『社会科学のためのベイズ統計モデリング』第 1 章

Hiroshi Hamada
Tohoku University

April, 2020
at Tohoku University

1.1 事象と標本空間

例：保険加入行動

集合

$$\Omega = \{\,\text{加入しない}, \text{加入する}\,\}$$

を**標本空間**（sample space）と呼ぶ．

定義 (部分集合)

集合 A の全ての要素が集合 B の要素であるとき，A を B の**部分集合**と呼び，$A \subset B$ と書く． 確率論では標本空間 Ω の部分集合を**事象**とよぶ．

1.1 事象と標本空間

確率は次の性質を満たす

1. 任意の事象 A に対して，$0 \leq P(A) \leq 1$
2. $P(\Omega) = 1$，$P(\emptyset) = 0$
3. $A \cap B = \emptyset$ ならば $P(A \cup B) = P(A) + P(B)$

1.2 確率変数

標本空間

$$\Omega = \{\,\text{加入しない}, \text{加入する}\,\}$$

の要素に，それぞれ数字の 0 と 1 を対応させます．

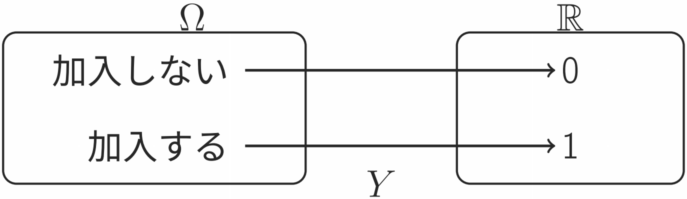
$$\text{加入しない} \to 0$$
$$\text{加入する} \to 1$$

標本空間 Ω の要素を数と対応させる関数を確率変数（random variable）という．

1.2 確率変数

- 確率変数は標本空間 Ω の全ての要素に《数》を対応させる関数
- 要素と対応する数字（0 や 1）を確率変数の実現値（realization）と呼ぶ
- 《確率変数》は対応のルールそのもの (大文字)
- 《確率変数の実現値》はルールによって標本空間の要素と対応する数値（小文字）

1.2 確率変数



確率変数 Y のイメージ

確率変数の実現値と確率の対応．

Ω の要素		実現値	確率
加入しない	$\to$	0	$1-q$
加入する	$\to$	1	q

- 《$Y=0$ になる確率》は $1-q$
- 《$Y=1$ になる確率》は q

式で書くと

$$P(Y(\text{加入しない})=0)=P(Y=0)=1-q$$
$$P(Y(\text{加入する})=1)=P(Y=1)=q.$$

1.3 確率分布

例 (ベルヌーイ分布)

確率変数 Y が確率 q で $Y = 1$，確率 $1 - q$ で $Y = 0$ となるとき，Y はベルヌーイ分布 (Bernoulli distribution) にしたがう，という． □

確率変数の分布を表す記号

$$X \sim \text{分布名 (パラメータ)}$$

ベルヌーイ分布

$$Y \sim \mathrm{Bernoulli}(q)$$

正規分布

$$X \sim \mathrm{Normal}(\mu, \sigma)$$

定義 (確率質量関数)

離散的な実現値の集合を S とおく．関数 $f : S \to \mathbb{R}$ が次の性質を満たすとき，確率質量関数（probability mass function）という．単に確率関数とも呼ぶ．

1. 任意の $x \in S$ について $f(x) \geq 0$
2. $\displaystyle\sum_{x \in S} f(x) = 1.$

離散確率変数 X の実現値が x となる確率を確率質量関数 $f(x)$ で

$$P(X = x) = f(x)$$

と定義する．$X = a$ である確率は $f(a)$ である

$\mathbb{R}$ は実数全体の集合

定義 (確率密度関数)

次の性質を満たす関数 $f : \mathbb{R} \to \mathbb{R}$ を**確率密度関数**（probability density function）という．

1. 任意の $x \in \mathbb{R}$ について $f(x) \geq 0$
2. $\displaystyle\int_{-\infty}^{\infty} f(x)dx = 1$

