Introduction to mathematical statistics ゼミ (第 1 回)

担当: 伊藤真道

7/17

1 Chap.1. Distributions of Random Variables

1.1 Introduction / 序章

- 標本空間 (sample space): ランダム試行 (random experiment) の起こりうる結果 (outcome) の全てを集めた空間のこと．以下 $\mathscr{C}$ で標本空間を表す．
- 相対頻度 (relative frequancy): $C \subset \mathscr{C}$(C を $\mathscr{C}$ の部分集合とする)，$\#(A)$ で事象 (event)A の要素数とする．相対頻度とは，N をランダム試行の回数, $f = \#(C)$ とした時，f/N で定義される量．N が増加するにつれて，相対頻度は安定することが知られている．頻度論では，"C の起こる確率" と呼ばれる．

1.2 Algebra of sets / 集合代数

定義 1.1 (部分集合 (subset))**.** もし，集合 A_1 の全ての要素が，集合 A_2 の要素でもある場合，"集合 A_1 は集合 A_2 の部分集合である" といい，$A_1 \subset A_2$ で表す．本書では，集合の次元が明らかな場合は，わざわざ書かないこととする．

¡note¿

定義 1.2 (空集合 (null set))**.** もし，集合 A が一つも要素を持たないとき，A は空集合 *(null set)* であるといい，ϕ で表す．

¡note¿

定義 1.3 (和集合 (union))**.** 少なくとも，集合 A_1，A_2 のいずれか一つ *(*両方でもいいよ*)* に属する全ての要素の集合のことを A_1 と A_2 の和集合 *(union)* といい，$A_1 \cup A_2$ で表す．いくつかの集合 $A_1, A_2, \ldots$ の和集合とは，いくつかの集合 $A_1, A_2, \ldots$ の少なくとも一つには属する全ての要素の集合のことである．この和集合は，$A_1 \cup A_2 \cup \ldots$ と表され，もし集合の個数が有限の k 個なら，$A_1 \cup A_2 \cup A_3 \ldots \cup A_k$ と表される．

¡note¿

定義 1.4 (共通部分 (intersection))**.** 集合 A_1 と A_2 の両方に属する全ての要素の集合のことを，A_1 と A_2 の共通部分 *(intersection)* といい，$A_1 \cap A_2$ で表す．いくつかの集合 $A_1, A_2, \ldots$ の共通部分とは，いくつかの集合 $A_1, A_2, \ldots$ の全てに属する要素全ての集合のことである．この和集合は，$A_1 \cap A_2 \cap \ldots$ と表され，もし集合の個数が有限の k 個なら，$A_1 \cap A_2 \cap A_3 \ldots \cap A_k$ と表される．

¡note¿

定義 1.5 (空間 (space))**.** ある特定の議論や，想定のもとで，その議論に関係する全ての要素全体を記述することができる．この，想定下の全ての要素の集合は空間 *(space)* と呼ばれる．本書では，空間を花文字 *(*$\mathscr{A}, \mathscr{B}, \mathscr{C}$*)* で表すことがおおい．

¡note¿

定義 1.6 (補集合)**.** $\mathscr{A}$ を空間，A をその部分集合とする．$\mathscr{A}$ の要素であり，A の要素ではないものの集合を A の *(*$\mathscr{A}$ *のもとでの)* 補集合と呼び，A^c で表す．特に，$\mathscr{A}^c = \phi$ である．

¡note¿

1.3 Set function / 集合関数

- 集合関数 (set function): ある特定の点のみではなく，点の集合全体に対して定義される関数．

¡note¿

1.4 The probability set function / 確率集合関数

定義 1.7 (確率集合関数)**.** もし，$P(C)$ が，標本空間 $\mathscr{C}$ の部分集合に対して定義されており，なおかつ，

1. $P(C) \geq 0$
2. 互いに排反な集合 C_i *(*$=C_i \cap C_j = \phi,\ i \neq j$*)* について，$P(C_1 \cup C_2 \cup C_3 \cup \cdots) = P(C_1) + P(C_2) + P(C_3) + \cdots$
3. $P(\mathscr{C}) = 1$

