https://doi.org/10.20965/jaciii.2019.p0715

Letter:

Index-Based Notation for Random Variable and Probability Space

Hiroki Shibata and Yasufumi Takama[†]

Graduate School of System Design, Tokyo Metropolitan University
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan
E-mail: ytakama@tmu.ac.jp
[†]Corresponding author

[Received October 1, 2018; accepted February 12, 2019]

In a conventional notation used in many studies, a probability space and state space of random variables is identified by its symbol. However, such a notation makes a formula ambiguous in a large equation. This letter proposes to use an index set to identify the probability space and state space of random variables. It is shown that the proposed notation can increase the generality of formulas without ambiguity.

Keywords: probability distribution, random variable, probability space, generative model, mathematical notation

1. Introduction

Generative models have been used in many domains, which include Boltzmann Machines (BM) [1, 2], Bayesian networks [3], and Markov random fields [4]. These models use a probability theory and its notation. However, all these works identify the probability space and state space by the symbol of random variables, which loses generality and strictness in mathematical analysis. Notation is important to analyze models as shown in the history of science. For example, Einstein proposed the notation called Einstein summation convention to simplify the formulas of tensors [5]. Newton introduced the notation of a differential equation to represent the laws of motion [6]. This letter proposes to use an index set to identify the probability space and state space of random variables. This letter takes several frequently used formulas and shows that proposed notation can represent those formulas in general and unambiguous forms.

In this letter, symbols used in formulas follow the standard of ISO 80000-2:2009 [7]. The concept and notation for set theory follow those introduced in [8]. This letter uses the concept of family [8] to clarify the definition and algebra of multi-dimensional variable. Family is the key concept of the proposed notation. If a variable x is a family defined on an index set Λ, it is defined as $x = (x_i | i \in \Lambda)$.[1] If there is a family of sets $X := (X_i | i \in \Lambda \wedge x_i \in X_i)$, the proposition $x \in \prod_{i\in\Lambda} X_i$ holds. The formula $\prod_{i\in\Lambda} X_i$ is called the direct product set of X, and if all terms of X is equal to Y, it can be written as $Y^{|\Lambda|}$, where $|\Lambda|$ denotes the number of elements in Λ. For example, in a 3-dimensional Euclidean space, let $\Lambda = \{\mathrm{x}, \mathrm{y}, \mathrm{z}\}$ denotes the index set representing axes. Its direct product set is defined as $\mathbb{R} \times \mathbb{R} \times \mathbb{R}$ or $\mathbb{R}^3$, and a variable a that belongs to this space can be defined as $a := (a_\mathrm{x}, a_\mathrm{y}, a_\mathrm{z}) \in \mathbb{R}^3$.

A notation $A = B$ for sets A and B is equivalent to the proposition $A \subset B \wedge B \subset A$. A notation $x = y$ for families x and y defined on the same index set Λ is equivalent to the proposition $\forall i \in \Lambda, x_i = y_i$.

2. Conventional Notation on Probability Theory

A probability space consists of 3 components: a sample space Ω, a set of events S, and a probability measure $P : S \to [0, 1]$ [9]. This letter deals with only the case that Ω is a countable set. In this case, S consists of all subsets of Ω. For a probability measure P, $\forall A, B \in S, A \cap B = \emptyset \Rightarrow P(A \cup B) = P(A) + P(B)$ and $P(\Omega) = 1$ hold.

Random variable is a surjective map of $\Omega \to \Omega'$, where Ω' is a state space. The probability of a random variable $\mathbf{x} : \Omega \to \Omega'$ taking x $(\in \Omega')$ can be represented as $P(\mathbf{x}^{-1}(x))$; $\mathbf{x}^{-1}$ is the inverse map of $\mathbf{x}$. The probability $P(\mathbf{x}^{-1}(x))$ is often represented as $p(\mathbf{x} = x)$ [10]; p is called a probability distribution of $\mathbf{x}$. Many studies [1, 2, 11] have used bold Roman fonts for random variables to distinguish it from its realization values in Ω'. Realization value x is often omitted; the value of the probability $\mathbf{x}$ is represented as $p(\mathbf{x})$. Components of a probability space can be retrieved using the probability distribution and the inverse map of a random variable defined on them. Therefore, usually those are not explicitly defined in a formulation.