連続確率変数 X の実現値が区間 $[a, b]$ 内におさまる確率 $P(a \leq X \leq b)$ を，次の確率密度関数の積分で定義する．

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

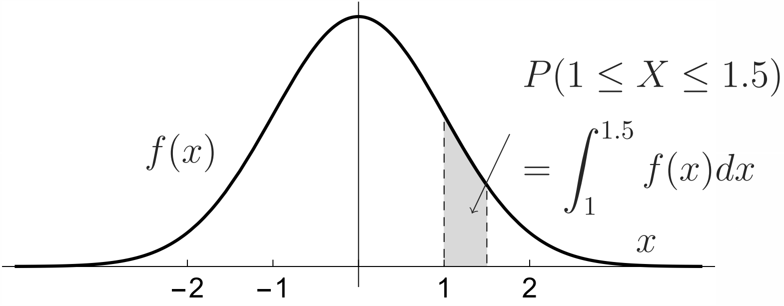


Figure: 確率密度関数 $f(x)$ のグラフと確率 $P(1 \le X \le 1.5)$ に対応する面積

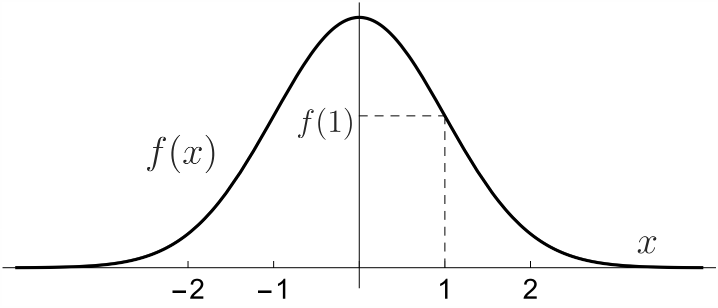


Figure: 確率密度関数 $f(x)$ のグラフと $x = 1$ における $f(1)$ の位置

任意の $a \in \mathbb{R}$ にたいして

$$P(X = a) = P(a \leq X \leq a) = \int_a^a f(x)dx = 0$$

定義 (確率変数の期待値)

離散確率変数 X の確率質量関数が
$f(x_i) = p_i, i = 1, 2, \ldots, n$ であるとき，次の総和

$$x_1 p_1 + x_2 p_2 + \cdots + x_n p_n = \sum_{i=1}^{n} x_i p_i$$

を確率変数 X の期待値（expectation）と呼ぶ．確率変数
X の期待値を $\mathbb{E}[X]$ と書く．

$$\mathbb{E}[X] = \sum_{i=1}^{n} x_i p_i = \sum_{i=1}^{n} x_i \cdot f(x_i).$$

定義 (連続確率変数の期待値)

連続確率変数 X の期待値は，次の積分

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

である．ただし $f(x)$ は X の確率密度関数である．

同時確率と確率変数の独立

定義 (n 個の離散確率変数の独立)

離散確率変数 $X_1, X_2, \ldots, X_n$ が独立であるとは、任意の実現値 $x_k (k = 1, 2, \ldots, n)$ について、

$$
\begin{aligned}
&P(X_1 = x_1, X_2 = x_2, \ldots, X_n = x_n) \\
&= P(X_1 = x_1)P(X_2 = x_2)\cdots P(X_n = x_n)
\end{aligned}
$$

が成立することをいう。

連続確率変数 X, Y の同時確率密度関数のグラフの例

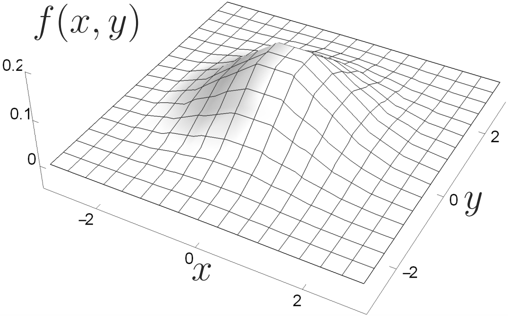


Figure: X, Y の同時確率密度関数 $f(x,y)$ のグラフ

山の体積は 1.

定義 (同時確率密度関数)

n 組の実数 $(x_1, x_2, \ldots, x_n)$ の関数 $q(x_1, x_2, \ldots, x_n) \geq 0$ が

$$\iint \cdots \int q(x_1, x_2, \ldots, x_n) dx_1 dx_2 \cdots dx_n = 1$$

を満たすとき，$q(x_1, x_2, \ldots, x_n)$ を同時確率密度関数という．

図に描けないから理解できない，と考えてはいけない．単に n 重積分で 1 になる関数と（抽象的に）考えればよい．

X, Y の同時確率密度関数が $p(x, y)$ であるとき，

$$p(x, y) = p(x)p(y)$$

ならば X, Y は独立．

定義 (n 個の連続確率変数の独立)