を満たすとき，$P(C)$ をランダム試行の結果に対する確率集合関数と呼ぶ．標本空間 $\mathscr{C}$ のそれぞれの部分集合 C における，$P(C)$ を”ランダム試行の結果が集合 C の要素である確率”や，”事象 C の確率”，”集合 C における確率測度”という．

¡note¿

定理 1.1. 全ての $C \subset \mathscr{C}$ に対して，$P(C) = 1 - P(C^c)$

証明 1.1. $\mathscr{C} = C \cup C^c, C \cap C^c$ とする．定義 *1.7* の *(3)*，*(2)* から，

$$1 = P(C) + P(C^c)$$

よって定理 *1* が得られた．

¡note¿

定理 1.2. 空集合の確率は *0*，つまり $P(\phi) = 0$ である．

証明 1.2. 定理 *1* にて，$C = \phi$ とすると，$C^c = \mathscr{C}$ である．結果として，

$$P(\phi) = 1 - P(\mathscr{C}) = 1 - 1 = 0 \text{ よって，定理 2 が示された．}$$

¡note¿

定理 1.3. もし，標本空間 $\mathscr{C}$ の部分集合 C_1 と C_2 が $C_1 \subset C_2$ であるならば，$P(C_1) \leq P(C_2)$

証明 1.3. 今，$C_2 = C_1 \cup (C_1^c \cup C_2)$，$C_1 \cap (C_1^c \cap C_2) = \phi$ とする．定義 *1.7* の *(2)* から，

$$P(C_2) = P(C_1) + P(C_1^c \cup C_2)$$

であるが，定義 *1.7(1)* から，$P(C_1^c \cup C_2) \geq 0$ である．以上から，$P(C_2) \geq P(C_1)$

¡note¿

定理 1.4. 任意の $C \subset \mathscr{C}$ に対して，$0 \leq P(C) \leq 1$

証明 1.4. $\phi \subset C \subset \mathscr{C}$ であるから，定理 *1.3* から，

$$P(\phi) \leq P(C) \leq P(\mathscr{C}) \; or \; 0 \leq P(C) \leq 1$$

¡note¿

定理 1.5. もし，$C_1, C_2 \subset \mathscr{C}$ なら，

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

証明 1.5. $C_1 \cup C_2$ と C_2 は互いに共通部分を持たない集合の和集合として，以下のように表すことができる．

$$\begin{aligned} C_1 \cup C_2 &= C_1 \cup (C_1^c \cap C_2) \\ C_2 &= (C_1 \cap C_2) \cup (C_1^c \cap C_2) \end{aligned}$$

これと，定義 *1.7(2)* から，

$$P(C_1 \cup C_2) = P(C_1) + P(C_1^c \cap C_2)$$

$$P(C_2) = P(C_1 \cap C_2) + P(C_1^c \cap C_2)$$

これらを解くと，

$$P(C_1 \cup C_2) = P(C_1) + P(C_2) - P(C_1 \cap C_2)$$

¡note¿
7/21 証明を訂正しました．

1.5 Random variables / 確率変数

定義 1.8 (確率変数 (random variable))**.** 標本空間 $\mathscr{C}$ で，ランダム試行が行われたとする．任意の $c \in \mathscr{C}$ に対して，$X(c) = x$ である，つまり，ある一つの実数値を割り当てる関数 X のことを確率変数 (random variable) と呼ぶ．X の空間は，実数値の集合 $\mathscr{A} = \{x | x = X(c), c \in \mathscr{C}\}$ である．