One of the problems of this notation is that whether $\mathbf{x}$ is a variable of one-dimension or multi-dimension cannot be identified.

Conventionally, a probability distribution is identified by the argument random variable of its function [1, 10, 11] and the same symbol such as p is used for different

1. This notation is equal to $(x_{i_1}, x_{i_2}, \ldots, x_{i_n})$; $\Lambda = \{i_j | j = 1, 2, ..., n\}$.

probability distribution. Therefore, for random variables $\mathbf{x}$ and $\mathbf{y}$, of which state spaces are the same, Eq. (1) does not always hold because of this convention.

$$\mathbf{x} = \mathbf{y} \Rightarrow p(\mathbf{x}) = p(\mathbf{y}). \quad (1)$$

That is, the distribution of $\mathbf{x}$ and $\mathbf{y}$ are different even if $\mathbf{x}$ and $\mathbf{y}$ take the same value. For example, with $p(\mathbf{x}=1) = 0.3,\ p(\mathbf{x}=2) = 0.7,\ p(\mathbf{y}=1) = 0.4,\ p(\mathbf{y}=2) = 0.6$, Eq. (1) does not hold. Namely, probability distributions are not identified without specifying its arguments.

As for a summation in terms of arbitrary function $f : \Omega' \to \mathbb{R}$ on a random variable $\mathbf{x}$ of which state space is Ω', the notation $\sum_{\mathbf{x}} f(\mathbf{x})$ is often used to simplify the representation of $\sum_{x \in \Omega'} f(x)$ [1, 11]. However, this notation identifies the state space used in summation with the symbol of a random variable that has its state space. Therefore, for random variables $\mathbf{x}$ and $\mathbf{y}$, Eq. (2) does not usually hold without additional note that $\mathbf{y}$ is a random variable of which its state space is the same as $\mathbf{x}$.

$$\sum_{\mathbf{x}} f(\mathbf{x}) \neq \sum_{\mathbf{y}} f(\mathbf{y}). \quad (2)$$

Because of these problems in Eqs. (1) and (2), the conditional distribution of $\mathbf{x}$ given $\mathbf{y}$ cannot be written as Eq. (3) .

$$p(\mathbf{x}|\mathbf{y}) = \frac{p(\mathbf{x},\mathbf{y})}{\sum_{\mathbf{x}} p(\mathbf{x},\mathbf{y})}. \quad (3)$$

In Eq. (3) , $\mathbf{x}$ in the denominator of the right side cannot be identified from $\mathbf{x}$ in the left side. Furthermore, another symbol cannot be used instead of $\mathbf{x}$ to represent the denominator because of the property related to Eqs. (1) and (2). To avoid this problem, a conditional distribution is usually written as Eq. (4) . This restriction makes a problem when we consider to reduce the common factors in the both denominator and numerator.

$$\begin{aligned} p(\mathbf{x}|\mathbf{y}) &= \frac{p(\mathbf{x},\mathbf{y})}{p(\mathbf{y})}, \\ p(\mathbf{y}) &= \sum_{\mathbf{x}} p(\mathbf{x},\mathbf{y}). \end{aligned} \quad (4)$$

Equation (3) can be written as Eq. (5) with the realization values of $x \in \Omega'_{\mathrm{x}}$ and $y \in \Omega'_{\mathrm{y}}$, where Ω'_{x} is the state space of $\mathbf{x}$, and Ω'_{y} for $\mathbf{y}$ [9, 10]. However, it is redundant and loses simplicity introduced in the notation by omitting realization values and state spaces.

$$p(\mathbf{x}=x|\mathbf{y}=y) = \frac{p(\mathbf{x}=x,\mathbf{y}=y)}{\sum_{z \in \Omega'_{\mathrm{x}}} p(\mathbf{x}=z,\mathbf{y}=y)}. \quad (5)$$

