

---

第 1 部では、情報幾何の議論に必要となる確率論と数理統計の前提知識を整理する。ただしここではごく基礎的な事項のみを述べ、それ以外の話題は第 2 部で情報幾何的な視点から扱うものとする。そもそも確率・統計を幾何学的に定式化しようとする際には、座標変換に関する不変性などを定義に繰り入れる必要がある。このとき定義が見かけ上煩雑になることで、確率・統計的な意味を一時的に見失ってしまうおそれがある。それを防ぐために、確率・統計的な意味を一旦明らかにしておくのもこの部の目的である。

# 第1章 確率論

確率論の基礎事項を整理する。

## 1.1 確率空間

**定義 1.1.1** (確率空間). 測度空間  $(\Omega, \mathcal{F}, P)$  であって

- (1) 各  $E \in \mathcal{F}$  に対し  $P(E) \geq 0$
- (2)  $P(\Omega) = 1$

をみたすものを**確率空間 (probability space)** といい、 $P$  を  $(\Omega, \mathcal{F})$  上の**確率測度 (probability measure)** あるいは**確率分布 (probability distribution)** という。

**定義 1.1.2** (確率変数).  $(\Omega, \mathcal{F}, P)$  を確率空間、 $(X, \mathcal{A})$  を可測空間とする。可測関数  $X: (\Omega, \mathcal{F}) \rightarrow (X, \mathcal{A})$  を  $(X, \mathcal{A})$  に値をもつ**確率変数 (random variable; r.v.)** という。とくに  $X = \mathbb{R}$ ,  $\mathcal{A} = \mathcal{B}(\mathbb{R})$  (Borel 集合族) のとき、 $X$  を単に**確率変数**という。

**定義 1.1.3** (確率変数の確率分布).  $X: (\Omega, \mathcal{F}) \rightarrow (X, \mathcal{A})$  を確率変数とする。このとき、写像

$$P^X: \mathcal{A} \rightarrow [0, +\infty], \quad E \mapsto P(X^{-1}(E)) \quad (E \in \mathcal{A}) \quad (1.1.1)$$

は  $(X, \mathcal{A})$  上の測度となる (このあと示す)。これを  $X$  の**確率分布 (probability distribution of  $X$ )** という。

$X$  の確率分布が  $(X, \mathcal{A})$  上のある確率分布  $\nu$  に等しいとき、 $X$  は  $\nu$  に従うという。

証明 [TODO]

□

確率論には「**0-1 法則 (zero-one law)**」と呼ばれるいくつかの定理がある。次に述べる Borel-Cantelli の補題はそのひとつである。

**定理 1.1.4** (Borel-Cantelli の補題). [TODO]

証明 [TODO]

□

**定理 1.1.5** (Jensen の不等式).  $(\Omega, \mathcal{F}, P)$  を確率空間、 $A \subset \mathbb{R}^n$  を凸集合、 $\Psi: A \rightarrow \mathbb{R}$  を凸関数とする。このとき、 $\Psi(f)$  が  $P$ -可積分となるような任意の  $P$ -可積分関数  $f$  に対し

$$\Psi\left(\int_{\Omega} f(x) P(dx)\right) \leq \int_{\Omega} \Psi(f(x)) P(dx) \quad (1.1.2)$$

が成り立つ。

証明 [TODO] cf. [?, p.153]

□

## 1.2 離散確率分布

**定義 1.2.1** (離散確率分布).  $X$  を高々可算集合、 $\mathcal{A} := 2^X$  とするとき、可測空間  $(X, \mathcal{A})$  を **離散確率空間 (discrete probability space)** といい、 $(X, \mathcal{A})$  上の確率分布を **離散確率分布 (discrete probability distribution)** という。

**定義 1.2.2** (確率質量関数).  $(X, \mathcal{A})$  を離散確率空間、 $\mu$  を  $(X, \mathcal{A})$  上の確率分布とする。このとき、 $(X, \mathcal{A})$  上の数え上げ測度に関する  $\mu$  の Radon-Nikodym 微分<sup>1)</sup>、すなわち

$$\mu(E) = \sum_{x \in E} p(x) \quad (E \in \mathcal{A}) \quad (1.2.1)$$

なる関数  $p: X \rightarrow [0, +\infty]$  を  $\mu$  の **確率質量関数 (probability mass function; PMF)** という。

**定義 1.2.3** (確率母関数). [TODO]

## 1.3 連続確率分布

$\mathbb{R}$  上の確率分布は確率論や統計学において重要である。

**定義 1.3.1** (連続確率分布).  $\mathcal{B}$  を  $\mathbb{R}$  の Borel 集合族とすると、可測空間  $(\mathbb{R}, \mathcal{B})$  を **連続確率空間 (continuous probability space)** といい、 $(\mathbb{R}, \mathcal{B})$  上の確率分布を **連続確率分布 (continuous probability distribution)** という。

**定義 1.3.2** (絶対連続分布). [TODO]

**定義 1.3.3** (確率密度関数).  $(\mathbb{R}, \mathcal{B})$  を連続確率空間、 $\mu$  を  $(\mathbb{R}, \mathcal{B})$  上の Lebesgue 測度に関し絶対連続な確率分布とする。このとき、 $\mu$  の Radon-Nikodym 微分、すなわち

$$\mu(E) = \int_E p(x) dx \quad (E \in \mathcal{B}) \quad (1.3.1)$$

なる関数  $p: \mathbb{R} \rightarrow [0, +\infty]$  を  $\mu$  の **確率密度関数 (probability density function; PDF)** という。

