A Generalized Fitting Algorithm Using the Kolmogorov-Smirnov Test

Naoshi Sakamoto

***Abstract*—In this paper, I propose a fitting algorithm for a probability distribution for observations by using the Kolmogorov-Smirnov test. Drezner *et al.* propose an algorithm that calculates the mean and standard deviation of a normal distribution of observations in order to minimize the KS statistic of the Kolmogorov-Smirnov test. I generalize the algorithm of Drezner *et al.* and obtain the necessary conditions that enable other probability distributions to be applied. I show that it may be applied to well-known probability distributions.**

***Index Terms*—Closest fit, kolmogorov-smirnov test, probabilistic distribution.**

I. Introduction

Consider various series of observations. The data generating process of each series is not necessarily known. In order to better understand the data generating process, it is common to assume a probability distribution.

Let a sorted series of observations be $x_1, \ldots, x_n$. Assuming that these values are generated by an i.i.d. probability model that conforms to a continuous cumulative distribution function (CDF) $F(x)$, there are several methods for calculating the discrepancy between the model and the observed series. The Kolmogorov-Smirnov test is a method that calculates the difference between the CDF of the assumed model and the CDF of the observations. The resulting KS statistic is an indicator of the appropriateness of the assumed distribution.

There are many situations that demand knowledge of the probability distribution for a set of observations. One example is the situation in which it is important to determine the parameters of the probability distribution. Another is when selecting a model from a set of candidates, where models of the candidates have parameters. For these situations, it is necessary to determine the most appropriate parameters for the given observations. The KS statistic indicates how well the assumed model conforms to the observations.

In order to achieve this, methods to estimate parameters by using Kolmogorov-Smirnov test have been developed. When discussing the estimation of parameters of a probability model, it is necessary to treat the model as a family of CDFs. Methods for fitting for a general family of CDFs are complicated. Györfi *et al*. [2] shows the condition for the consistency of a family of probability distributions with respect to Kolmogorov-test.

On the other hand, Drezner *et al*. [1] focuses on the normal distribution and proposes an efficient fitting algorithm (DTZ algorithm) using the Kolmogorov-Smirnov test.

In this study, I analyze and generalize the DTZ algorithm, show necessary conditions to apply the generalized algorithm, and show some common probability distributions to which it can be applied.

The remainder of this paper is organized as follows: Section II discusses notation and the Kolmogorov-Smirnov test. Section III discusses the related literature. Section IV shows the generalized algorithm and the conditions. Section V concludes this paper.

II. Preliminaries

A. Notation

Let $\mathfrak{R}$ be the set of real numbers. Let $\exp(x)$ denote the exponential function with argument x. For $Dom_i \subseteq \mathfrak{R}$ (i=1, 2, ..., k), and multivariate function $F: Dom_1 \times Dom_2 \times Dom_3 \times \ldots \times Dom_k \to [0, 1]$ ($k \geq 2$), define the inverse function for the second argument $\mathrm{inv}_2 F: Dom_1 \times [0,1] \times Dom_3 \times \ldots \times Dom_k \to Dom_2$ as (1):

$$\mathrm{inv}_2 F(x_1; u, x_3, \ldots, x_k) = \inf\{y \mid F(x_1; y, x_3, \ldots, x_k) = u\} \quad (1)$$

Let $N(\mu, \sigma^2)$ denote the normal distribution with mean μ and standard deviation σ.

Let $\Phi(x)$ denote the CDF of $N(0, 1^2)$ (2), and let $\mathrm{erfc}(x)$ denote the complementary error function (3):

$$\Phi(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{t^2}{2}\right) dt \quad (2)$$

$$\mathrm{erfc}(x) = \frac{2}{\sqrt{\pi}} \int_{x}^{\infty} \exp(-t^2)\, dt \quad (3)$$

A real number function $f(x)$ is polynomial time computable if there exists a polynomial $p(n, m)$ where, for the fixed accuracy with length of m and any x with a length n, the value with the accuracy of $f(x)$ can be calculated in $p(n, m)$ steps.

