

測度論的確率論

目次

- [初頭的な話\(古典的確率論\)](#)
- [ここからの話の前提知識](#)
- [標本点、事象、確率の定義](#)
- [確率変数](#)
- [確率の表記方法](#)
- [確率変数から生成される \$\sigma\$ 加法族](#)
- ["初頭的"な条件付確率](#)
- [ベイズの定理と独立](#)
- [確率分布\(または単に分布\)](#)
- [分布関数、同時分布関数、周辺分布関数](#)
- [分布関数の性質](#)
- [離散型、連続型、密度関数](#)
- [確率変数、確率分布、分布関数の演繹的な関係](#)
- [確率変数、確率分布、分布関数の帰納的？な関係](#)
- [期待値](#)
- [初頭的な条件付き期待値](#)
- [例題1](#)
- [期待値による確率変数の生成](#)

初頭的な話(古典的確率論)

初頭的な話は高校の時にやっているので省略。忘れていたら(そのままだでも構わないが)マセマの確率統計第1章などに書いてある。

ここからの話の前提知識

初歩的な集合論や位相空間論、測度論、ルベーグ積分などは既習とする。

標本点、事象、確率の定義

(Ω, F, P) なる測度空間を考える。ここでさらに $P(\Omega) = 1$ の時、 (Ω, F, P) を **確率空間** と呼び、 P を **確率(測度)**、 F の元を **事象**、 Ω の元を **標本点** と呼ぶ。また、 $P(A)$ を **事象Aの確率**と呼ぶ。

また、写像 $P : F \rightarrow R$ を **確率法則** と呼ばれる。

確率変数

(Ω, F, P) なる確率空間と $(R^n, B(R^n))$ なる測度空間を考える。さらに写像 $X : \Omega \rightarrow R$ を考えこれが F -可測関数ならば、 X を **(d次元)確率変数** という。 $n=1$ の場合特に **実確率変数** という。 F -可測関数とは数式にすると($n=1$ の場合、)

1. f が可測関数である。つまり、任意の $B \in B(R)$ の逆像 $f^{-1}(B)$ が $f^{-1}(B) \in F$
2. 任意の区間 I の逆像 $f^{-1}(I)$ が $f^{-1}(I) \in F$
3. 任意の $a \in R$ に対し、 $\{x \in X | f(x) > a\} \in F$ 。つまり $f^{-1}((a, \infty)) \in F$
4. 任意の $a \in R$ に対し、 $\{x \in X | f(x) \geq a\} \in F$ 。つまり $f^{-1}([a, \infty)) \in F$
5. 任意の $a \in R$ に対し、 $\{x \in X | f(x) < a\} \in F$ 。つまり $f^{-1}((-\infty, a)) \in F$
6. 任意の $a \in R$ に対し、 $\{x \in X | f(x) \leq a\} \in F$ 。つまり $f^{-1}((-\infty, a]) \in F$

確率の表記方法

確率変数を用いたよく使われる確率の表記方法を紹介する。 (Ω, F, P) なる確率空間と確率変数 X を考える。

$$\begin{aligned} P(X = x_i) &:= P(X^{-1}(\{x_i\})) \\ P(X \in A) &:= P(X^{-1}(A)) \\ P(a \leq X \leq b) &:= P(X^{-1}([a, b])) \end{aligned}$$

確率変数から生成される σ 加法族

(Ω, F, P) なる確率空間と $(R^d, B(R^d))$ なる測度空間を考える。まず、確率変数 $X : \Omega \rightarrow R^d$ は可測関数の条件より任意の $B \in B(R^d)$ に対し

$$X^{-1}(B) = \{x \in \Omega | f(x) \in B^d\} \in F$$

の性質を満たす。よって逆に確率変数 X が与えられたとき

$$\sigma(X) := \{X^{-1}(B) | B \in B(R^d)\}$$

と定義すると $\sigma(X)$ は σ 加法族になる。証明は逆像の性質と σ 加法族の性質を用いれば簡単にできる。よって **確率変数 X で記述される事象は $\sigma(X)$ で最小である。**

