

# UC San Diego

## UC San Diego Previously Published Works

### Title

Error estimate and convergence analysis of moment-preserving discrete approximations of continuous distributions

### Permalink

<https://escholarship.org/uc/item/6jr3t66z>

### Journal

AIP Conference Proceedings, 1636(1)

### ISSN

0094-243X

### Authors

Tanaka, Ken'ichiro  
Toda, Alexis Akira

### Publication Date

2014-12-05

### DOI

10.1063/1.4903706

Peer reviewed

# Error Estimate and Convergence Analysis of Moment-Preserving Discrete Approximations of Continuous Distributions

Ken'ichiro Tanaka\* and Alexis Akira Toda†

\*School of Systems Information Science, Future University Hakodate, Hokkaido, Japan

†Department of Economics, University of California San Diego, 9500 Gilman Dr, La Jolla, CA 92093

**Abstract.** We propose a numerical method to approximate a given continuous distribution by a discrete distribution with prescribed moments. The approximation is achieved by minimizing the Kullback-Leibler information of the unknown discrete distribution relative to the known continuous distribution (evaluated at given discrete points) subject to some moment constraints. We study the theoretical error bound and the convergence property of the method. The order of the theoretical error bound of the expectation of any bounded measurable function with respect to the approximating discrete distribution is never worse than the integration formula we start with, and therefore the approximating discrete distribution weakly converges to the given continuous distribution. Moreover, we present some numerical examples that show the advantage of our method.

**Keywords:** probability distribution, discrete approximation, generalized moment, integration formula, Kullback-Leibler information, Fenchel duality, error estimate, convergence analysis

**PACS:** 01.30.Cc, 02.50.Cw, 02.60.Gf, 02.60.-x

## INTRODUCTION

In various fields, it is often necessary to numerically compute an expectation  $E[g(X)]$  for a function  $g: \mathbb{R}^K \rightarrow \mathbb{R}$  and a random variable  $X$  with a given continuous distribution. Let  $f: \mathbb{R}^K \rightarrow \mathbb{R}$  be the probability density function of the given distribution. If the function  $g$  is explicitly known, it suffices to use some integration formula

$$E[g(X)] = \int_{\mathbb{R}^K} g(x) f(x) dx \approx \sum_{i=1}^M w_{i,M} g(x_{i,M}) f(x_{i,M}) \quad (1)$$

for the computation, where  $M$  is the number of integration points and  $w_{i,M}$  is the weight on point  $x_{i,M}$ . However, especially in economics, the function  $g$  often contains some unknown parameters  $\theta$  as  $g = g(\cdot, \theta)$  and we want to determine the parameter value  $\theta^*$  from some given conditions for  $E[g(X, \theta)]$ , say  $\theta^* = \arg \max_{\theta} E[g(X, \theta)]$ . Furthermore, the values  $g(x, \theta)$  or  $f(x)$  may be available only on some prescribed discrete set of  $x$ 's, say  $D \subset \mathbb{R}^K$ . Such a restriction may arise when  $g(\cdot, \theta)$  or  $f$  is obtained by some statistical estimate from real data, or when the application requires  $D$  to be a particular set, say lattice points. As a concrete example for the situation above, consider an optimal portfolio problem in economics. The problem is to determine the portfolio  $\theta = (\theta_1, \dots, \theta_J)$  that maximizes the expectation  $\frac{1}{1-\gamma} E[R(\theta)^{1-\gamma}]$  subject to the constraint  $\sum_{j=1}^J \theta_j = 1$ , where  $\gamma > 0$  is the relative risk aversion coefficient,  $R(\theta) = \sum_{j=1}^J R_j \theta_j$  is the return on portfolio, and  $R_j$ 's are random variables expressing the gross returns of financial assets. In such a situation, we need to compute  $E[g(X, \theta)]$  many times for different  $\theta$ 's using only the  $x$ 's in  $D$ . Then, it is desirable to use a highly accurate integration formula with light computing load, say, small  $M$  in (1). Some popular formulas, such as the Newton-Cotes type or the Gauss type formulas, are suitable for such a purpose. These formulas, however, are not necessarily available if the integration points  $x_{i,M}$ 's are restricted as  $\{x_{i,M}\} \subset D$ . In that case we need some recipes for the approximation in (1). Note that such an approximation is equivalent to finding a discrete distribution  $\{w_{i,M} f(x_{i,M})\}$  approximating the given distribution in the sense of the weak topology.