連続確率変数 $X_1, X_2, \ldots, X_n$ が独立であるとは，同時確率密度関数について

$$q(x_1, x_2, \ldots, x_n) = q(x_1)q(x_2) \cdots q(x_n)$$

が成立することをいう

サンプルと真の分布

- n を自然数として，ある同一の確率分布 $q(x)$ に独立に従う確率変数の組

$$X_1, X_2, \dots, X_n$$

をサンプルと呼ぶ
- サンプルの実現値は $x_1, x_2, \dots, x_n$
- n 個のサンプルの実現値

$$x^n = (x_1, x_2, \dots, x_n)$$

- サンプルの実現値は直積の要素

$$x^n = (x_1, x_2, \ldots, x_n) \in \underbrace{\mathbb{R} \times \mathbb{R} \times \cdots \times \mathbb{R}}_{n} = \mathbb{R}^n$$

- サンプルは，同一の分布に，独立にしたがうと仮定する．independent and identically distributed
- i.i.d. サンプルの確率密度関数

$$q(x_1, x_2, \ldots, x_n) = \prod_{i=1}^{n} q(x_i)$$

- i.i.d. でないサンプルを扱うモデルは馬場 (2019:267-332) などを参照

例 (サンプルの期待値)

n 個のサンプル（確率変数）について，その関数

$$f(X^n) = f(X_1, X_2, \dots, X_n)$$

が与えられたとき，その平均をとる操作 $\mathbb{E}[\quad]$ を

$$\mathbb{E}[f(X^n)] = \int\int\cdots\int f(x_1, x_2, \dots, x_n)\prod_{i=1}^{n} q(x_i)dx_i$$

と表す $\square$

統計的推測

- 観察したデータはサンプル（確率変数）の実現値であり，観察した実現値から観察できない真の分布を推測することを，統計的推測ないし統計的学習という
- 《真の分布がデータを生成する》という仮定の下で，真の分布はおそらくこうではないかという予想を，予測分布という形で分析者が定式化

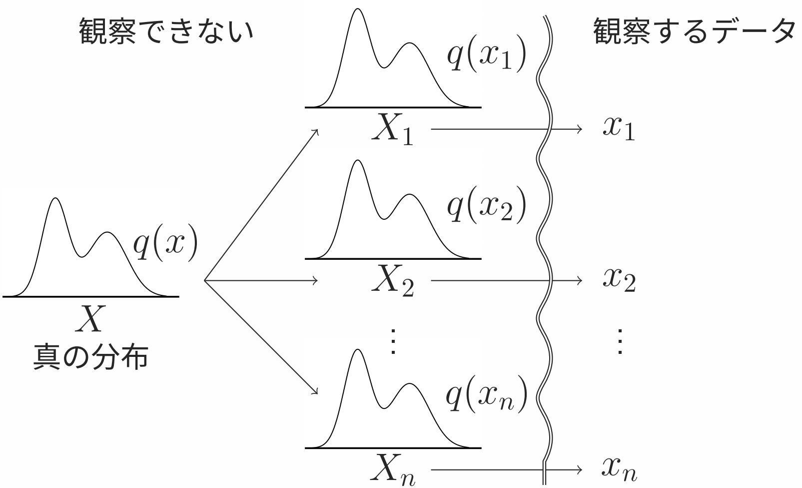
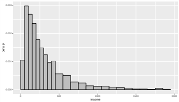


Figure: 真の分布・サンプル（確率変数）・実現値のイメージ

観察したデータ $x_1, x_2, \ldots, x_n$ からそれを生成した未知の分布 $q(x)$ を推測する

現実



日本の所得分布 $q(x)$

母集団

ランダム・サンプリング

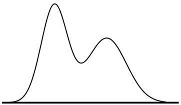
$x_1 \quad x_2 \quad \cdots \quad x_n$

実現値（データ）

推定値 $\hat{\theta} = f(x_1, x_2, \ldots, x_n)$

母集団の推測

モデルの世界



未知の分布 $q(x)$

i.i.d.

ズレの確認

汎化損失

$X_1 \quad X_2 \quad \cdots \quad X_n$

確率変数

推定量 $\hat{\Theta} = f(X_1, X_2, \ldots, X_n)$

確率モデル $p(x|\theta) \neq q(x)$

予測分布

$p(x|\hat{\theta})$ や $\int p(x|\theta) \underbrace{p(\theta|x^n)}_{\text{posterior}} d\theta$

1 章まとめ

- 試行の結果として起こる可能性をすべて集めた集合を標本空間という．その部分集合を事象という
- 標本空間の要素を数に対応させるルール（関数）を確率変数という
- 独立に同一の分布にしたがう n 個の確率変数の組をサンプルという
- データとは確率変数（サンプル）の実現値である
- 現実に観察したデータから，観察不可能な真の分布を推測することが統計学の目的．