"初頭的"な条件付確率

(Ω, F, P) なる確率空間に対し、ある事象 B を条件づけたときの、事象 A の **条件付き確率** $P(A|B)$ は

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$

と定義される。ただし、 $P(B) > 0$

$P(A|B)$ の直感的な意味は、事象Bが起きるものとした時の事象Aが起こる確率となる。

また、実際に計算すると、 (Ω, F, P) における条件付き確率 $P(\cdot|B)$ は確率測度となっているため $(\Omega, F, P(\cdot|B))$ の期待値も当然同じ式で求めることが出来る。

ベイズの定理と独立

条件付き確率の定義より以下の定理が成り立つ。

$$P(B|A) = \frac{P(A|B)}{P(A)}P(B)$$

この定理は以下の意味で重要である。

- $P(B)$ は事象Aが起きる前の事象Bの確率(事前確率)
- $P(B|A)$ は事象Aが起きた後での事象Bの確率(事後確率)

と見る事が出来る。また、 Ω の直和分割 $B_1, \dots, B_n \in F$ とすると $P(A) = \sum_{i=1}^n P(B_i)P(A|B_i)$ なので

$$P(B_j|A) = \frac{P(B_j)P(A|B_j)}{P(A)} = \frac{P(B_j)P(A|B_j)}{\sum_{i=1}^n P(B_i)P(A|B_i)}$$

また、以下の式が成り立つ時、事象AとBは独立であるという。 $(P(B)>0$ とすると)

$$P(A|B) = P(A) \Leftrightarrow P(A \cap B) = P(A)P(B)$$

が成り立つ時である。これは事象Bが起きたのを知っても事象Aの確率(事後確率)が変化しないという=事象Aと事象Bは独立している意味になる。

確率分布(または単に分布)

(Ω, F, P) なる確率空間と $(R^d, B(R^d))$ なる測度空間を考える。さらに(d次元)確率変数 $X = (X_1, X_2, \dots, X_d)$ が定義されているとする。

ここで測度 P^X を以下のように構成する。 $(P^X$ は P_X と書かれる本もあり。)

$$P^X : B(R^d) \rightarrow [0, 1] \quad P^X(B) = P(X \in B) = P(X^{-1}(B))$$

とする。するとこれは測度の条件を満たすため測度となりさらに確率測度となる。この P^X は**Xの確率分布**、または単に**Xの分布**という。なお、測度論の言葉を使うなら(自分の測度論のサイトには載せてない

が)像測度という言葉に対応する。ちなみにディラック測度($\delta_a(A) = 1 \quad (a \in A)$)なる測度を用いると P^X の表記が楽になることがある。

これによって (Ω, F, P) なる確率空間から確率変数 $X : \Omega \rightarrow R^n$ を用いて、 $(R^d, B(R^d), P^X)$ なる確率空間を定義できた。

これは **確率変数とその値になる確率** ということもできる。確率変数が取らない集合を確率分布に入れば0になるためである。

分布関数、同時分布関数、周辺分布関数

(Ω, F, P) なる確率空間と $(R^d, B(R^d), P^X)$ なる確率空間を考える。さらに(d次元)確率変数 $X = (X_1, X_2, \dots, X_d)$ が定義されているとする。

ここで $x \in R^d$ に対し、 $F : R^d \rightarrow [0, 1]$ を

$$\begin{aligned} F(x) &:= P^X((-\infty, x_1] \times \dots \times (-\infty, x_d]) \\ &= P^X(\{X \in R^n \mid X_1 \leq x_1, \dots, X_n \leq x_n\}) \end{aligned}$$