Definition 1. *For $k \geq 2$, let $\Theta = \prod_{i=2}^{k} Dom_i$. A k-variate function $F(x_1; x_2, \ldots, x_k) = F(x_1; \theta)$ ($\theta \in \Theta$) is CDF-capable for x_1 if the following conditions are satisfied:*

1) For all $\theta \in \Theta$, $0 \leq F(x; \theta) \leq 1$;
2) For all $\theta \in \Theta$, $F(x; \theta)$ is non-decreasing in x;
3) For all $\theta \in \Theta$, $F(x; \theta)$ is right continuous in x;
4) For all $\theta \in \Theta$, $\lim_{x \to -\infty} F(x; \theta) = 0$;
5) For all $\theta \in \Theta$, $\lim_{x \to \infty} F(x; \theta) = 1$.

B. The Kolmogorov-Smirnov Test

The Kolmogorov-Smirnov test was proposed by Kolmogorov and later improved by Smirnov. Knuth [3] explains the importance of this test. For a continuous probability distribution, if the observations $x_1, \ldots, x_n$ obey a

Manuscript received October 5, 2015; revised January 8, 2016.

N. Sakamoto is with the Department of Information and Communication Engineering, Tokyo Denki University, Tokyo, Japan (e-mail: sakamoto@c.dendai.ac.jp).

DOI: 10.7763/IJCTE.2017.V9.1127

cumulative distribution function F, then $F(x_1), \ldots, F(x_n)$ obey the uniform distribution with range from 0 to 1. Thus, the KS statistic is the maximum value of the gap between the distribution of these and 1/*n*, 2/*n*, ..., *n*/*n*. This indicates how closely the empirical distribution conforms to the assumed probability distribution.

For the sorted observations $x_{1,\ \ldots,}\ x_n$ ($x_1 \leq \ldots \leq x_n$) and an assumed CDF $F(x)$, the KS statistic is defined by (4):

$$KS^+ = \max_{1 \leq k \leq n} \left\{ \frac{k}{n} - F(x_k) \right\},$$
$$KS^- = \max_{1 \leq k \leq n} \left\{ F(x_k) - \frac{k-1}{n} \right\}, \tag{4}$$
$$KS = \max\{KS^+, KS^-\}.$$

III. Related Literature

Győrfi *et al.* [2] introduces a notion of consistency for a parameterized family of probability distributions with Kolmogorov distance (5), yields conditions for consistency, and shows that some well-known probability distributions are consistent. The conditions that are applied in this study are the same as those in Győrfi *et al.* [2].

$$K(F, G) = \sup_x |F(x) - G(x)| \tag{5}$$

Győrfi *et al.* [2] shows that the normal, log-normal, gamma, beta, binomial, Poisson, geometric, and negative binomial distributions are consistent.

Weber *et al.* [4] evaluates the degree to which an evolutionary program is suitable for fitting continuous CDFs by using Kolmogorov-Smirnov test. Their study uses a software tool called the minimum Kolmogorov-Smirnov estimation fitter. This is based on the bell-curve based evolutionary optimization algorithm [5]. The fitting algorithm may be applied to many common probability distributions, including those discussed in this study. Noting that evolutionary programs are generally not universal, Weber *et al.* [5] expresses that some evolutionary programs may be applied to many probability distributions. This implies that some probability distributions might share a property that makes the search for suitable parameters efficient.

Drezner, Turel and Zerom [1] proposes a fast algorithm that seeks the μ and σ that minimize the KS statistic of the normal distribution $N(\mu, \sigma^2)$ with respect to sorted observations $x_{1,\ \ldots,}\ x_n$ ($x_1 \leq \ldots \leq x_n$). Their algorithm is as follows. Let Φ(x) denote the CDF of $N(0, 1^2)$. Then, denote the CDF of $N(\mu, \sigma^2)$ as $\Phi(\frac{x-\mu}{\sigma})$. Thus, for a mean μ and a standard deviation σ, the KS statistic is defined as:

$$KS(\mu, \sigma) = \max_{1 \leq k \leq n} \left\{ \frac{k}{n} - \Phi\left(\frac{x_k - \mu}{\sigma}\right), \Phi\left(\frac{x_k - \mu}{\sigma}\right) - \frac{k-1}{n} \right\}$$

Suppose that L bounds the KS statistic. Then we obtain (7) by solving (6) for μ:

$$F(\sigma, L) = \min_{k > nL} \left\{ x_k - \Phi^{-1}\left(\frac{k}{n} - L\right)\sigma \right\}$$
$$- \max_{k < n(1-L)+1} \left\{ x_k - \Phi^{-1}\left(L + \frac{k-1}{n}\right)\sigma \right\}. \tag{7}$$

The KS statistic is less than L iff F(σ, L) is not negative. Since, both the minimum and the maximum operators are applied to linear functions, solving for the maximum value of F(σ, L) becomes a linear programming problem. Thus, an algorithm that solves a linear programming problem, such as that of Megiddo [6], can determine the relative values of the KS statistic and L. Then, a binary search using this information can seek the minimum L. This algorithm is denoted the DTZ algorithm (see Fig. 1).