Several methods for discrete approximations of continuous distributions have been proposed in the literature [1, 2, 3, 4]. Tauchen [1] and Adda and Cooper [2] adopt simple partitions of the domain of the distribution function of a given distribution and assign the true probability to a representative point of each partitioned domain. Although their methods are intuitive, simple, and work in any dimension, their methods are not so accurate, in particular, they generate discrete distributions with only approximate moments. Miller and Rice [3] and Devuyt and Preckel [4] discretize the density function of a given distribution using the weights and the points of the Gaussian integration and

its generalization to multi-dimensions, respectively. Although their methods are often more accurate and can match prescribed polynomial moments exactly, they do not allow for the restriction  $\{x_{i,M}\} \subset D$  and cannot be applied to non-polynomial moments. Furthermore, the multi-dimensional method by Devuyt and Preckel [4] is computationally intensive and does not have a theoretical guarantee for the existence of the discretization, error bounds, or convergence.

As a remedy for these methods, in Tanaka and Toda [5], we proposed an approximation method based on Jaynes [6]’s maximum entropy principle (MaxEnt) that matches prescribed moments exactly. Starting from any integration formula, we “fine-tuned” the given probabilities by minimizing the Kullback-Leibler information (relative entropy) of the unknown probabilities relative to the given probabilities subject to the prescribed moment constraints. Furthermore, we proved the existence and the uniqueness of the solution of the minimization problem and that the solution can be easily computed by solving the dual problem. Our method is computationally very simple and works on any discrete set  $D$  of any dimension with any prescribed moments (not necessarily polynomials).

This paper has three contributions. First, we evaluate the theoretical error of our method. We show that the order of the theoretical error estimate is at most that of the initial integration formula. Thus our method does not compromise the order of the error at the expense of matching moments. Second, as a theoretical consequence of the estimate, we show the weak convergence of the discrete distribution generated by our method to the given continuous distribution. This means that, for any bounded measurable function  $g$ , the expectation of  $g$  with respect to the approximating discrete distribution converges to the exact one with respect to the given distribution. This convergence property is also practically important because it guarantees that our method never generates a pathological discrete distribution with exact moments which has extremely different probability from the given distribution on some domain, at least when the discrete set is large enough. Third, we present some numerical examples that show the advantage of our method.

## THE APPROXIMATION METHOD

In this section, we review the approximation method proposed in Tanaka and Toda [5]. Let  $f$  be a probability density function on  $\mathbb{R}^K$  and assume that some generalized moments

$$\bar{T} = \int_{\mathbb{R}^K} f(x) T(x) dx \tag{2}$$

are given, where  $T : \mathbb{R}^K \rightarrow \mathbb{R}^L$  is a continuous function. Moreover, for each positive integer  $M$ , assume that a finite discrete set  $D_M = \{x_{i,M} \mid i = 1, \dots, M\} \subset \mathbb{R}^K$  is given. As an example of  $D_M$ , for a real number  $h > 0$  and a positive integer  $N$ , we can consider the lattice  $D_M = \{(n_1 h, n_2 h, \dots, n_K h) \mid n_1, n_2, \dots, n_K = 0, \pm 1, \dots, \pm N\}$ , in which case  $M = (2N + 1)^K$ . Our aim is to determine a discrete probability distribution  $P_M = \{p(x_{i,M}) \mid x_{i,M} \in D_M\}$  on  $D_M$  with exact moments  $\bar{T}$  which approximates  $f$ .