When $\mathbf{x}$ is a D-dimensional variable as $\mathbf{x} = (\mathbf{x}_1, \ldots, \mathbf{x}_D)$,[2] the conditional distribution of an element $\mathbf{x}_i$ given the other elements of $\mathbf{x}$ is often written as Eq. (6) [12].

$$p(\mathbf{x}_i|\mathbf{x}_{-i}) = \frac{p(\mathbf{x})}{p(\mathbf{x}_{-i})}. \quad (6)$$

2. It can be also defined as a family with the index set $\{1,2,\ldots,D\}$.

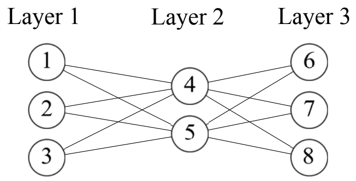

Fig. 1. Example of a 3-layer BM with indices for its units.

This representations lose generality when handling conditional distributions of more variables because all the variables should be written independently. When formulating a generative model like BM, different multi-dimensional variables are defined for different layers. For example, the distribution of BM described in **Fig. 1** is defined as Eq. (7) using three multi-dimensional variables $\mathbf{x} = (\mathbf{x}_1, \mathbf{x}_2, \mathbf{x}_3)$, $\mathbf{y} = (\mathbf{y}_1, \mathbf{y}_2)$, $\mathbf{z} = (\mathbf{z}_1, \mathbf{z}_2, \mathbf{z}_3)$ [1]. For simplicity, this formulation omits bias terms.

$$\begin{aligned} p(\mathbf{x},\mathbf{y},\mathbf{z}) &= \frac{1}{Z_{\mathrm{p}}} \exp(-\phi(\mathbf{x},\mathbf{y},\mathbf{z})), \\ Z_{\mathrm{p}} &= \sum_{\mathbf{x}} \sum_{\mathbf{y}} \sum_{\mathbf{z}} \exp(-\phi(\mathbf{x},\mathbf{y},\mathbf{z})), \\ \phi(\mathbf{x},\mathbf{y},\mathbf{z}) &= -\sum_{i=1}^{3} \sum_{j=1}^{2} \mathbf{x}_i w_{ij} \mathbf{y}_j \\ &\quad - \sum_{j=1}^{2} \sum_{k=1}^{3} \mathbf{y}_j v_{jk} \mathbf{z}_k, \end{aligned} \quad (7)$$

where Z_{p} is a partition function and ϕ is an energy function, w and v are weight coefficients between units respectively.

Obviously, the scalability of this formulation is low, because it becomes more complicated when it has more layers.

Furthermore, as the summation terms are separately established for each layer, the conditional distribution of a certain unit's state being 1 given the others' states has to be written separately as Eq. (8) even though it takes almost the same form.[3]

$$\begin{aligned} p(\mathbf{x}_i = 1|\mathbf{y}) &= f_\sigma\left(\sum_{j=1}^{2} w_{ij}\mathbf{y}_j\right), \quad i = 1,2,3, \\ p(\mathbf{y}_j = 1|\mathbf{x},\mathbf{z}) &= f_\sigma\left(\sum_{i=1}^{3} \mathbf{x}_i w_{ij} + \sum_{k=1}^{3} v_{jk}\mathbf{z}_k\right), \\ &\qquad j = 1,2, \\ p(\mathbf{z}_k = 1|\mathbf{y}) &= f_\sigma\left(\sum_{j=1}^{2} \mathbf{y}_j v_{jk}\right), \quad k = 1,2,3, \end{aligned} \quad (8)$$

where $f_\sigma : x \mapsto 1/(1+\exp(-x)), x \in \mathbb{R}$ is the logistic function.

The problems mentioned in this section are caused by the rule that identifies a probability space and a state

3. This formulation of BM assumes that its units' state space is $\{0,1\}$, which is called Bernoulli BM.

space by the symbol of a random variable as mentioned at Eqs. (1) and (2).

3. Proposed Notation

This letter proposes to use an index set to identify the probability space and state space of random variables with the notation of set theory. For that purpose, it also extends the concept of family as followings.

