

$$\begin{aligned}
 V(X - E(X)) &= V(X) \\
 (&= E\{(X - E(X))^2\} - (E\{X - E(X)\})^2 \\
 &= E\{X^2 - 2X \cdot E(X) + (E(X))^2\} - 0 \\
 &= E(X^2) - 2(E(X))^2 + (E(X))^2)
 \end{aligned} \tag{29}$$

と変わらない。

X の分散によって,

$$Z = \frac{1}{\sqrt{V(X)}} \cdot \{X - E(X)\} \tag{30}$$

と定義すると, $X' = X - E(X)$ について,

$$\begin{aligned}
 E(Z) &= 0 \\
 (&= \frac{1}{\sqrt{V(X)}} \cdot E(X - E(X))) \\
 V(Z) &= 1 \\
 (&= \frac{1}{V(X)} \cdot V(X - E(X)) \\
 &= \frac{V(X)}{V(X)})
 \end{aligned} \tag{31}$$

となる。

いかなる確率変数も, その期待値を引き, さらにその標準偏差の尺度で割れば, 期待値は 0 に, 分散は 1 に調整される。この変換を標準化といい, Z を標準化変数という。

5.3 モーメント (moment) と モーメント母関数 (moment generating function)

5.3.1 歪度 (歪度係数)(skewness)

確率分布の非対称性の方向，およびその程度を表すには

$$\begin{aligned}\alpha_3 &= E(X - \mu)^3 / \sigma^3 \\ &= (E(X^3) - 3\mu E(X^2) + 2\mu^3) / \sigma^3\end{aligned}\quad (32)$$

と呼ぶ。

- $\alpha_3 > 0$ ならば右の裾が長い
- $\alpha_3 < 0$ ならば左の裾が長い
- $|\alpha_3|$ がその程度を表す。

5.3.2 尖度 (kurtosis)

確率分布の、尖りの程度を表す指標，すなわち，四乗の期待値

$$\begin{aligned}\alpha_4 &= E(X - \mu)^4 / \sigma^4 \\ &= (E(X^4) - 4\mu E(X^3) + 6\mu^2 E(X^2) - 2\mu^4) / \sigma^4\end{aligned}\quad (33)$$

は、中心の周囲の部分のとり具合を表す。正規分布の $\alpha_4 = 3$ と比較して， $\alpha_4 - 3$ を X の確率分布の尖度 (超過係数 (coefficient of excess)) という。

- $\alpha_4 - 3 > 0$ ならば正規分布より尖っている。
- $\alpha_4 - 3 < 0$ ならば正規分布より丸く鈍い形をしている。

5.3.3 モーメント

確率分布の形は $E(X - \mu)^r$ なる量で決まってくることが分かる。一般に、

$$\mu_r = E(X^r) \quad (34)$$

を， X の (原点の周りの) r 次の (モーメント)，または (積率) といい。

$$\mu'_r = E(X - \mu)^r \quad (\text{ただし } \mu = E(X)) \quad (35)$$

を， X の (期待値 (平均) の周りの) r 次の (モーメント) という。また

$$\alpha_r = E\{(X - \mu)/\sigma\}^r \quad (36)$$

は, X の r 次の (標準化モーメント) といわれる。ここで,

$$\begin{aligned}\mu_0 &\equiv 1, & \mu'_1 &\equiv 0 \\ \mu_1 &= E(X), & \mu'_2 &= V(X)\end{aligned}\quad (37)$$

歪度 α_3 , 尖度を定義するときの α_4 などは, 高次の標準化モーメントである。

5.3.4 モーメント母関数

全ての次数のモーメントを指定すれば, それにより一つの確率分布が決定されるはずである。全てのモーメントを生成するモーメント母関数 (積率母関数) を

$$M_X(t) = E(e^{tx}) \quad (38)$$

と定義する。計算は,

$$\begin{aligned}M_X(t) &= \sum_x e^{tx} f(x) && \text{(離散型)} \\ M_X(t) &= \int_{-\infty}^{\infty} e^{tx} f(x) dx && \text{(連続型)}\end{aligned}\quad (39)$$

による。式 (37) から, モーメント母関数を繰り返し微分して 0 とおいた導関数は

$$M'_x(0) = \mu_1, \quad M''_x(0) = \mu_2, \quad M'''_x(0) = \mu_3, \dots \quad (40)$$

などとなる。一般にモーメント母関数の r 階導関数から

$$M_x^{(r)}(0) = \mu_r \quad (41)$$

のように, 各次数のモーメントが分かる。

指数分布のモーメント母関数

$$\begin{aligned}M_X(t) &= \int_0^{\infty} e^{tx} \lambda e^{-\lambda x} dx \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x} dx \\ &= \lambda/(\lambda - t) && (t < \lambda \text{ のとき})\end{aligned}\quad (42)$$

となる。

$$\begin{aligned}M'_x(0) = \mu_1 &= [\lambda/(\lambda - t)^2]_0 = 1/\lambda, & M''_x(0) = \mu_2 &= [\lambda \cdot 2/(\lambda - t)^3]_0 = 2/(\lambda^2), \\ M'''_x(0) = \mu_3 &= [(\lambda \cdot 2 \cdot 3)/(\lambda - t)^4]_0 = 6/\lambda^3, \\ \dots\dots, M_x^{(r)}(0) = \mu_r &= [(\lambda \cdot r!)/(\lambda - t)^{r+1}]_0 = r!/\lambda^r, \dots\dots\end{aligned}\quad (43)$$

などと求められる。

5.4 チェビシェフの不等式 (Chebyshev's inequality)

標準偏差 $\sigma = D(x)$ を単位にとるとき，一般に次の不等式が成り立つ。これをチェビシェフの不等式という。確率分布の式が具体的に分かっていなくとも，期待値と分散さえ分かっているならば，確率の値が不等式の形で示される。

いかなる確率変数 X に対しても

$$P(|X - \mu| \geq k\sigma) \leq 1/k^2 \quad (44)$$

となる。ただし， $\mu = E(X)$ ， $\sigma^2 = V(X)$ である。(証明は教科書 105 ページ、「入門統計学」85 ページ参照)

5.5 確率変数の変換

一般に $Y = \phi(X)$ ， ϕ は単調増加とおこう。 $y = \phi(x)$ ， $\Delta y = \phi(x + \Delta x) - \phi(x)$ とする ($\Delta x, \Delta y$ はいずれも微小とする)。したがって， $x \leq X \leq x + \Delta x$ と $y \leq Y \leq y + \Delta y$ は論理的に同値である。 $P(y \leq Y \leq y + \Delta y) = P(x \leq X \leq x + \Delta x)$ とわかる。

X, Y の密度関数を $f(x), g(y)$ で表せば，式 (8) より $g(y) \cdot \Delta y = f(x) \cdot \Delta x$ 。 $y = \phi(x)$ を x について解いた関数 (逆関数) を $x = \varphi(y)$ として代入すると， Y の密度関数は

$$g(y) = f(x) \Delta x / \Delta y \doteq f(x) (dx/dy) = f(\varphi(y)) \cdot |d\varphi(y)/dy| \quad (45)$$

参考文献

- [1] 橋本 智雄「入門統計学」 共立出版 (2004)
- [2] 垂水 共之・飯塚 誠也「R/S-PLUS による統計解析入門」共立出版 (2006)