確率統計

Probability and Statistics

第3回講義資料

Lecture notes 3

多変数確率分布

Multi-variate Probability Distributions

豊橋技術科学大学

Toyohashi University of Technology

電気・電子情報工学系

Department of Electrical and Electronic Information Engineering

准教授　竹内啓悟

Associate Professor　Keigo Takeuchi

がん検診(Cancer screening)

がんである(ない)という事象を$\omega_{11}(\omega_{12})$、検診の結果が陽性(陰性)であるという事象をω_{21} (ω_{22})とする。

Let ω_{11} (ω_{12}) and ω_{21} (ω_{22}) denote events indicating that a person has a (no) cancer, and that an examination alerts positive (negative).

確率空間(Probability space) $(\Omega, \mathcal{F}, P)$

標本空間(Sample space)： $\Omega = \Omega_1 \times \Omega_2, \qquad \Omega_i = \{\omega_{i1}, \omega_{i2}\}$

σ-集合体(σ-field)：

$$
\begin{aligned}
\mathcal{F} = \{&\emptyset, (\omega_{11}, \omega_{21}), (\omega_{11}, \omega_{22}), (\omega_{12}, \omega_{21}), (\omega_{12}, \omega_{22}),\\
&\{(\omega_{11}, \omega_{21}), (\omega_{11}, \omega_{22})\}, \{(\omega_{11}, \omega_{21}), (\omega_{12}, \omega_{21})\},\\
&\{(\omega_{11}, \omega_{21}), (\omega_{12}, \omega_{22})\}, \{(\omega_{11}, \omega_{22}), (\omega_{12}, \omega_{21})\},\\
&\{(\omega_{11}, \omega_{22}), (\omega_{12}, \omega_{22})\}, \{(\omega_{12}, \omega_{21}), (\omega_{12}, \omega_{22})\},\\
&\{(\omega_{11}, \omega_{21}), (\omega_{11}, \omega_{22}), (\omega_{12}, \omega_{21})\},\\
&\{(\omega_{11}, \omega_{21}), (\omega_{11}, \omega_{22}), (\omega_{12}, \omega_{22})\},\\
&\{(\omega_{11}, \omega_{21}), (\omega_{12}, \omega_{21}), (\omega_{12}, \omega_{22})\},\\
&\{(\omega_{11}, \omega_{22}), (\omega_{12}, \omega_{21}), (\omega_{12}, \omega_{22})\}, \Omega\}
\end{aligned}
$$

確率分布(Probability distribution)：P

同時分布(Joint distribution)：

$P(\omega_{11}, \omega_{21}) = pq$ 「がんである」かつ「陽性である」(“w cancer” and “positive”)

$P(\omega_{11}, \omega_{22}) = p(1-q)$ 「がんである」かつ「陰性である」(“w cancer” and “negative”)

$P(\omega_{12}, \omega_{21}) = r - pq$ 「がんでない」かつ「陽性である」(“w/o cancer” and “positive”)

「がんでない」かつ「陰性である」(“w/o cancer” and “negative”)

$$P(\omega_{12}, \omega_{22}) = P(\Omega) - P(\omega_{11}, \omega_{21}) - P(\omega_{11}, \omega_{22}) - P(\omega_{12}, \omega_{21})$$
$$= 1 - pq - p(1-q) - (r - pq) = 1 - r - p + pq.$$

周辺分布(Marginal distribution)：

「がんであるかつ陽性」または「がんであるかつ陰性」＝「がんである」

“w cancer and positive” or “w cancer and negative” = “w cancer”

$$P(\omega_{11}, \Omega_2) \overset{\text{def}}{=} P(\{(\omega_{11}, \omega_{21}), (\omega_{11}, \omega_{22})\}) = P\big((\omega_{11}, \omega_{21}) \cup (\omega_{11}, \omega_{22})\big)$$
$$= pq + p(1-q) = p.$$

条件付き分布(Conditional distribution)：

「がんである」という条件の下で「陽性である」(“positive” conditioned on “w cancer”)

$$P(\omega_{21} | (\omega_{11}, \Omega_2)) \overset{\text{def}}{=} \frac{P(\omega_{11}, \omega_{21})}{P(\omega_{11}, \Omega_2)} = \frac{pq}{p} = q.$$