1. Let $L_{min} = \frac{1}{n}$ and L_{max}=1. Let ε denote an arbitrarily small positive value.
2. If $L_{max} - L_{min} < \varepsilon$, output L_{max} as the KS statistic, then terminate.
3. Let $L = (L_{min} + L_{max})/2$. By applying the algorithm of Megiddo [6] for solving a linear programming problem for a finite degree, compute the σ that maximizes F(σ,L).
4. If F(σ,L) is negative, let $L_{min} = L$. Otherwise, let $L_{max} = L$.
5. Go to 2.

Fig. 1. DTZ algorithm.

IV. Proposed Algorithm

The aim of this study is to apply the DTZ algorithm to probability distributions other than the normal distribution. Thus, the condition that allows for the algorithm to be applied to other probability distributions is imposed.

First, the DTZ algorithm is abstracted.

Definition 2. *A family of probability distributions $\mathcal{F}$ is DTZ-applicable if there exists a polynomial time computable decision function G*(L)*, where G*(L)*≥0, iff there exists $F \in \mathcal{F}$ such that the KS statistic of F(x) is less than L with respect to given sorted observations $x_{1,\ \ldots,}\ x_n$.*

If F is DTZ-applicable, then the minimum value of L can be determined with a binary search by applying the abstract DTZ algorithm (Fig. 2).

A. CDFs with a Single Parameter

This section discusses families of probability distributions with a single parameter, such as the exponential distribution.

Definition 3. *A CDF-capable bivariate function $F(x; y)$ ($F: Dom_1 \times Dom_2 \to [0, 1]$) has Property A if the following conditions hold:*

1) For all x, $F(x; y)$ is non-increasing in y.
2) *For* all x, $F(x; y)$ is right continuous in y.
3) $\mathrm{inv}_2 F(x; u)$ is polynomial time computable.

1. Let $L_{min} = \frac{1}{n}$ and L_{max}=1. Let ε denote an arbitrarily small positive number.
2. If $L_{max} - L_{min} < \varepsilon$, output L_{max} as the KS statistic, then terminate.
3. Let $L = (L_{min} + L_{max})/2$.
4. If $G(L)$ is negative, let $L_{min} = L$. Otherwise, let $L_{max} = L$.
5. Go to 2.

Fig. 2. Abstract DTZ algorithm.

Theorem 1: *Let Dom*$_1$ *and Dom*$_2$ *be subsets of* $\Re$*. If a*

CDF-capable bivariate function $F(x;y)$ $(F: Dom_1 \times Dom_2 \rightarrow [0,1])$ *has Property A, then* $\mathcal{F} = \{F(x;y)|y \in Dom_2\}$ *is DTZ-applicable.*

With respect to the function of the minimum of the inverse, I show the following:

Lemma 2. *If F(x; y) is non-increasing and right continuous in* y *for all* x, *then* $\mathrm{inv}_2F(x;u)$ *is decreasing in* u, *for all x.*

Proof: For any x, let $Y_x(u)=\{y \mid F(x;y)=u\}$.

For $u_0, u_1 \in [0,1]$, $\mu_0 < \mu_1$, $Y_x(\mu_0)$ and $Y_x(\mu_1)$ each contains its respective minimum because they are right continuous. Let $y_0 = \min Y_x(u_0)$ and $y_1 = \min Y_x(u_1)$. Since $Y_x(u_0) \cap Y_x(u_1) = \emptyset$, this yields $y_0 \neq y_1$. Therefore $y_0 > y_1$.

Proof of Theorem 1: For a CDF family $\mathcal{F} = \{F(x;y)|y \in Dom_2\}$ with Property A, if a polynomial time computable decision function $G(L)$ for F can be derived, the KS statistic can be computed by applying the abstract DTZ algorithm.