1. The notation $\mathbf{x}_\Lambda$ denotes a family defined on an index set Λ: $\mathbf{x}_\Lambda := (\mathbf{x}_i | i \in \Lambda)$. Such variables belong to the same direct product set defined on Λ. At the same time, it corresponds to a state space if $\mathbf{x}_\Lambda$ is a random variable.
2. An index set of a random variable identifies the probability space to which the random variable belong. However, for the convenience, when only their probability measures are different, it is represented by using different symbols for those probability distribution functions, e.g., $p(\mathbf{x}_\Lambda)$ and $q(\mathbf{x}_\Lambda)$.
3. A subset of index set that defines a random variable also composes a probability space and state space, which relate to its original probability space. For the convenience, a random variable on an index set of which size is 1 can be represented by its element: $\mathbf{x}_{\{\lambda\}}$ has the same meanings as $\mathbf{x}_\lambda$.

4. Application of Proposed Notation

Letting Λ and Γ are index sets, the following properties are derived.

Whereas conventional notation does not explicitly express the dimension of a random variable, the dimension of $\mathbf{x}_\Lambda$ is obviously $|\Lambda|$, based on the item 1 in Section 3.

Equation (9) holds for any random variables $\mathbf{x}_\Lambda$ and $\mathbf{y}_\Lambda$, based on the item 2 in Section 3. By replacing q with p, it corresponds to Eq. (1) . Note that the proposition $q = p$ is always true when q and p are represented by the same symbol.

$$(q = p) \wedge (\mathbf{x}_\Lambda = \mathbf{y}_\Lambda) \Rightarrow q(\mathbf{x}_\Lambda) = p(\mathbf{y}_\Lambda). \quad \ldots \quad (9)$$

As for summation mentioned at Eq. (2) , Eq. (10) also holds because the random variables indexed by the same index set have the same state space regardless of their symbols. It is based on the item 1 in Section 3.

$$\Lambda = \Gamma \Rightarrow \sum_{\mathbf{x}_\Lambda} f(\mathbf{x}_\Lambda) = \sum_{\mathbf{y}_\Lambda} f(\mathbf{y}_\Lambda). \quad \ldots \quad (10)$$

Using the property that Eqs. (9) and (10) hold, the conditional distribution of $\mathbf{x}_\Lambda$ given $\mathbf{y}_\Gamma$ is written as Eq. (11) in an unambiguous way, without dividing a formula like Eq. (4) .

$$p(\mathbf{x}_\Lambda | \mathbf{y}_\Gamma) = \frac{p(\mathbf{x}_\Lambda, \mathbf{y}_\Gamma)}{\sum_{\mathbf{z}_\Lambda} p(\mathbf{z}_\Lambda, \mathbf{y}_\Gamma)}. \quad \ldots \quad (11)$$

The conditional distribution of $\mathbf{x}_i$, where $i \in \Lambda$, which corresponds to Eq. (6) , is written as Eq. (12) based on the item 3 in Section 3.

$$p\left(\mathbf{x}_i | \mathbf{x}_{\Lambda\setminus\{i\}}\right) = \frac{p(\mathbf{x}_i)}{\sum_{\mathbf{y}_i} p\left(\mathbf{y}_i, \mathbf{x}_{\Lambda\setminus\{i\}}\right)}. \quad \ldots \quad (12)$$

For $M \subset \Lambda$, the conditional distribution of $\mathbf{x}_M$ given $\mathbf{x}_{\Lambda\setminus M}$ can be written as Eq. (13) based on the item 3 in Section 3. Note that when $M = \{i\}$, it corresponds to Eq. (12) . By replacing M with Λ, and $\Lambda \setminus M$ with Γ, it also corresponds to Eq. (11) .

$$p\left(\mathbf{x}_M | \mathbf{x}_{\Lambda\setminus M}\right) = \frac{p(\mathbf{x}_\Lambda)}{\sum_{\mathbf{y}_M} p\left(\mathbf{y}_M, \mathbf{x}_{\Lambda\setminus M}\right)}. \quad \ldots \quad (13)$$