独立性(Independence)

周辺分布(Marginal distribution)：

「がんであるかつ陽性」または「がんでないかつ陽性」＝「陽性である」

"w cancer and positive" or "w/o cancer and positive" = "positive"

$$P(\Omega_1, \omega_{21}) \stackrel{\text{def}}{=} P(\{(\omega_{11}, \omega_{21}), (\omega_{12}, \omega_{21})\}) = P\big((\omega_{11}, \omega_{21}) \cup (\omega_{12}, \omega_{21})\big)$$

$$= pq + r - pq \ = r.$$

独立性(Independence)：

「陽性である」確率が、「がんである」という条件に依存いない。

The probability of "positive alert" is independent of conditioning "w cancer."

$$P(\omega_{21} | (\omega_{11}, \Omega_2)) = P(\Omega_1, \omega_{21}).$$

$$\Leftrightarrow P(\omega_{11}, \omega_{21}) = P(\omega_{11}, \Omega_2) P(\Omega_1, \omega_{21})$$

$r = q$はこの場合における独立性の必要十分条件である。

The independence in this case holds if and only if $r = q$.

がんであろうがなかろうが、同じ確率で陽性と診断される検査を受ける意味はない。

Regardless of whether or not a person has a cancer, it makes no sense for him/her to take an examination that alerts "positive" with an identical probability.

同時確率分布関数(Joint probability distribution function)

$\mathbb{R}^n$上の確率空間(Probability space on $\mathbb{R}^n$) $(\Omega = \mathbb{R}^n, \mathcal{F}, P)$

確率ベクトル(Probability vector)：$\boldsymbol{X}(\boldsymbol{\omega}) = \left(X_1(\boldsymbol{\omega}), \dots, X_n(\boldsymbol{\omega})\right)^{\mathrm{T}}$

任意のベクトル$\boldsymbol{x} = (x_1, \dots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$に対して、以下を満たす実ベクトル値関数$\boldsymbol{X}$を確率ベクトルと呼ぶ。

A real vector-valued function $\boldsymbol{X}$ is called a random vector if the following holds for any vector $\boldsymbol{x} = (x_1, \dots, x_n)^{\mathrm{T}} \in \mathbb{R}^n$:

$$\{\boldsymbol{\omega} \in \Omega : X_1(\boldsymbol{\omega}) \le x_1, \dots, X_n(\boldsymbol{\omega}) \le x_n\} \subset \mathcal{F}.$$

同時確率分布関数(Joint probability distribution function)

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P(\{\boldsymbol{\omega} \in \Omega : X_1(\boldsymbol{\omega}) \le x_1, \dots, X_n(\boldsymbol{\omega}) \le x_n\})$$

同時確率分布関数は同時分布関数あるいは同時分布とも呼ばれ、確率ベクトル$\boldsymbol{X}$の統計的性質のすべてを決定する。

The joint probability distribution function is also called joint distribution function or joint distribution, and determines all statistical properties of the random vector $\boldsymbol{X}$.

略記(Abbreviation)

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = P(X_1 \le x_1, \dots, X_n \le x_n) \overset{\text{def}}{=} P(\boldsymbol{X} \le \boldsymbol{x}).$$

周辺確率分布関数(Marginal probability distribution functions)

番号集合の部分集合$I \subset \{1, \dots, n\}$に対して、確率ベクトル$\boldsymbol{X}$を二つのグループに分割する。

For a subset $I \subset \{1, \dots, n\}$ of the set of indices, decompose the random vector $\boldsymbol{X}$ into disjoint two groups.

$$\boldsymbol{X}_I = \{X_i : i \in I\}, \quad \boldsymbol{X}_{I^c} = \{X_i : i \in I^c\}.$$

ベクトル$\boldsymbol{x}_I \in \mathbb{R}^{|I|}$も同様に定義する。

The vector $\boldsymbol{x}_I \in \mathbb{R}^{|I|}$ is also defined in the same manner.

周辺確率分布関数(Marginal probability distribution functions)

$$P_{\boldsymbol{X}_I}(\boldsymbol{x}_I) \overset{\mathrm{def}}{=} P(\{X_i \le x_i : i \in I\}) = P_X(\{\boldsymbol{x} \in \mathbb{R}^n : x_i = \infty \text{ for } i \in I^c\}).$$

周辺分布は同時分布から計算できるが、逆はできない。

Any marginal distribution can be computed from the joint distribution, while the converse is not correct.