Consider the sorted observations $x_1, \ldots, x_n$ $(x_1 \leq \ldots \leq x_n)$. If, for some y and some L, the KS statistic of $F(x; y)$ is less than or equal to L, then (8) and (9) hold for all k $(1 \leq k \leq n)$:

$$\frac{k}{n} - F(x_k;y) \leq L \tag{8}$$

$$F(x_k;y) - \frac{k-1}{n} \leq L \tag{9}$$

Since the range of F is [0, 1], (8) holds if $k/n \leq L$, and (9) trivially holds if $1-(k-1)/n \leq L$. Thus, (10) and (11) hold:

$$F(x_k;y) \geq \frac{k}{n} - L \quad \text{for } k > nL \tag{10}$$

$$F(x_k;y) \leq L + \frac{k-1}{n} \quad \text{for } k < n(1-L)+1 \tag{11}$$

According to Property A, because F is non-increasing in y, the direction of the sign of each inequality reverses when solving (10) and (11) for y:

$$\mathrm{inv}_2F\left(x_k;\frac{k}{n} - L\right) \leq y \quad \text{for } k > nL$$
$$\mathrm{inv}_2F\left(x_k;L + \frac{k-1}{n}\right) \geq y \quad \text{for } k < n(1-L)+1$$

Applying the maximum and the minimum operators yields (12) and (13):

$$\max_{k>nL}\left\{\mathrm{inv}_2F\left(x_k;\frac{k}{n} - L\right)\right\} \leq y \tag{12}$$

$$\min_{k<n(1-L)+1}\left\{\mathrm{inv}_2F\left(x_k;L + \frac{k-1}{n}\right)\right\} \geq y \tag{13}$$

By deleting y and subtracting (12) from (13), the $G(L)$ (14) must always be greater than or equal to zero.

$$G(L) = \min_{k<n(1-L)+1}\left\{\mathrm{inv}_2F\left(x_k;L + \frac{k-1}{n}\right)\right\} - \max_{k>nL}\left\{\mathrm{inv}_2F\left(x_k;\frac{k}{n} - L\right)\right\} \tag{14}$$

Thus, $G(L) \geq 0$ holds iff L bounds KS statistic. Recall that $\mathrm{inv}_2F(x;u)$ is polynomial time computable. Because $G(L)$ can be computed by evaluating the maximum and the minimum for each set of n observations, $G(L)$ is also polynomial time computable.

Therefore, F is DTZ-applicable.

B. CDF with Two Parameters

Definition 4. *A CDF-capable trivariate function* $F(x;y,z)$ $(F: Dom_1 \times Dom_2 \times Dom_3 \rightarrow [0,1])$ *has Property B if the followings are satisfied:*

1) For all x and z, $F(x; y,z)$ is monotone non-increasing in y.
2) For all x and z, $F(x; y, z)$ is right continuous for y.
3) For all x and u, $\mathrm{inv}_2F(x;\ u,z)$ is a linear function for z. That is, it can be described as $c_0(x; u)+ c_1(x; u)z$, where c_0 and c_1 are polynomial time computable.

Theorem 3. *If a CDF-capable trivariate function* $F(x;y,z)$ $(F: Dom_1 \times Dom_2 \times Dom_3 \rightarrow [0,1])$ *has Property B, then the family of probability distributions* $\mathcal{F} = \{F(x;y,z)|y \in Dom_2, z \in Dom_3\}$ *is DTZ-applicable.*

Lemma 4. *For* $f_k : Dom_3 \rightarrow Dom_2$ $(1 \leq k \leq n)$, *if all* f_k *are convex, then* $\min_{1\leq k\leq n}\{f_k(z)\}$ *is also convex.*

Proof: For some x_0, x_1, and t, assuming that $g(tx_0 +$ *1−tx1<tgx0+1−tg(x1)* yields a contradiction.

For i such that $f_i(tx_0 + (1-t)x_1) = g(tx_0 + (1-$ *tx*1), imposing the assumption implies $f_i(tx_0 + (1-t)x_1) < tf_i(x_0) + (1-t)f_i(x_1)$. Thus, $g(x_0) \leq f_i(x_0)$ and $g(x_1) \leq f_i(x_1)$ yield $g(tx_0 +$ 1−*tx*1≥*tgx*0+1−*tgx*1. This contradicts the assumption.

Lemma 5. *If all* f_k *are concave, then* $\max_{1\leq k\leq n}\{f_i(z)\}$ *is also concave.*

Proof: Similar to the proof of Lemma 4.