The distribution of BM described in **Fig. 1** is written as Eq. (14) .

$$\begin{aligned} p(\mathbf{x}_\Lambda) &= \frac{1}{Z_\mathrm{p}} \exp(-\phi(\mathbf{x}_\Lambda)), \\ \phi(\mathbf{x}_\Lambda) &= -\sum_{(i,j)\in c^{-1}(\mathscr{E})} \mathbf{x}_i w_{c(i,j)} \mathbf{x}_j, \quad \ldots \quad (14) \end{aligned}$$

where $\mathscr{V}_1 = \{1,2,3\}, \mathscr{V}_2 = \{4,5\}, \mathscr{V}_3 = \{6,7,8\}, \Lambda = \bigcup_{l=1}^{3} \mathscr{V}_l$. $\mathscr{E}$ is the index set of weight coefficient w, and $c : \Lambda \times \Lambda \to \mathscr{E}$ is the surjective map that satisfies $c^{-1}(\mathscr{E}) = \bigcup_{l=1}^{2} \mathscr{V}_l \times \mathscr{V}_{l+1}$. This formulation of BM is simpler and more general than that of Eq. (7) in the sense that different layers are represented in a unified manner. Note that only one index can exist in a formulation to avoid ambiguity based on item 3 in Section 3. As 2 index sets $\mathscr{E}$ and Λ are used in this formulation, the condition $\mathscr{E} \cap \Lambda = \emptyset$ must be satisfied.

The conditional distribution of $\mathbf{x}_i = 1$ where $i \in \mathscr{V}_k$, which corresponds to Eq. (8) , can be obtained as following. By substituting Eq. (14) in Eq. (12) , and assuming $\mathscr{V}_0 = \mathscr{V}_4 = \emptyset$, Eq. (15) is obtained. Note that Z_p in Eq. (14) is reduced.

$$p\left(\mathbf{x}_i = 1 | \mathbf{x}_{\Lambda\setminus\{i\}}\right) = \frac{\exp\left(-\phi'\left(\mathbf{x}_i, \mathbf{x}_{\Lambda\setminus\{i\}}\right)\right)}{\sum_{\mathbf{y}_i} \exp\left(-\phi'\left(\mathbf{y}_i, \mathbf{x}_{\Lambda\setminus\{i\}}\right)\right)}, \quad (15)$$

where $\phi'\left(\mathbf{z}_i, \mathbf{z}_{\Lambda\setminus\{i\}}\right)$ is defined as Eq. (16) with $E(i) = \mathscr{V}_{k-1} \times \{i\} \cup \{i\} \times \mathscr{V}_{k+1}$. Although ϕ' is almost the same as ϕ, it is used because we have to deal with the element indexed by i separately from the others.

$$\begin{aligned} \phi'\left(\mathbf{z}_i, \mathbf{z}_{\Lambda\setminus\{i\}}\right) = &-\sum_{(h,j)\in E(i)} \mathbf{z}_h w_{c(h,j)} \mathbf{z}_j \\ &-\sum_{(h,j)\in c^{-1}(\mathscr{E})\setminus E(i)} \mathbf{z}_h w_{c(h,j)} \mathbf{z}_j. \quad (16) \end{aligned}$$

Considering the state space $\{0,1\}$ of $\mathbf{y}_i$ and that the second summation term of Eq. (16) exists commonly in the both denominator and numerator in the right hand of Eq. (15) , Eq. (15) can be written as Eq. (17) !ß&

$$p\left(\mathbf{x}_i = 1 | \mathbf{x}_{\Lambda\setminus\{i\}}\right) =$$
$$f_\sigma\left(\sum_{h\in\mathscr{V}_{k-1}} \mathbf{x}_h w_{c(h,i)} + \sum_{j\in\mathscr{V}_{k+1}} w_{c(i,j)}\mathbf{x}_j\right). \quad . \quad . \quad (17)$$

Eq. (17) takes the same form independent of the graphical structure of BM. On the other hand, Eq. (8) , which uses conventional notation, depends on the graphical structure.