同時分布から周辺分布を計算する操作のことを「周辺化」と呼ぶことがある。

We may refer to operations to compute marginal distributions from the joint distribution as "marginalization."

独立性(Independence)

組ごとの独立性(Pairwise Independence)

$$P_{X_i,X_j}(x_i,x_j) = P_{X_i}(x_i)P_{X_j}(x_j) \text{ for all } i \neq j.$$

独立性(Independence)

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} P_{X_i}(x_i)\,.$$

独立性は組ごとの独立性を示唆するが、逆は必ずしも正しくない。

Independence implies pairwise independence, while the converse is not necessarily correct.

∵ 独立性の定義式をX_iとX_jを除いて周辺化すると、

Marginalizing the definition of independence over all random variables with the exception of X_i and X_j, we have

$$P_{X_i,X_j}(x_i,x_j) = \lim_{\{x_k\to\infty : k\neq i,j\}} P_{\boldsymbol{X}}(\boldsymbol{x}) = P_{X_i}(x_i)P_{X_j}(x_j)\prod_{k\neq i,j}\lim_{x_k\to\infty} P_{X_k}(x_k)$$

$$= P_{X_i}(x_i)P_{X_j}(x_j).$$

離散分布(Discrete distributions)

確率ベクトル$\boldsymbol{X}$の同時分布が高々可算個の点でのみ増加するとき、$\boldsymbol{X}$を離散確率ベクトルと呼ぶ。

A random vector $\boldsymbol{X}$ is called discrete if the joint distribution of $\boldsymbol{X}$ increases only in the set of at most countable points.

周辺確率分布(Marginal probability distribution)

離散確率ベクトル$\boldsymbol{X}$の同時分布$P(\boldsymbol{X}=\boldsymbol{x})$に対して、周辺分布$P(\boldsymbol{X}_I=\boldsymbol{x}_I)$は以下で与えられる。

For a joint distribution $P(\boldsymbol{X}=\boldsymbol{x})$ of a discrete random vector $\boldsymbol{X}$, the marginal distribution $P(\boldsymbol{X}_I=\boldsymbol{x}_I)$ is given by

$$P(\boldsymbol{X}_I=\boldsymbol{x}_I)=\sum_{\boldsymbol{x}_{I^c}} P(\boldsymbol{X}=\boldsymbol{x}),$$

ここで、総和は$\boldsymbol{x}_{I^c}$のすべての組み合わせに関して取られる。

where the summation is over all possible combinations of $\boldsymbol{x}_{I^c}$.

∵ 周辺分布関数の定義から従う。(Due to the definition of marginal distribution functions.)

この総和操作のことを「周辺化」と呼ぶことがある。

This summation operation may be called “marginalization.”

条件付き確率分布(Conditional probability distributions)

For $P(\boldsymbol{X}_{I^c} = \boldsymbol{x}_{I^c}) \neq 0,\quad P(\boldsymbol{X}_I | \boldsymbol{X}_{I^c} = \boldsymbol{x}_{I^c}) \stackrel{\text{def}}{=} \dfrac{P(\boldsymbol{X}_I, \boldsymbol{X}_{I^c} = \boldsymbol{x}_{I^c})}{P(\boldsymbol{X}_{I^c} = \boldsymbol{x}_{I^c})}.$

注意1(Remark 1)

この定義は離散分布のみに対して意味を持つ。

This definition makes sense only for discrete distributions.

注意2(Remark 2)

周辺分布$P(\boldsymbol{X}_{I^c})$と条件付き分布$P(\boldsymbol{X}_I | \boldsymbol{X}_{I^c})$が与えられたときの同時分布$P(\boldsymbol{X})$の定義式ともみなせる。

It can be regarded as a definition of the joint distribution $P(\boldsymbol{X})$ when marginal and conditional distributions $P(\boldsymbol{X}_{I^c})$ and $P(\boldsymbol{X}_I | \boldsymbol{X}_{I^c})$ are given.

$$P(\boldsymbol{X}_I | \boldsymbol{X}_{I^c}) P(\boldsymbol{X}_{I^c}) = P(\boldsymbol{X}).$$

応用上は、周辺分布と条件付き分布が与えられることが多い。

Marginal and conditional distributions are usually given in practice.