¡note¿

定義 1.9 (2 つの確率変数の空間)**.** ある標本空間 $\mathscr{C}$ で，ランダム試行が行われたとする．今，任意の $c \in \mathscr{C}$ に対して，ある一つの実数値を割り当てる，つまり，$X_1(c) = x_1, X_2(c) = x_2$ であるような順序の存在する二つの確率変数 X_1, X_2 を考える．X_1 と X_2 の空間は，$\mathscr{A} = \{(x_1, x_2) | x_1 = X_1(c), x_2 = X_2(c), c \in \mathscr{C}\}$ のような，x_1, x_2 を並べたものを要素に持つような集合である．

¡note¿

定義 1.10 (n 個の確率変数の空間)**.** ある標本空間 $\mathscr{C}$ で，ランダム試行が行われたとする．$X_i(i =$

$1, 2, \cdots, n)$ を，任意の $c \in \mathscr{C}$ に対して，ある一つの実数値を割り当てる，つまり，$X_i(c) = x_i$ であるような確率変数とする．これらの確率変数の空間は，$\mathscr{A} = \{(x_1, x_2, \ldots, x_n) | x_1 = X_1(c), \ldots, x_n = X_n(c), c \in \mathscr{C}\}$ のような，順序のある長さ n の組の集合である．さらに，$A \subset \mathscr{A}$ とすると，$Pr[(X_1(c), \ldots, X_n(c)) \in A] = P(C)$．ただし，ここで，$C = \{c | c \in \mathscr{C} and\ [X_1(c), \ldots, X_n(c)] \in A\}$ である．

¡note¿

1.6 The probability density function / 確率密度関数

確率変数の種類によって，確率密度関数 (probability density function) は，主に二つのタイプに分類される．

- 離散型 (discrete type) の確率変数:

  X を 1 次元空間 $\mathscr{A}$ 上の確率変数とする．空間 $\mathscr{A}$ を，全ての有限の区間に最大でも有限個の $\mathscr{A}$ の点が存在するような点の集合とする．そのような集合 $\mathscr{A}$ は，離散的な点の集合と呼ばれる．

  関数 $f(x)$ を $f(x) > 0, x \in \mathscr{A}$ であり，なおかつ，

$$\sum_{\mathscr{A}} f(x) = 1$$

  であるような関数とする，どんな，$A \subset \mathscr{A}$ に対しても，確率集合関数 $P(A)$ が

$$P(A) = Pr(X \in A) = \sum_{A} f(x)$$

  のように表されるとき，X を**離散型の確率変数**であると呼び，X は離散型の分布を持つという．

- 連続型 (continuous type) の確率変数:

  1 次元空間 $\mathscr{A}$ を，$f(x) > 0, x \in \mathscr{A}$ に対して，Rieman integral が定義され，

$$\int_{\mathscr{A}} f(x)dx = 1$$

  であり，なおかつ，$f(x)$ が $\mathscr{A}$ の部分集合である，全ての有限の区間に，不連続な点を最大でも有限個しか持たないとする．もし，$\mathscr{A}$ が確率変数 X の空間であり，確率集合関数

$P(A), A \subset \mathscr{A}$ が $f(x)$ を用いて，

$$P(A) = Pr(X \in A) = \int_A f(x)dx$$

と表される時，X を**連続型の確率変数**と呼び，連続型の分布を持つという．

もし，$A = \{x|a < x < b\}$ である時，事象 A が生起する確率 $P(A) = Pr(X \in A)$ は，

$$Pr(a < X < b) = \int_a^b f(x)dx$$

と表され，さらに，$A = \{x|x = a\}$ である時，

$$P(A) = Pr(X \in A) = Pr(X = a) = \int_a^a f(x)dx = 0$$

となる．X が連続型の確率変数の時，一つの点から構成される集合の確率は 0 となる．

確率密度関数は以下の性質を満たす．逆に，これらを満たすものを確率密度関数と定義しても良い (稲垣，1990)

1. $f(x) \geq 0, x \in \mathscr{A}$
2. $\int_{\mathscr{A}} f(x)dx = 1$(連続型確率変数の場合, 離散型の場合は，$\int$ が $\sum$ に)
3. 分布関数は $F(x) = \int_{-\infty}^{x} f(x)dx$