とおく。この $F(x)$ を**Xの分布関数**という。特に $d \geq 2$ の時の $F(x)$ を**Xの同時分布関数**という。

また、 X_i に対する分布関数を**X_iの分布関数**という。

定義を見ると、 $F(x)$ は確率変数 X が $X < x$ になる確率と見ることが出来る。このような見方が出来れば逆に確率変数 X が x になる確率は $F(X) - F(X_-)$ となるのは彰がだろう。つまり分布関数が(右連続は前提として)不連続でないと $X = x$ となる確率は0である。逆に不連続なら $X = x$ となる確率は0ではないのも明らかだろう。

分布関数の性質

1. F は単調非減少関数
2. $F(\infty) = 1$ かつ $F(-\infty) = 0$
3. $F(x)$ は右連続関数
4. $P(a < X \leq b) = F(b) - F(a)$ 、 $P(a \leq X \leq b) = F(b) - F(a_-)$ 。ただし、 $F(a_-) = \lim_{x \rightarrow -0+a} F(x)$
5. F の不連続点はたかだか可算個

要はルベーグ-スティルチェス測度を P とした時の F かな？(裏が取れたら追記)

離散型、連続型、密度関数

$1_A(x) = 1 \quad (x \in A), 0 \quad (OtherWise)$ なる関数を **定義関数** という。また、定義より直ちに $1_\emptyset(x) = 0$ となることも念頭に置こう。

$p(a) := P(X = a)$ なる関数 $p : A \rightarrow [0, 1]$ を **確率関数** と呼ぶ。または $f_x(a) := P(X = a)$ と書いて **確率質量関数** と呼ばれる。

d次元ベクトル x, y が $x \leq y$ とは各 x_i, y_i に対し、 $x_i \leq y_i$ が成り立つとする。

たかだか可算な集合 $A \subset R^d$ に対し、 $P(X \in A) = 1$ が成り立つとき、**Xは離散型** という。つまり高々可算個の1点集合の確率変数の確率の和が1になるということである。数式にすると $\sum_{a \in A} p(a) = 1$ のAが高々可算可算であると言っている。また、**分布関数の取る値が高々可算個(ならば→不連続(逆は成り立たない))**と言ってもいい。Aを高々可算個とした場合Aの任意の元にNでナンバリング出来るのでそれくお a_i とする。この時分布関数は以下のように書ける。

$$\begin{aligned} F(x) &= \sum_{i: a_i \leq x} P(X = a_i) = \sum_{i=1}^{\infty} p(a_i) 1_{\{b|a_i \leq b \leq x\}}(x) \\ &= \sum_{i: a_i \leq x} P^X(\{a_i\}) \end{aligned}$$

ここで $1_{\{b|a_i \leq b \leq x\}}(x)$ は $x \leq a_i$ ならば $1_\emptyset(x)$ となるため常に0となり左辺と等しいのは分かるだろう。さらに確率分布 P^X も

$$P^X(B) = \sum_{a_i \in A} p(a_i) \delta_{a_i}(B)$$

となる。よって離散型だと確率(質量)関数 $p(x)$ が重要になる。

また、分布関数 F が R^d 上連続の時 **Xは連続型** という。特に $x \in R^d$ に対し非負関数 $f(x)$ 、ただし $\int_{R^d} f(x) dx = 1$ とする。これを用いて

$$F(x) = \int_{(-\infty, x_1] \times \cdots \times (-\infty, x_d]} f(z) dz = \int_{R^d} f(z) 1_{\{z \leq x\}} dz$$

と書けるとき、Fは**絶対連続型** といい、 f を(確率変数Xの分布関数Fの)確率密度関数 または単に Xの密度関数 という。

確率変数、確率分布、分布関数の演繹的な関係

今までの流れを整理する。

- (Ω, F, P) なる確率空間がまず 与えられている。
- さらにそこに $X : \Omega \rightarrow R^d$ なる確率変数(可測関数)を定義する。
- すると P^X なるXの分布($(R^d, B(R^d))$ の確率測度)が求まる。つまり $(R^d, B(R^d), P^X)$ なる確率空間が導ける。