**定義 1.3.4** (モーメント母関数). [TODO]

モーメント母関数と異なり、特性関数は常に存在する。

**定義 1.3.5** (特性関数). [TODO]

1) 数え上げ測度に関する  $\mu$  の Radon-Nikodym 微分はつねに存在する。なぜなら、 $(X, \mathcal{A})$  上のいかなる確率測度も  $(X, \mathcal{A})$  上の数え上げ測度に関して絶対連続だからである。

## 1.4 多変量分布

[TODO]

## 1.5 確率変数の収束

定理 1.5.1 (大数の法則). [TODO]

証明 [TODO]

□

定理 1.5.2 (中心極限定理). [TODO]

証明 [TODO]

□

## 第2章 推定

この章では、統計的推論の基本的な問題のひとつである推定について述べる。

### 2.1 十分統計量

標本 (= 確率変数) [TODO] 標本とは確率変数のことなのか?  $\Omega$  の元ではないのか? の実現値に対しその特徴を要約した値を割り当てる関数を統計量という。事象は統計量を通して「観測」される。もちろん標本それ自体も統計量である。

**定義 2.1.1** (統計量). [TODO]

統計量を用いて母集団 (= 確率分布) のパラメータを推定することを考える。

**定義 2.1.2** (点推定). 標本  $X$  から母集団のパラメータ  $\theta$  を特定することを**推定 (estimate)** といい、とくに一意に推定することを**点推定 (point estimation)** という。点推定において、標本  $X$  の実現値  $x$  に対し  $\theta$  の**推定値 (estimate)**  $\hat{\theta}(x)$  を割り当てる関数  $\hat{\theta}$  を  $\theta$  の**推定量 (estimator)** という。

**注意 2.1.3.** 今後、点推定のことを単に推定ということにする。

推定のために"十分"な情報を含んだ統計量を十分統計量という。

**定義 2.1.4** (十分統計量). [TODO]

**例 2.1.5** (Bernoulli 分布の例). [TODO]

次の定理はある統計量が十分統計量であるための必要十分条件を与え、具体的な判定に役立つ。

**定理 2.1.6** (Fisher-Neyman の分解定理). [TODO]

証明 [TODO]

□

**例 2.1.7** (正規分布の例). [TODO]

### 2.2 指数型分布族

指数型分布族について述べる。

**定義 2.2.1** (指数型分布族).  $(X, \mathcal{A})$  を可測空間、 $\mu$  を  $(X, \mathcal{A})$  上の  $\sigma$ -有限測度、 $\mathcal{P} = (P_\theta)_{\theta \in \Theta}$  を  $(X, \mathcal{A})$  上の確率

分布族とする。ここで、各  $P_\theta$  が  $\mu$  に関し絶対連続で、Radon-Nikodym 微分が

$$\frac{dP_\theta}{d\mu}(x) = g(x) \exp \left( \sum_{i=1}^m a_i(\theta) T_i(x) - \psi(\theta) \right) = g(x) \exp(a(\theta) \cdot T(x) - \psi(\theta)) \quad \mu\text{-a.e. } x \in \mathcal{X} \quad (2.2.1)$$

の形に表せるとき、 $\mathcal{P}$  を**指数型分布族 (exponential family)** という。ただし、 $T_i, g$  は  $(\mathcal{X}, \mathcal{A})$  上の可測関数、 $a_i, \psi$  は  $\Theta$  上の実数値関数である。とくに

$$\frac{dP_\theta}{d\mu}(x) = g(x) \exp \left( \sum_{i=1}^m \theta_i T_i(x) - \psi(\theta) \right) = g(x) \exp(\theta \cdot T(x) - \psi(\theta)) \quad \mu\text{-a.e. } x \in \mathcal{X} \quad (2.2.2)$$

の形を**正準形 (canonical form)** という<sup>2)</sup>。

以下に指数型分布族に関する具体例を挙げる。

例 2.2.2 (正規分布). [TODO]

命題 2.2.3. 指数型分布族の定義の  $T = (T_1, \dots, T_m)$  は  $\theta$  の十分統計量である。

証明 Fisher-Neyman の分解定理より従う。 □

## 2.3 混合型分布族

定義 2.3.1 (混合型分布族). [TODO]

## 2.4 点推定の手法

定義 2.4.1 (最尤推定). [TODO]

定義 2.4.2 (モーメント法). [TODO]

定義 2.4.3 (最小 2 乗法). [TODO]

## 2.5 推定量の性質

定義 2.5.1 (不偏性). [TODO]

定義 2.5.2 (一様最小分散不偏推定量; UMVUE). [TODO]

2) 指数型分布族は常に正準形で書くことができる。実際、 $\theta$  の代わりに  $a(\theta)$  をパラメータとすればよい。

## 2.6 推定量の評価

定義 2.6.1 (スコア関数). [TODO]

Fisher 情報量を定義する。

定義 2.6.2 (Fisher 情報量). [TODO]

不偏推定量の分散の下界を Fisher 情報量の言葉で与えるのが Cramer-Rao 不等式である。[TODO] Cramer-Rao 不等式は Fisher "計量" に対しどのような意味を持つ？

命題 2.6.3 (Cramer-Rao 不等式). [TODO]

証明 [TODO]

□

定義 2.6.4 (有効性). [TODO]

## 2.7 漸近理論

漸近理論について述べる。一般に最尤推定量は有効性をみたすとは限らないが、漸近的には有効性をみたすことが知られている<sup>1)</sup>。

[TODO]

---

1) ただし Neyman-Scott 問題においては最尤推定量は漸近有効とは限らない。

## 第 3 章 仮説検定

この章では、統計的推論の基本的な問題のひとつである仮説検定について述べる。

### 3.1 仮説検定

[TODO]