¡note¿

1.7 The distribution function / 分布関数

確率変数 X が，1 次元の集合 A 上で確率集合関数 $P(A)$ を持つとする．$A = \{x| -\infty < x\}$ とすると，$P(A) = Pr(X \in A) = Pr(X \leq x)$ となる．これは，点 x のみによって定まる関数であり，分布関数 (時に累積分布関数) と呼ばれ，特別に

$$F(x) = Pr(X \leq x)$$

と表される．分布関数は，確率密度関数 $f(x)$ を用いて，

$$F(x) = \sum_{w \leq x} f(w)(\text{離散型})$$

,

$$F(x) = \int_{-\infty}^{x} f(w)dw(\text{連続型})$$

のように表せる．分布関数が離散型か連続型のいずれかになるかは，確率変数が，そのいずれの型かに依存する．

次に分布関数のいくつかの重要な性質について，列挙する．

1. 分布関数 $F(x)$ は，至る所で連続
2. 分布関数 $F(x)$ の x に関する微分は存在し，それは確率密度関数に一致する．つまり，

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

   である．
3. $0 \leq F(x) \leq 1$
4. $F(x)$ は，x の非減少関数である．
5. $F(\infty) = 1$ であり，$F(-\infty) = 0$ である．
6. $F(x)$ は右側連続である．

4 番目の性質から，

$$Pr(a < X \leq b) = F(b) - F(a)$$

である．

¡note¿

1.8 Certain probability model / 特定の確率モデル

実数直線上の閉区間 $[a, b]$ から，一つの点を選ぶ試行を考える．標本空間 $\mathscr{C}$ は $[a, b]$ であり，確率変数 X は，$\mathscr{C}$ 上で定義される恒等関数とする．そのため，X の空間は，$\mathscr{A} = \mathscr{C}$ である．試行の性質から，区間 A が $\mathscr{A}$ の部分空間である時，事象 A の起こる確率は，区間 A の相対的な長さとするのが自然であろう．つまり，もし，$A = [a, x], x \leq b$ の時，比率定数 c を用いて，

$$P(A) = Pr(X \in A) = Pr(a \leq X \leq x) = c(x - a)$$

となる．上記の表現で，$x = b$ とすると，

$$1 = Pr(a \leq X \leq b) = c(b - a)$$

であり，$c = 1/(b - a)$ であることがわかる．よって，分布関数 $F(x) = Pr(X \leq x)$ を，

$$\begin{aligned} F(x) &= 0, (x < a) \\ &= \frac{x - a}{b - a}, (a \leq x \leq b) \\ &= 1, (b < x) \end{aligned}$$

のようにとると，適切な確率モデルを得ることができるだろう．結果として，X の確率密度関数は，

$$\begin{aligned} f(x) = F'(x) &= \frac{1}{b - a}, (a \leq x \leq b) \\ &= 0(elsewhere) \end{aligned}$$

となる．この例では，X は区間 $[a, b]$ 上で**一様分布**をもつという．

例 2(超幾何分布)
100 本のヒューズで構成されている，あるロットが以下の手順で検査されるとする．100 本のうち，ランダムに 5 本のヒューズを選び，検査する．もし，5 本のヒューズ全てが適切な電流量で飛ぶならば，そのロットは採用される．もし，ここでロットを構成する 100 本のヒューズのうち 20 本が欠陥品だとした時，ロットが採用される確率は，適切な条件下で，

$$\frac{\binom{80}{5}}{\binom{100}{5}} = 0.32$$

である．より一般的に，確率変数 X が検査される 5 本のうち欠陥品であるヒューズの個数とする．X の空間は，$\mathscr{A} = \{x | x = 0, 1, 2, 3, 4, 5\}$ であり，確率密度関数は，

$$f(x) = Pr(X = x) = \frac{\binom{20}{x}\binom{80}{5-x}}{\binom{100}{5}},\ (x = 0, \ldots, 5) \tag{1}$$