To match the moments  $\bar{T}$  with  $P_M = \{p(x_{i,M}) \mid x_{i,M} \in D_M\}$ , it suffices to assign  $p(x_{i,M})$ ’s such that

$$\sum_{i=1}^M p(x_{i,M}) T(x_{i,M}) = \bar{T}. \tag{3}$$

Note that this equation is often ill-posed because in general the number of unknowns  $p(x_{i,M})$ ’s, namely  $M$ , is much larger than the number of equations (moments),  $L + 1$ .<sup>1</sup> To obtain  $P_M$  with (3) approximating  $f$ , we first choose a numerical integration formula by setting positive weights  $w_{i,M}$  ( $i = 1, 2, \dots, M$ ):

$$\int_{\mathbb{R}^K} f(x) g(x) dx \approx \sum_{i=1}^M w_{i,M} f(x_{i,M}) g(x_{i,M}), \tag{4}$$

where  $g$  is an arbitrary function whose expectation with respect to the density  $f$  we want to compute. For instance, if  $K = 1$  and  $D_M = \{nh \mid k = 0, \pm 1, \dots, \pm N\}$  for  $h > 0$ , we can choose the  $(2N + 1)$ -point trapezoidal formula for a univariate function on  $\mathbb{R}$  by setting  $w_{i,M} = h$ . Then, we obtain  $P_M = \{p(x_{i,M}) \mid x_{i,M} \in D_M\}$  by the following

---

<sup>1</sup> The “+1” comes from accounting the probabilities  $\sum_{i=1}^M p(x_{i,M}) = 1$ .

optimization problem:

$$\min_{\{p(x_{i,M})\}} \sum_{i=1}^M p(x_{i,M}) \log \frac{p(x_{i,M})}{w(x_{i,M})f(x_{i,M})} \text{ subject to } \sum_{i=1}^M p(x_{i,M})T(x_{i,M}) = \bar{T}, \sum_{i=1}^M p(x_{i,M}) = 1, p(x_{i,M}) \geq 0. \quad (\text{P})$$

The problem (P) is equivalent to the minimization problem of the Kullback-Leibler information of  $P_M$  relative to the discrete distribution proportional to  $\{w(x_{i,M})f(x_{i,M}) \mid x_{i,M} \in D_M\}$ . Note that the problem (P) has a unique solution if  $\bar{T} \in \text{conv} T(D_M)$ , where  $\text{conv} T(D_M)$  is the convex hull of  $T(D_M)$ , because in that case the constraint set is nonempty, compact, convex, and the objective function is continuous (by adopting the convention  $0 \log 0 = 0$ ) and strictly convex.

To characterize the solution of (P), we consider the Fenchel dual<sup>2</sup> of (P), which can be written as

$$\bar{\lambda}_M = \arg \min_{\lambda \in \mathbb{R}^L} \left[ -\langle \lambda, \bar{T} \rangle + \log \left( \sum_{i=1}^M w_{i,M} f(x_{i,M}) e^{\langle \lambda, T(x_{i,M}) \rangle} \right) \right], \quad (\text{D})$$

where  $\langle \cdot, \cdot \rangle$  denotes the inner product in  $\mathbb{R}^L$ . Note the simplicity of the dual problem (D) compared to the primal problem (P); the dual (D) is an *unconstrained* optimization problem with typically a small number of unknowns  $L$  whereas the primal problem (P) is a *constrained* optimization problem with typically a large number of unknowns  $M$ .

The following theorem in [5] shows that the solution of (P) are fine-tuned values of  $w(x_{i,M})f(x_{i,M})$ .