- さらにそれ(P^X)を用いると X の分布関数 $F(x)$ が求まる。分布関数の取る値がただか可算個なら離散型、全ての点で連続なら連続型である。
- ただし、場合によっては分布関数の取る値が非可算個なのに全ての点では連続にならないケースもちろんあるだろう。これは混合分布と呼ばれる。

となる。混合型になる例を一つ上げる。

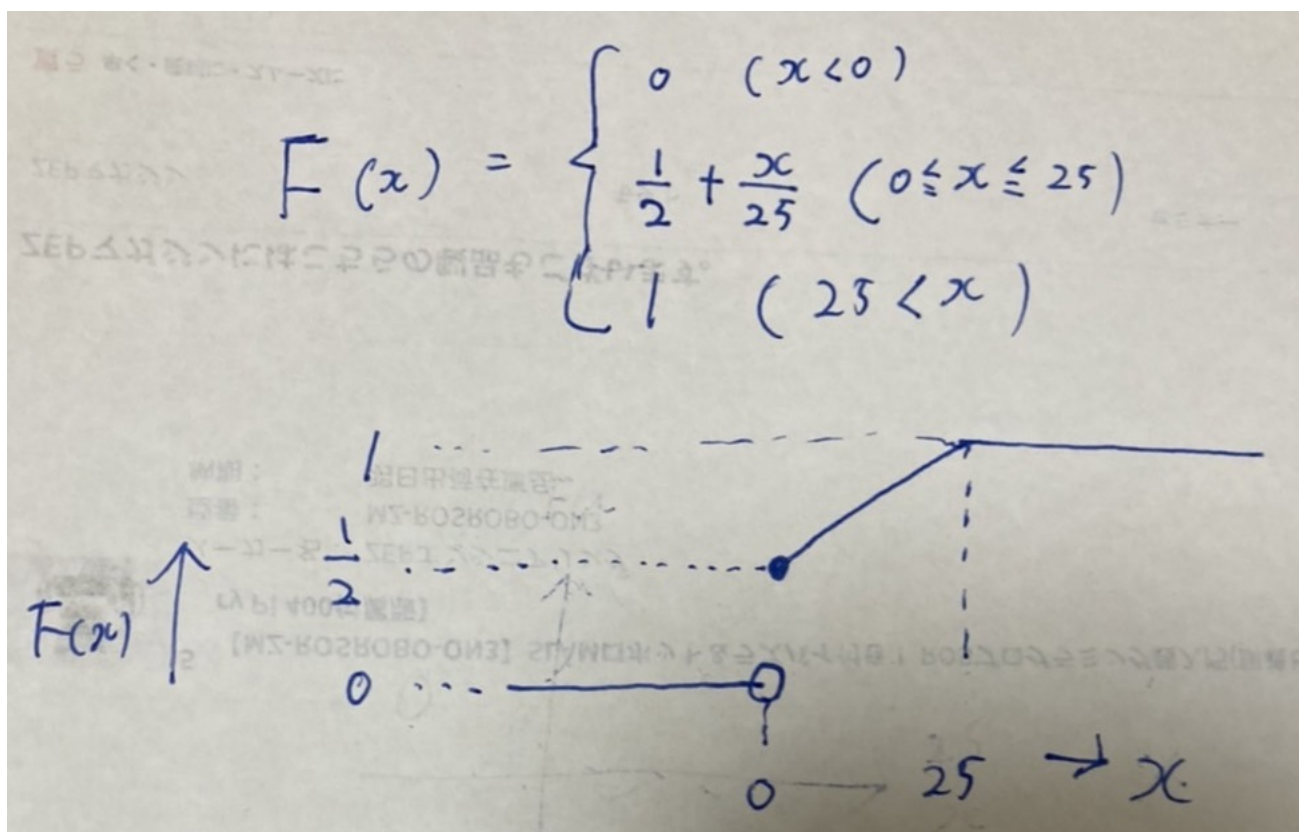
今、車がA市にいる。午後0時から午後1時の間をランダムに出発する。車は25km離れたB市に時速50kmで向かったとする。ここで午後1時置ける車のB市からの距離の分布関数を求めよ。