で与えられる．この離散型確率分布の例は，**超幾何分布**と呼ばれる．
¡note¿

1.9 Mathematical Expectation / 数学的期待

確率変数の分布に関する，より有用な概念に，期待値 (Expectation value) がある．X を確率密度関数 $f(x)$ を持つ確率変数，$u(X)$ を X の関数とする．期待値は，以下のように定義される．

$$E[u(X)] = \int_{-\infty}^{\infty} u(x)f(x)dx \text{ (連続型確率変数)} \tag{2}$$

$$E[u(X)] = \sum_{x} u(x)f(x) \text{ (離散型確率変数)} \tag{3}$$

上記の定義は，確率変数が一次元空間上で定義されている場合である．確率変数が n 次元空間上で定義されている場合は，その分，$\int$ や $\sum$ が増える．以下，期待値の有用な性質を列挙する．

1. もし，k が定数の場合，$E[k] = k$
2. もし，k が定数，v が関数の場合，$E[kv] = kE[v]$
3. もし，k_1, k_2 が定数，v_1, v_2 が関数の場合，$E[k_1v_1 + k_2v_2] = k_1E[v_1] + k_2E[v_2]$

確率変数の線形和の期待値は，期待値の定数倍の和になるが，確率変数の積の期待値は，期待値の席にはならないことに注意.

¡note¿

1.10 Some special mathematical expectation / いくつかの特別な数学的期待

先ほどと，同じセッティングで，離散型確率変数の場合を考える．離散型確率変数の期待値は

$$E[X] = \sum_x x f(x)$$

であり，もし，確率密度関数の正の部分の空間の離散点が，$a_1, a_2, a_3, \ldots,$ である時，期待値は，

$$E[X] = a_1 f(a_1) + a_2 f(a_2) + a_3 f(a_3) + \cdots .$$

と表すことができる．これは，値 $a_1, a_2, a_3, \ldots,$ を $f(a_1), f(a_2), f(a_3), \ldots,$ で重みづけた重み付き平均とみなすことができる．このことから，$E[X]$ は，算術平均 (arithmetic mean)，もしくは，平均値 (mean value) と呼ばれることがある．確率変数の平均 μ が存在するとき，確率変数が連続型，離散型にかかわらず，

$$\mu = E[X]$$

で定義される．

¡note¿

もう一つの特別な期待値は，$u(X) = (X - \mu)^2$ とした時に得られる．これは，X が離散型の時，

$$E[(X - \mu)^2] = \sum_x (x - \mu)^2 f(x) \tag{4}$$

$$= (a_1 - \mu)^2 f(a_1) + (a_2 - \mu)^2 f(a_2) + (a_3 - \mu)^2 f(a_3) + \cdots . \tag{5}$$

と表される．これは，"$a_1, a_2, a_3, \ldots,$ の平均 μ からのズレの二乗"を，$f(a_1), f(a_2), f(a_3), \ldots,$ で重みづけた重み付き平均とみなすことができる．この平均からの確率変数 X のずれの二乗の平均を分散 (variance) と呼び，σ^2 で表す．また，分散の正の二乗根を標準偏差 (standard deviation) と呼ぶ．これらを改めて数式で書くと，

$$\sigma^2 = E[(X - \mu)^2]$$

,

$$\sigma = \sqrt{\sigma^2}$$

である．分散の計算は，

$$\sigma^2 = E[X^2] - \mu^2$$

でやるのが楽な場合がある．簡単なので，各自示してみましょう．

¡note¿

次に積率母関数 (moment generating function) について紹介する．積率母関数は以下のように定義される期待値のことである．

$$E[e^{tX}] = \int_{-\infty}^{\infty} e^{tx} f(x) dx \ (\text{連続型確率変数}) \tag{6}$$