**Theorem 1** (Tanaka and Toda [5]). *Suppose that  $\bar{T} \in \text{conv} T(D_M)$ . Then the solution of (P) is given by  $p(x_{i,M}) = C w_{i,M} f(x_{i,M}) e^{\langle \bar{\lambda}_M, T(x_{i,M}) \rangle}$ , where  $C > 0$  is a normalizing constant and  $\bar{\lambda}_M$  is determined by (D).*

Theorem 1 indicates that the solution of (P) can be explicitly obtained if a solution  $\bar{\lambda}_M$  of (D) exists. Theorem 2 below guarantees the existence of a solution  $\bar{\lambda}_M$  of (D). Here, in order to guarantee the uniqueness of the solution as well, we adopt a stronger assumption  $\bar{T} \in \text{int}(\text{conv} T(D_M))$ , where “int” denotes the set of the interior points of a region.

**Theorem 2** (Tanaka and Toda [5]). *Suppose that  $\bar{T} \in \text{int}(\text{conv} T(D_M))$ . Then (i) the objective function in (D) is continuous and strictly convex, and (ii) the solution  $\bar{\lambda}_M$  uniquely exists.*

## ERROR BOUND AND CONVERGENCE PROPERTY

In this section we give theoretical estimate for the accuracy of our method. Let  $g : \mathbb{R}^K \rightarrow \mathbb{R}$  be a bounded measurable function. Under appropriate assumptions, we first estimate the error

$$E_{g,M} = \left| \int_{\mathbb{R}^K} f(x) g(x) dx - \sum_{i=1}^M p(x_{i,M}) g(x_{i,M}) \right|, \quad (5)$$

where  $p(x_{i,M})$ 's are determined by Theorem 1. Next, we show the weak convergence of  $P_M$  to  $f$ , i.e.,  $E_{g,M} \rightarrow 0$  ( $M \rightarrow \infty$ ) for any  $g$ . Throughout this paper,  $\langle \cdot, \cdot \rangle$  and  $\|\cdot\|$  denote the inner product and the Euclidean norm of  $\mathbb{R}^L$ , respectively.

Since  $f(x)$  is a probability density function, the moment condition (2) is equivalent to  $\int_{\mathbb{R}^K} f(x)(T(x) - \bar{T}) dx = 0$ . Hence by redefining  $T(x) - \bar{T}$  as  $T(x)$ , without loss of generality we may assume  $\bar{T} = 0$ . We keep this convention throughout the remainder of this section.

We consider the error estimate and the convergence analysis under the following two assumptions. The first assumption states that the moment defining function  $T$  has no degenerate components and the moment  $\bar{T} = 0$  can be expressed also as an expectation on the discrete set  $D_M$ .

**Assumption 1.** *The components of the moment defining function  $T$  are affine independent as functions both on  $\mathbb{R}^L \cap \text{supp} f$  and  $D_M \cap \text{supp} f$  for any positive integer  $M$ . Namely, for any  $0 \neq (\lambda, \mu) \in \mathbb{R}^L \times \mathbb{R}$ , there exists  $x_{i,M} \in D_M$  such that  $\langle \lambda, T(x_{i,M}) \rangle + \mu \neq 0$ . Furthermore,  $T$  and  $D_M$  satisfy  $0 \in \text{int}(\text{conv} T(D_M))$  for any positive integer  $M$ .*

<sup>2</sup> See [7] for an application of the Fenchel duality to entropy-like minimization problems.

The second assumption concerns the convergence property of the integration formula (4).

**Assumption 2.** For any bounded measurable function  $g$  on  $\mathbb{R}^K$ , we have

$$\lim_{M \rightarrow \infty} \sum_{i=1}^M w_{i,M} f(x_{i,M}) g(x_{i,M}) = \int_{\mathbb{R}^K} f(x) g(x) dx. \quad (6)$$

Furthermore, the integration formula applies to  $\|T(x)\|$  as well:

$$\lim_{M \rightarrow \infty} \sum_{i=1}^M w_{i,M} f(x_{i,M}) \|T(x_{i,M})\| = \int_{\mathbb{R}^K} f(x) \|T(x)\| dx =: I_{\|T\|} < \infty. \quad (7)$$

Since (6) merely states that the integration formula converges to the true value, (7) is the only essential assumption. The following theorem gives an error bound of the discrete approximation.