まず、出発時間 $\omega \in [0, 1]$ に対し、確率変数 $X : \omega \rightarrow R$ を午後1時置ける車のB市からの距離とする。すると問題の設定より明らかに

$$P^X(B) = \frac{1}{2}\delta_0(B) + \frac{1}{2}\frac{\mu_{[0,25]}(B)}{25}$$

となる。ここで $\delta_0(B)$ は $0 \in B$ の時だけ1になりそれ以外は0となる関数(ディラック測度)である。

$\mu_{[0,25]}(B)$ は制限されたルベーグ測度 $\mu_{[0,25]} := \mu(B \cap [0, 25])$ である。つまり最大で25である。よって分布関数は画像のものになる。



確率変数、確率分布、分布関数の帰納的？な関係

まず、確率分布 P^X と分布関数 $F(x)$ は1対1となる。 P^X を $F(x)$ で定まるルベーグ-スティルチェス測度と考えればまあ納得である。

また離散分布では確率(質量)関数, 連続分布では確率密度関数 $f(x)$ が重要な役割を果たす。なぜなら、

期待値

(Ω, F, P) なる確率空間と $X : \Omega \rightarrow R$ なる確率変数を考える。確率変数 X が P -可積分である時

$$E[X] := \int_{\Omega} X(\omega) P(d\omega)$$

と書いて、 $E[X]$ を X の期待値という。 X が P -可積分でない時には期待値を持たないという。なお、 $B \in F$ なる B に積分範囲を制限した期待値を

$$E[X; B] = E[1_B X] = \int_{\Omega} 1_B(\omega) X(\omega) P(d\omega)$$

と書く。 1_B とは B の元に対してだけ1、それ以外は0を取る関数である。

初頭的な条件付き期待値

(Ω, F, P) なる確率空間に対し、条件付き確率 $P(\cdot | B)$ を定義した時その期待値を $E[X | B]$ と表し、**確率変数 X による B に関する条件つき期待値** という。

例題1

サイコロを1回投げることを考える。

$$\Omega = \{i \text{ の目が出る} | i = 1, 2, 3, 4, 5, 6\}$$

ここで集合 A_i を

$$A_i = \{i \text{ の目が出る}\} \quad (\subset \Omega = \{1 \text{ の目が出る}, \dots, 6 \text{ の目が出る}\})$$

と置き、 $F = 2^{\Omega}$ とする。さらに測度 P を

$$P(A_i) = \frac{1}{6}$$

とおく。以上より確率空間 (Ω, F, P) が定まった。さらに確率変数 $X : \Omega \rightarrow \{1, 2, 3, 4, 5, 6\}$ を以下のように定める。

$$X^{-1}(\{i\}) = A_i \quad (i \in F)$$

これは以下と同値である。

$$X(i\text{の目が出る}) = i$$

ここで「偶数の目が出る」という事象 $B = A_2 \cup A_4 \cup A_6$ を考える。すると条件付き確率 $P(\cdot|B)$ は

$$P(A_i|B) = 0 (i \in \{1, 3, 5\} \text{の時}), \frac{1}{3} (i \in \{2, 4, 6\} \text{の時})$$

となる。よって期待値 $E[X]$ 、 $E[X|B]$ は

$$\begin{aligned} E[X] &= \sum_{i=1}^6 X(A_i)P(A_i) \\ &= 1 \cdot \frac{1}{6} + 2 \cdot \frac{1}{6} + \cdots + 6 \cdot \frac{1}{6} \\ &= \frac{7}{2} \end{aligned}$$

$$\begin{aligned} E[X|B] &= \sum_{i=1}^6 X(A_i)P(A_i|B) \\ &= 1 \cdot 0 + 2 \cdot \frac{1}{3} + \cdots + 5 \cdot 0 + 6 \cdot \frac{1}{3} \\ &= 4 \end{aligned}$$

期待値による確率変数の生成