$$E[e^{tX}] = \sum_{x} e^{tx} f(x) \ (\text{離散型確率変数}) \tag{7}$$

積率母関数はしばしば，

$$M(t) = E[e^{tX}]$$

のように，$M(t)$ で表される．以下，積率母関数の性質を列挙する．

1. 積率母関数は確率変数の分布を完全に決定する．(=分布形と積率母関数が一対一の対応)
2. 積率母関数の m 回微分を $M^{(m)}(t)$ とすると，

   $$M^{(m)}(t)|_{t=0} = E[X^m]$$

   である．(=m 回微分して $t = 0$ とした時，それは確率変数 m 上の期待値に一致する．これを確率変数の m 次のモーメントと呼ぶことがある．)

全ての分布が積率母関数を持つわけではないことに注意されたい．
¡note¿

次に紹介するのは特性関数 (characteristic function) である．多くの分布は，積率母関数を持たない．しかし，特性関数

$$\psi(t) = E[e^{itX}]$$

は全ての分布に対して存在する．

証明)

まず，特性関数の絶対値を考えると，

$$|\psi(t)| = |\int_{-\infty}^{\infty} e^{itx} f(x) dx| \leq \int_{-\infty}^{\infty} |e^{itx} f(x)| dx$$

となるが，$f(x)$ は非負の関数なので $|f(x)| = f(x)$，また，

$$|e^{ixt}| = |\cos tx + i \sin tx| = 1$$

である．よって，

$$|\psi(t)| \leq \int_{-\infty}^{\infty} f(x) dx = 1$$

となり，1 によって上から押さえつけられる．

特性関数も，積率母関数と同様に，分布と 1 対 1 対応しており，微分して $t = 0$ と置くことで，X のモーメントを計算することができる．

¡note¿

1.11 Chebyshev's inequality / チェビシェフの不等式

定理 1.6 (マルコフの不等式)**.** $u(X)$ を確率変数に対する非負値関数とする．もし，$E[u(X)]$ が存在するなら，任意の正定数 c に対して，

$$Pr[u(X) \geq c] \leq \frac{E[u(X)]}{c} \tag{8}$$

が成立する．

証明 1.6. ここでは，確率変数 X が連続型として証明する．$A = \{x | u(x) \geq c\}$，$f(x)$ を確率密度関数とする．すると，

$$E[u(X)] = \int_{-\infty}^{\infty} u(x) f(x) dx = \int_{A} u(x) f(x) dx + \int_{A^c} u(x) f(x) dx$$

である．最右辺の $\int_A u(x) f(x) dx, \int_{A^c} u(x) f(x) dx$ は非負だから，

$$E[u(X)] \geq \int_{A} u(x) f(x) dx$$

である．ここで，$A = \{x | u(x) \geq c\}$ であることに注意すると，

$$E[u(X)] \geq \int_A u(x) f(x) dx \geq c \int_A f(x) dx$$

となる．また，

$$\int_A f(x) dx = Pr[X \in A] = Pr[u(X) \geq c]$$

である．以上から，

$$E[u(x)] \geq cPr[u(X) \geq c]$$

が成立する．

¡note¿

定理 1.7 (チェビシェフの不等式 (Chebyshev's inequality))**.** 確率変数 X が確率分布を持つとし，有限の分散 σ^2 を持つとする．任意の $k > 0$ に対し，

$$Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

，同様のことだが，

$$Pr(|X - \mu| < k\sigma) \geq 1 - \frac{1}{k^2}$$

が成立する．

証明 1.7. 定理 *1.6* において，$u(X) = (X - \mu)^2, c = k^2\sigma^2$ とおく．すると，

$$Pr[(X - \mu)^2 \geq k^2\sigma^2] \leq \frac{E[(X - \mu)^2]}{k^2\sigma^2} = \frac{1}{k^2}$$

となる．よって，

$$Pr(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

が得られる．

チェビシェフの不等式はあくまで，上限であり，実際の確率はもっと低いこともあるということに注意されたい．
¡note¿