**Theorem 3.** Let Assumptions 1 and 2 be satisfied,  $g$  be a measurable function with  $|g(x)| \leq G$  ( $x \in \mathbb{R}^K$ ), and  $\alpha > 0$  be large enough such that

$$C_\alpha := \inf_{\lambda \in \mathbb{R}^L, \|\lambda\|=1} \frac{1}{2} \int_{\mathbb{R}^K} f(x) (\max\{0, \min\{\langle \lambda, T(x) \rangle, \alpha\}\})^2 dx > 0.$$

Then, for any  $\varepsilon$  with  $0 < \varepsilon < C_\alpha$ , there exists a positive integer  $M_\varepsilon$  such that for any  $M$  with  $M \geq M_\varepsilon$ , we have

$$E_{g,M} \leq E_{g,M}^{(a)} + G \left( E_{\mathbf{1},M}^{(a)} + 6 \frac{I_{\|T\|} + E_{\|T\|,M}^{(a)}}{C_\alpha - \varepsilon} E_{T,M}^{(a)} \right), \quad (8)$$

where  $E_{g,M}^{(a)}$ ,  $E_{\mathbf{1},M}^{(a)}$ ,  $E_{T,M}^{(a)}$  and  $E_{\|T\|,M}^{(a)}$  are the errors of the integration formula for the given functions  $g$ ,  $\mathbf{1}$ ,  $T$ , and  $\|T\|$ .

*Proof.* See the full paper [8]. □

Note that  $E_{g,M}$  is bounded by a formula consisting of  $E_{g,M}^{(a)}$ ,  $E_{\mathbf{1},M}^{(a)}$ ,  $E_{T,M}^{(a)}$  and  $E_{\|T\|,M}^{(a)}$ . Since all of them converge to zero as  $M \rightarrow \infty$ , it follows from (8) that  $E_{g,M} = O\left(\max\{E_{g,M}^{(a)}, E_{\mathbf{1},M}^{(a)}, E_{T,M}^{(a)}\}\right)$  as  $M \rightarrow \infty$ , so the error  $E_{g,M}$  is at most of the same order as the error of the initial integration formula. Thus our method does not compromise the order of the error at the expense of matching moments.

Using Theorem 3, we immediately obtain our main result that  $E_{g,M} \rightarrow 0$  ( $M \rightarrow \infty$ ), i.e., the weak convergence of the approximating discrete distribution  $P_M = \{p(x_{i,M})\}$  to  $f$ .

**Theorem 4.** Let Assumptions 1 and 2 be satisfied. Then, for any bounded measurable function  $g$ , we have

$$\lim_{M \rightarrow \infty} \sum_{i=1}^M p(x_{i,M}) g(x_{i,M}) = \int_{\mathbb{R}^K} f(x) g(x) dx, \quad (9)$$

i.e., the discrete distribution  $P_M$  weakly converges to the exact continuous distribution  $f$ .

## NUMERICAL EXPERIMENTS

In this section, we present some numerical examples that compare the accuracy of approximate expectations computed by an initial integration formula and its modifications by our method. All computations in this section are done by MATLAB programs with double precision floating point arithmetic on a PC.

## Gaussian and beta distributions

We choose the Gaussian and the beta distributions as examples of continuous distributions and adopt the trapezoidal formula as an initial integration formula to compute the expectation  $E[g(X)]$ , where  $g(x) = e^x \mathbf{1}_{[-10,10]}(x)$  for  $X \sim N(0, 1)$  and  $g(x) = e^x$  for  $X \sim Be(2, 4)$ . In the following, let  $M$  be an odd integer with  $M = 2N + 1$  ( $N = 1, 2, \dots, 12$ ).