This derivation is based on the notation of set theory and the proposed notation. As the notation of set theory is common, it is supposed that the proposed notation is acceptable to those who uses modern mathematics such as probability theory and probabilistic models.

5. Conclusion

This letter proposed the notation for a random variables and its probability space based on an index set. It was shown that proposed notation can represent several frequently used formulas in general and unambiguous forms. This letter also showed that the formulation of BM is written in a simple and general way by applying the proposed notation.

Acknowledgements

This work was partly supported by Grant-in-Aid for Research on Priority Areas, Tokyo Metropolitan University, "Research on social big data" and JSPS KAKENHI Grant Numbers JP15H02780 and JP16K12535.

References:

[1] R. Salakhutdinov and G. Hinton, "Deep Boltzmann machines," 12th Int. Conf. on Artificial Intelligence and Statistics, pp. 448-445, 2009.
[2] N. T. Kuong, E. Uchino, and N. Suetake, "IVUS Tissue characterization of coronary plaque by classification restricted Boltzmann machine," J. Adv. Comput. Intell. Intell. Inform., Vol.21, No.1, pp. 67-73, 2017.
[3] R. Salakhutdinov and A. Mnih, "Bayesian probabilistic matrix factorization using Markov chain Monte Carlo," 25th Int. Conf. on Machine Learning, pp. 880-887, 2008.
[4] A. A. Efros and T. K. Leung, "Texture synthesis by non-parametric sampling," 7th IEEE Int. Conf. on Computer Vision, Vol.2, pp. 1033-1038, 1999.
[5] A. Einstein, "The foundation of the general theory of relativity," Annalen der Physik, Vol.49, No.7, pp. 769-822, 1916.
[6] D. E. Smith, "History of mathematics," Dover Publications, Vol.2, p. 692, 1958.
[7] The International Organization for Standardization, "ISO 80000-2:2009(E)," 47 pp., 2009, https://people.engr.ncsu.edu/jwilson/files/mathsigns.pdf [accessed June 17, 2019]
[8] Mathematical Society of Japan, "Encyclopedic dictionary of mathematics," MIT Press, 2nd edition, 3rd printing, 1993.
[9] Y. V. Prohorov and Y. A. Rozanov, "Probability theory," Springer-Verlag, 1969.
[10] S. Geman and D. Geman, "Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images," IEEE Trans. on Pattern Analysis and Machine Intelligence, Vol.PAMI-6, No.6, pp. 724-741, 1984.
[11] C. M. Bishop, "Pattern recognition and machine learning," Springer, 8th printing, 2006.
[12] C. Lin and Y. He, "Joint sentiment/topic model for sentiment analysis," 18th ACM Conf. on Information and Knowledge Management, pp. 375-384, 2009.

Name:
Hiroki Shibata

Affiliation:
Graduate School of System Design, Tokyo Metropolitan University

Address:
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Brief Biographical History:
2014 B.S. dgree from Tokyo Metropolitan University
2016 M.S. dgree from Tokyo Metropolitan University

Membership in Academic Societies:
- The Institute of Electronics, Information and Communication Enginneers (IEICE)
- The Japanese Society of Artificial Intelligence (JSAI)
- Association for Computing Machinery (ACM)

Name:
Yasufumi Takama

Affiliation:
Graduate School of System Design, Tokyo Metropolitan University

Address:
6-6 Asahigaoka, Hino, Tokyo 191-0065, Japan

Brief Biographical History:
1994 B.S. degree from The University of Tokyo
1996 M.S. degree from The University of Tokyo
1999 Dr.Eng. dgree from The University of Tokyo
2014- Professor, Graduate School of System Design, Tokyo Metropolitan University

Main Works:
- Y. Takama et al., "Classification of Street Lighting Conditions for a Community-Centric System," J. Adv. Comput. Intell. Intell. Inform., Vol.20, No.6, pp. 875-881, 2016.

Membership in Academic Societies:
- IEEE Computer Society
- The Institute of Electronics, Information and Communication Enginners (IEICE)
- The Japanese Society of Artificial Intelligence (JSAI)
- Information Processing Society of Japan (IPSJ)