二項分布(Binomial distribution)

$\{X_i\}_{i=1}^n$を独立同一分布($P(X_i=1)=p, P(X_i=0)=1-p$)に従うベルヌーイ確率変数列とする。確率変数$Y=\sum_{i=1}^n X_i$が二項分布に従うことを示せ。

Let $\{X_i\}_{i=1}^n$ denote independent and identically distributed (i.i.d.) Bernoulli random variables. Show that $Y=\sum_{i=1}^n X_i$ is binomial-distributed.

$$P(Y=k)=\sum_{\boldsymbol{x}\in\{0,1\}^n:\sum_{i=1}^n x_i=k} P(\boldsymbol{X}=\boldsymbol{x}) = \sum_{\boldsymbol{x}\in\{0,1\}^n:\sum_{i=1}^n x_i=k} p^k(1-p)^{n-k}.$$

最後は、1がk回、0が$n-k$回出る確率は$p^k(1-p)^{n-k}$であることから従う。

The last follows from the fact that k ones and $n-k$ zeros occur with probability $p^k(1-p)^{n-k}$.

総和における組み合わせ数を数え上げると、

Counting the number of combinations in the summation, we obtain

$$P(Y=k)=\binom{n}{k}p^k(1-p)^{n-k}.$$

周辺化には膨大な数の和の計算が必要なことを理解せよ。

Understand the requirement of a huge number of additions in marginalization.

連続分布(Continuous distributions)

確率ベクトル$\boldsymbol{X}$の同時分布が同時確率密度関数$p_{\boldsymbol{X}}(\boldsymbol{x})$を使って以下で表現できるとき、$\boldsymbol{X}$を連続確率ベクトルと呼ぶ。

A random vector $\boldsymbol{X}$ is called continuous if the joint distribution of $\boldsymbol{X}$ has the following representation based on a joint probability density function (pdf) $p_{\boldsymbol{X}}(\boldsymbol{x})$:

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{\cap_{i=1}^{n}(-\infty, x_i]} p_{\boldsymbol{X}}(\boldsymbol{y}) d\boldsymbol{y} \quad \text{for all } \boldsymbol{x} \in \mathbb{R}^n.$$

性質1(Property 1) $$p_{\boldsymbol{X}}(\boldsymbol{x}) \geq 0.$$

性質2(Property 2) $$\int_{\mathbb{R}^n} p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x} = 1.$$

周辺確率密度関数(Marginal pdf)

$$P_{\boldsymbol{X}_I}(\boldsymbol{x}_I) = \int_{\cap_{i \in I}(-\infty, x_i]} p_{\boldsymbol{X}_I}(\boldsymbol{y}_I) d\boldsymbol{y}_I \quad \text{for all } x_I \in \mathbb{R}^{|I|}.$$

性質3(Property 3)

$$p_{\boldsymbol{X}_I}(\boldsymbol{x}_I) = \int_{\mathbb{R}^{|I^c|}} p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}_{I^c} .$$

∵ 周辺分布関数の定義から従う。(Due to the definition of marginal distribution functions.)

この積分操作のことを「周辺化」と呼ぶことがある。

This integration operation may be called “marginalization.”

条件付き確率密度関数(Conditional pdfs)

For $p_{\boldsymbol{X}_{I^c}}(\boldsymbol{x}_{I^c}) \neq 0$, $\quad p_{\boldsymbol{X}_I|\boldsymbol{X}_{I^c}}(\boldsymbol{x}_I|\boldsymbol{x}_{I^c}) \overset{\text{def}}{=} \dfrac{p_{\boldsymbol{X}}(\boldsymbol{x})}{p_{\boldsymbol{X}_{I^c}}(\boldsymbol{x}_{I^c})}$.