For the Gaussian distribution, we set  $h_M^{(1)} = 1/\sqrt{N}$ ,

$$f_1(x) = \frac{1}{\sqrt{2\pi}} \exp\left(-\frac{x^2}{2}\right) \quad (x \in (-\infty, \infty)),$$

$$w_{i,M}^{(1)} = \begin{cases} h_M^{(1)} & (i \neq 1, M), \\ h_M^{(1)}/2 & (i = 1, M), \end{cases} \quad x_{i,M}^{(1)} = (i - N - 1)h_M^{(1)} \quad (i = 1, \dots, M = 2N + 1),$$

which means that we let  $D = \{nh_M^{(1)} \mid n = 0, \pm 1, \dots, \pm N\}$  and approximate the integral  $\int_{-\infty}^{\infty}$  by  $\int_{-\sqrt{N}}^{\sqrt{N}}$ . Then  $E[g(X)] = (\operatorname{erf}(9/\sqrt{2}) + \operatorname{erf}(11/\sqrt{2}))\sqrt{e}/2$  for  $X \sim N(0, 1)$ .

For the beta distribution, we set

$$f_2(x) = x(1-x)^3/B(2, 4) \quad (x \in [0, 1]),$$

$$w_{i,M}^{(2)} = \begin{cases} h_M^{(2)} & (i \neq 1, M), \\ h_M^{(2)}/2 & (i = 1, M), \end{cases} \quad x_{i,M}^{(2)} = (i-1)h_M^{(2)} \quad (i = 1, \dots, M),$$

where  $B(\cdot, \cdot)$  is the beta function and  $h_M^{(2)} = 1/(M-1)$ . Then  $E[g(X)] = 20(49 - 18e)$  for  $X \sim Be(2, 4)$ .

For numerical experiments, we compute  $E[g(X)]$  using five formulas; the trapezoidal formula

$$E[g(X)] \approx \sum_{k=1}^M w_{i,M}^{(k)} f_k(x_{i,M}^{(k)}) g(x_{i,M}^{(k)}) \quad (k = 1, 2), \quad (10)$$

its modifications by our method with exact polynomial moments  $E[X^l]$  up to 2nd order ( $l = 1, 2$ ), 4th order ( $l = 1, \dots, 4$ ), and 6th order ( $l = 1, \dots, 6$ ), and Simpson's formula with the number of grid points  $M = 2N + 1$  ( $N = 1, 2, \dots, 12$ ). (See [9] for more details on quadrature formulas.) Here, we intend to observe the relative errors of the computed values for small  $M$ 's. We do not compare to other more sophisticated methods such as the Gaussian quadrature since for such cases we cannot freely choose the integration points.

In order for (P) to have a solution, it is necessary that there are at least as many unknown variables ( $p(x_{i,M})$ 's, so in total  $M$ ) as the number of constraints ( $L$  moment constraints and  $+1$  for probabilities to add up to 1, so  $L + 1$ ). Thus we need  $M \geq L + 1$ .<sup>3</sup> A sufficient condition for the existence of a solution is  $\bar{T} \in \operatorname{conv} T(D)$  (Theorem 1), which we can easily verify in the current application. We numerically solve the dual problem (D) by the Newton-Raphson algorithm starting from  $\lambda = 0$ .

The results are shown in Figure 1. Our method excels the trapezoidal and Simpson's formulas in the accuracy, at least when the number of grid points  $M$  is not too small. The errors basically decrease as the order of the moment increases.