Proof of Theorem 3: The proof is similar to the proof of Theorem 1 except that the number of parameters increases by one. That is, suppose that KS statistic is less than L, there exists a decision function, such as (15), such that $\mathrm{inv}_2F(x;u,z)$ is the minimum of the inverse function of $F(x; y, z)$ for y:

$$G(z,L) = \min_{k<n(1-L)+1}\left\{\mathrm{inv}_2F\left(x_k;L + \frac{k-1}{n}, z\right)\right\} - \max_{k>nL}\left\{\mathrm{inv}_2F\left(x_k;\frac{k}{n} - L, z\right)\right\} \tag{15}$$

Applying Property B yields (16):

$$\mathrm{inv}_2F\left(x_k;\frac{k}{n} - L, z\right) = c_0\left(x_k;\frac{k}{n} - L\right) + c_1\left(x_k;\frac{k}{n} - L\right)z \tag{16}$$

Because $\mathrm{inv}_2F(x;u,z)$ is a linear function in z, it is both convex and concave. Thus, by applying Lemmas 4 and 5, $G(z, L)$ is convex in z. By applying the algorithm of Megiddo [6] that solves linear programming problems with constant degrees, the z that maximizes $G(z, L)$ can be computed. Moreover, it can be determined if $G(z, L) \geq 0$ holds in polynomial time.

Therefore, $\mathcal{F}$ is DTZ-applicable.

C. CDF with More than Two Parameters

For CDFs with more than two parameters, if $\mathrm{inv}_2F(x_1;u,x_3,\ldots,x_k)$ can be denoted as a linear function $c_0(x_1;u) + \sum_{i=3}^{k} c_i(x_1;u)\cdot x_i$ for each parameter x_i (i=3,..., k), then the algorithm may be applied. However, this algorithm has not yet been applied to the multinomial distribution, the hyper geometric distribution, and others.

D. Application to Typical Probability Distributions

Tables I and II summarize DTZ-applicability for common probability distributions listed by Wikipedia [7].

TABLE I: ONE-PARAMETER PROBABILITY DISTRIBUTION

Distribution	$CDF(x;y)$	$\mathrm{inv}_2 CDF(x;u)$
Geometric	$1-\exp((x+1)y)$	$\frac{\log(1-u)}{x+1}$
Exponential	$1-\exp(-yz)$	$-\frac{\log 1-u}{x}$

The followings are examples:

1) *Geometric*

Equation (17) shows the geometric distribution. Denoting p as $1-\exp(y)$ and k as x, the algorithm presented in this study may be applied.

$$\Pr[X \le k] = F(k) = 1-(1-\mathrm{p})^{k+1}. \tag{17}$$

2) *Pareto*

Equation (18) shows the Pareto distribution. Denoting a as $1/z$ and b as $\exp(y)$, the algorithm presented in this study may be applied.

$$F(x) = 1-\left(\frac{b}{x}\right)^{a}. \tag{18}$$

3) *Weibull*

Equation (19) shows the Weibull distribution. Denoting $1/z$ as k and $\exp(y)$ as λ, the algorithm presented in this study may be applied.

$$F(x) = \begin{cases} 1-\exp\left(-\left(\frac{x}{\lambda}\right)^{k}\right) & \text{if } x \ge 0, \\ 0 & \text{o.w.} \end{cases} \tag{19}$$

4) *Uniform*

The uniform distribution has three conditions for x. However, since observations must be contained within the range, the second condition of the CDF is always chosen. Thus, the uniform distribution is DTZ-applicable.

The following probability distributions are not seemed to be DTZ-applicable: Poisson, Erlang, gamma, binomial, Student's t, beta, and x^2.