独立性(Independence) 連続確率ベクトル$\boldsymbol{X}$の要素が独立なとき、

When all elements of a continuous random vector $\boldsymbol{X}$ are independent,

$$p_{\boldsymbol{X}}(\boldsymbol{x}) = \prod_{i=1}^{n} p_{X_i}(x_i), \quad p_{\boldsymbol{X}_I|\boldsymbol{X}_{I^c}}(\boldsymbol{x}_I|\boldsymbol{x}_{I^c}) = \prod_{i\in I} p_{X_i}(x_i).$$

∵ 独立性の定義から、(From the definition of independence, we have)

$$P_{\boldsymbol{X}}(\boldsymbol{x}) = \int_{\cap_{i=1}^{n}(-\infty,x_i]} p_{\boldsymbol{X}}(\boldsymbol{y})d\boldsymbol{y} = \prod_{i=1}^{n} P_{X_i}(x_i) = \prod_{i=1}^{n} \int_{-\infty}^{x_i} p_{X_i}(y_i)dy_i.$$

両辺を$x_i (i = 1, \dots, n)$に関して偏微分することで、前半の主張を得る。

The former is obtained via the partial derivatives of both sides with respect to x_i for $i = 1, \dots, n$.

条件付き確率密度関数の定義式に前半の主張を代入すると、後半の主張が従う。

Substituting the former statement into the definition of the conditional pdf yields the latter statement.

期待値(Expectation)

関数$g: \mathbb{R}^n \to \mathbb{R}$に対して$\mathbb{E}[|g(\boldsymbol{X})|] < \infty$のとき、

離散確率ベクトル(Discrete random vector) X

$$\mathbb{E}[g(\boldsymbol{X})] \stackrel{\text{def}}{=} \sum_{\boldsymbol{x}} g(\boldsymbol{x}) P(\boldsymbol{X} = \boldsymbol{x}),$$

ここで、総和は$\boldsymbol{x}$のすべての組み合わせに関して取られる。

where the summation is over all possible combinations of $\boldsymbol{x}$.

連続確率ベクトル(Continuous random vector) X

$$\mathbb{E}[g(\boldsymbol{X})] \stackrel{\text{def}}{=} \int_{\mathbb{R}^n} g(\boldsymbol{x}) p_{\boldsymbol{X}}(\boldsymbol{x}) d\boldsymbol{x}.$$

特に、$\mathbb{E}[X_i]$、$\mathbb{V}[X_i] \stackrel{\text{def}}{=} \mathbb{E}[(X_i - \mathbb{E}[X_i])^2]$、$\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$をそれぞれ$X_i$の平均、$X_i$の分散、$i \neq j$に対して$X_i$と$X_j$の共分散と呼ぶ。

In particular, $\mathbb{E}[X_i]$, $\mathbb{V}[X_i] \stackrel{\text{def}}{=} \mathbb{E}[(X_i - \mathbb{E}[X_i])^2]$, and $\mathbb{E}[(X_i - \mathbb{E}[X_i])(X_j - \mathbb{E}[X_j])]$ are called the mean of X_i, the variance of X_i, and the covariance of X_i and X_j for $i \neq j$, respectively.

条件付き期待値(Conditional expectation)

関数$g\colon \mathbb{R}^{|I|} \to \mathbb{R}$に対して$\mathbb{E}[|g(\boldsymbol{X}_I)|] < \infty$のとき、

離散確率ベクトル(Discrete random vector) X

$$\mathbb{E}[g(\boldsymbol{X}_I)|\boldsymbol{X}_{I^c}] \stackrel{\text{def}}{=} \sum_{\boldsymbol{x}_I} g(\boldsymbol{x}_I) P(\boldsymbol{X}_I = \boldsymbol{x}_I | \boldsymbol{X}_{I^c}),$$

ここで、総和は$\boldsymbol{x}_I$のすべての組み合わせに関して取られる。

where the summation is over all possible combinations of $\boldsymbol{x}_I$.

連続確率ベクトル(Continuous random vector) X

$$\mathbb{E}[g(\boldsymbol{X}_I)|\boldsymbol{X}_{I^c}] \stackrel{\text{def}}{=} \int_{\mathbb{R}^{|I|}} g(\boldsymbol{x}_I) p_{\boldsymbol{X}_I|\boldsymbol{X}_{I^c}}(\boldsymbol{x}_I) d\boldsymbol{x}_I .$$

計算上は、条件として与えられた確率変数を決定論的な変数とみなして期待値を計算すればよい。

In computation, we can regard conditioned random variables as deterministic variables.

TOYOHASHI UNIVERSITY OF TECHNOLOGY