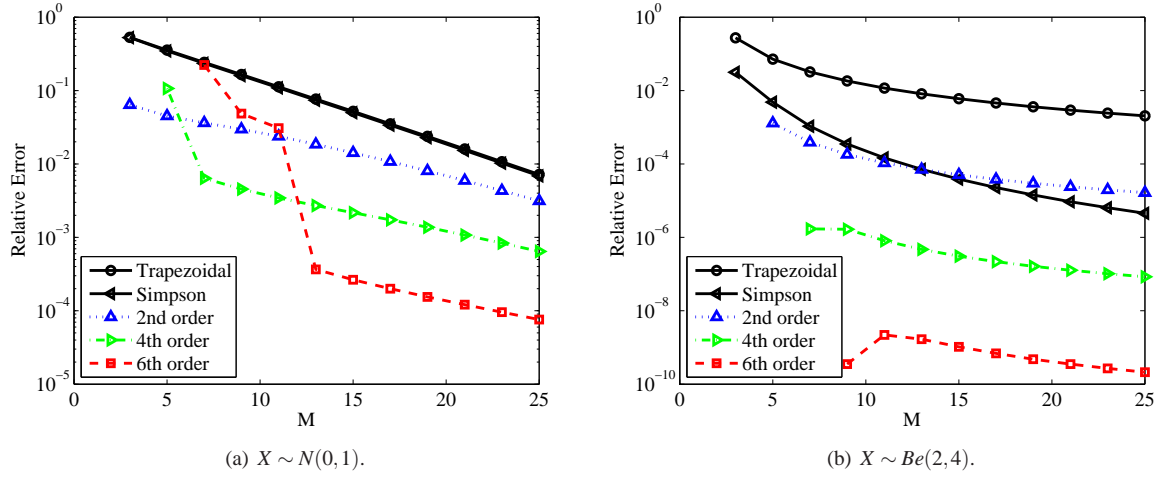
## Optimal portfolio problem

In this section we numerically solve the optimal portfolio problem briefly discussed in the introduction (see [5] for more details). Suppose that there are two assets, stock and bond, with gross returns  $R_1, R_2$ . Asset 1 (stock) is stochastic and lognormally distributed:  $\log R_1 \sim N(\mu, \sigma^2)$ , where  $\mu$  is the expected return and  $\sigma$  is the volatility. Asset 2 (bond) is risk-free and  $\log R_2 = r$ , where  $r$  is the (continuously compounded) interest rate. The optimal portfolio  $\theta$  is determined by the optimization

$$U = \max_{\theta} \frac{1}{1-\gamma} E[(R_1\theta + R_2(1-\theta))^{1-\gamma}], \quad (11)$$

---

<sup>3</sup> Since the beta density is zero at  $x = 0, 1$ , which are included in  $x_{i,M}^{(2)}$ 's, we necessarily have  $p(x_{i,M}^{(2)}) = 0$  for  $i = 1, M$ . Thus, the number of unknown variables is  $M - 2$ , so we need  $M - 2 \geq L + 1 \iff M \geq L + 3$  in the case of the beta distribution.



**FIGURE 1.** Relative errors of the computed values of  $E[g(X)]$ , where  $g(x) = e^x \mathbf{1}_{[-10,10]}(x)$  for  $X \sim N(0,1)$  and  $g(x) = e^x$  for  $X \sim Be(2,4)$ . The legend “Trapezoidal” and “Simpson” represent the relative errors by the trapezoidal and Simpson’s formulas, and “2nd order”, “4th order”, and “6th order” represent those by our method with exact polynomial moments  $E[X^l]$  up to 2nd order ( $l = 1, 2$ ), 4th order ( $l = 1, \dots, 4$ ), and 6th order ( $l = 1, \dots, 6$ ), respectively.

where  $\gamma > 0$  is the relative risk aversion coefficient. We set the parameters such as  $\gamma = 3$ ,  $\mu = 0.07$ ,  $\sigma = 0.2$ , and  $r = 0.01$ . We numerically solve the optimal portfolio problem (11) by applying the trapezoidal and Simpson formulas and our method. To approximate the lognormal distribution, let  $M = 2N + 1$  be the number of grid points ( $N$  is the number of positive grid points) and  $D = \{nh \mid n = 0, \pm 1, \dots, \pm N\}$ , where  $h = 1/\sqrt{N}$  is the grid size. Let  $p(x)$  be the approximating discrete distribution of  $N(0, 1)$  as in the previous subsection (trapezoidal, Simpson, or our method with various moments). Then we put probability  $p(x)$  on the point  $e^{\mu + \sigma x}$  for each  $x \in D$  to obtain the approximate stock return  $R_1$ .