TABLE II: TWO-PARAMETER PROBABILISTIC DISTRIBUTION

Distribution	$CDF(x;y,z)$	$c_0(x;u)$	$c_1(x;u)$
Cauchy	$\frac{1}{\pi}\arctan\left(\frac{x-y}{z}\right)+\frac{1}{2}$	x	$-\tan\left(\pi u-\frac{\pi}{2}\right)$
Gumbel	$\exp\left(-\exp\left(-\frac{x-y}{z}\right)\right)$	x	$\log\log\frac{1}{u}$
Normal	$\Phi\left(\frac{x-y}{z}\right)$	x	$-\Phi^{-1}(u)$
Laplace	$\begin{cases} \frac{1}{2}\exp\left(\frac{x-y}{z}\right) & \text{if } x<y \\ 1-\frac{1}{2}\exp\left(-\frac{x-y}{z}\right) & \text{if } x \ge y \end{cases}$	x	$\begin{cases} -\log 2u & \text{if } u<\frac{1}{2} \\ \log(2-2u) & \text{if } u \ge \frac{1}{2} \end{cases}$
Log Normal	$\Phi\left(\frac{\log x-y}{z}\right)$	$\log x$	$-\Phi^{-1}(u)$
Lévy	$\mathrm{erfc}\left(\sqrt{\frac{z}{2(x-y)}}\right)$	x	$\frac{1}{2\mathrm{erfc}^{-1}(u)^2}$
Logistic	$\frac{1}{1+\exp\left(-\frac{x-y}{z}\right)}$	x	$\log\left(\frac{1}{u}-1\right)$
Palate	$1-\left(\frac{\exp(y)}{x}\right)^{\frac{1}{z}}$	$\log x$	$\log(1-u)$
Uniform	$\begin{cases} 0 & \text{if } x<z \\ \frac{x-z}{y-z} & \text{if } z \le x \le y \\ 1 & if\ x>y \end{cases}$	$\frac{x}{u}$	$\frac{1-u}{u}$
Weibull	$\begin{cases} 1-\exp\left(-\left(\frac{x}{\exp(y)}\right)^{\frac{1}{z}}\right) & \text{if } x \ge 0 \\ 0 & \text{if } x<0 \end{cases}$	$\log x$	$-\log\log\frac{1}{1-u}$

V. CONCLUSION

This study analyzes the fitting algorithm proposed by Drezner *et al.*, abstracts it, derives the necessary conditions to apply it to probability distributions other than the normal, and determines if some common probability distributions satisfy the condition.

While the derived conditions are strict, they are applicable to many probability distributions. However, the conditions cannot be applied to the Poison, Gamma, and other distributions listed above. In the future, I would like to design a fast fitting algorithm for these probability distributions.

REFERENCES

[1] Z. Drezner, O. Turel, and D. Zerom, "A modified Kolmogorov-Smirnov test for normality," *Communications in Statistics-Simulation and Computation*, vol. 39, no. 4, pp. 693–704, 2010.

[2] L. Györfi, I. Vajda, and E. van der Meulen, "Minimum Kolmogorov distance estimates of parameters and parametrized distributions," *Metrika*, vol. 43, pp. 237–255, 1996.

[3] D. E. Knuth, *The Art of Computer Programming: Seminumerical Algorithms*, 3rd ed. USA: Addison-Wesley Longman Publishing Co., Inc., 1997, vol. 2.

[4] M. D. Weber, L. M. Leemis, and R. K. Kincaid, "Minimum Kolmogorov-Smirnov test statistic parameter estimates," *Journal of Statistical Computation and Simulation*, vol. 76, pp. 196–206, 2006.

[5] J. Sobieszczanski-Sobieski, K. Laba, and R. Kincaid, "Bell-curve based evolutionary optimization algorithm," in *7th AIAA/USAF/NASA/ISSMO Symposium on Multidisciplinary Analysis and Optimization*, September 1998, pp. 2083–2096.

[6] N. Megiddo, "Linear-time algorithms for linear programming in R^3 and related problems," *SIAM J. Comput.*, vol. 12, no. 4, pp. 759–776, 1983. [Online]. Available: http://dx.doi.org/10.1137/0212052

[7] Wikipedia, Probability distribution — Wikipedia, the free encyclopedia. [Online]. Available: http://en.wikipedia.org/wiki/Probability_distribution

[8] N. Megiddo, "Linear programming in linear time when the dimension is fixed," *J. ACM*, vol. 31, no. 1, pp. 114–127, Jan. 1984.

Naoshi Sakamoto was born in Japan in 1964, he is currently a professor at Tokyo Denki University. He graduated from the University of Electro-Communication in 1987 and received his M.S. and Dr.S. degrees from the Tokyo Institute of Technology in 1989 and 2001, respectively.

From 1992 to 1997, he was a research assistant at the Computer Center of Hitotsubashi University. From 1997 to 2001, he was a research assistant at the Tokyo Institute of Technology. From 2001 to 2014, he was an associate professor at Tokyo Denki University and became a full professor in 2014.

He received the 1999 IEICE Best Paper Award and his paper was selected as one of the best papers in the 2015 SNPD. His research interests are distributed algorithms, randomized algorithms, and complexity theory.