Table 1 shows the optimal portfolio  $\theta$  and its relative error for various moments  $L$  and number of points  $M = 2N + 1$ . The result is somewhat surprising. Even with 3 approximating points ( $N = 1$ ), our method derives an optimal portfolio that is off by only 0.5% to the true value, whereas the trapezoidal and Simpson methods are off by 127% and 220%, respectively. While our method virtually obtains the true value with 9 points ( $N = 4$ , especially when the 4th moment is matched), the trapezoidal and Simpson’s method still have more than 20% of error.

**TABLE 1.** Optimal portfolio and relative error for the trapezoidal, Simpson’s, and our method.  $N$ : number of positive grid points,  $M = 2N + 1$ : total number of grid points,  $L$ : maximum order of moments matched.

Number of grid points		$L = 0$ (Trapezoidal)		Simpson		$L = 2$		$L = 4$	
$N$	$M = 2N + 1$	$\theta$	Error (%)	$\theta$	Error (%)	$\theta$	Error (%)	$\theta$	Error (%)
$1^2 = 1$	3	1.5155	127	2.1377	220	0.6717	0.54	-	-
$2^2 = 4$	9	0.8246	23.4	0.8192	22.6	0.6694	0.20	0.6680	-0.015
$3^2 = 9$	19	0.6830	2.24	0.6821	2.11	0.6684	0.044	0.6681	0
$4^2 = 16$	33	0.6687	0.088	0.6687	0.088	0.6682	0.015	0.6681	0
$5^2 = 25$	51	0.6681	0	0.6681	0	0.6681	0	0.6681	0

The reason why the trapezoidal and Simpson’s methods give poor results when the number of approximating points are small is because the moments are not matched. To see this, taking the first-order condition for the optimal portfolio problem (11), we obtain  $E[(\theta X + R_2)^{-\gamma} X] = 0$ , where  $X = R_1 - R_2$  is the excess return on the stock. Taylor expanding  $(\theta X + R_2)^{-\gamma}$  around  $E[\theta X + R_2]$  and solving for  $\theta$ , after some algebra we get  $\theta = \frac{R_2 E[X]}{\gamma \text{Var}[X] - E[X]^2}$ . Therefore the (approximate) optimal portfolio depends on the first and second moments of the excess return  $X$ . Our method is accurate precisely because we match the moments. In complex economic problem, oftentimes we cannot afford to use many integration points, in which case our method might be useful to obtain accurate results.

## ACKNOWLEDGMENTS

KT was partially supported by JSPS KAKENHI Grant Number 24760064.

## REFERENCES

1. G. Tauchen, *Economics Letters* **20**, 177–181 (1986).
2. J. Adda, and R. W. Cooper, *Dynamic Economics: Quantitative Methods and Applications*, MIT Press, Cambridge, MA, 2003.
3. A. C. Miller, III, and T. R. Rice, *Management Science* **29**, 352–362 (1983).
4. E. A. DeVuyst, and P. V. Preckel, *Mathematical and Computer Modelling* **45**, 787–794 (2007).
5. K. Tanaka, and A. A. Toda, *Economics Letters* **118**, 445–450 (2013).
6. E. T. Jaynes, *Physical Review* **106**, 620–630 (1957).
7. J. M. Borwein, and A. S. Lewis, *SIAM Journal on Control and Optimization* **29**, 325–338 (1991).
8. K. Tanaka, and A. A. Toda, Error estimate and convergence analysis of moment-preserving discrete approximations of continuous distributions (2013), URL <http://arxiv.org/abs/1308.3753>, revise and resubmit in *SIAM Journal on Numerical Analysis*.
9. P. J. Davis, and P. Rabinowitz, *Methods of Numerical Integration*, Academic Press, Orlando, FL, 1984, second edn.