Economics Letters 118 (2013) 445–450

Contents lists available at SciVerse ScienceDirect

Economics Letters

journal homepage: www.elsevier.com/locate/ecolet

Discrete approximations of continuous distributions by maximum entropy

Ken'ichiro Tanaka[a], Alexis Akira Toda[b,*]

[a] School of Systems Information Science, Future University Hakodate, Japan

[b] Department of Economics, Yale University, United States

ARTICLE INFO

Article history:
Received 18 July 2012
Received in revised form
5 December 2012
Accepted 14 December 2012
Available online 22 December 2012

JEL classification:
C63
C65

Keywords:
Discrete approximation
Fenchel duality
Kullback–Leibler information
Maximum entropy principle

ABSTRACT

In numerically implementing the optimization of an expected value in many economic models, it is often necessary to approximate a given continuous probability distribution by a discrete distribution. We propose an approximation method based on the principle of maximum entropy and minimum Kullback–Leibler information, which is computationally very simple. Our method is not intended to replace existing methods but to complement them by "fine-tuning" probabilities so as to match prescribed (not necessarily polynomial) moments exactly.

1. Introduction

How can we find a discrete distribution that approximates a given distribution?

To motivate this question, consider the following problem. There is an investor who lives for two periods with utility function

$$\frac{1}{1-\gamma}\left(c_0^{1-\gamma}+\beta \mathrm{E}[c_1^{1-\gamma}]\right), \tag{1}$$

where $\beta>0$ is the discount factor, $\gamma>0$ is the relative risk aversion coefficient, and c_0, c_1 are consumption for today and tomorrow. The investor is endowed with initial wealth $w>0$ today but nothing tomorrow. After deciding how much to consume today, he can invest his remaining wealth $w-c_0$ in J assets indexed by $j=1,\dots,J$. Asset j has gross return $R_j\geq 0$, which is a random variable. Let θ_j be the fraction of the remaining wealth invested in asset j,[1] $\theta=(\theta_1,\dots,\theta_J)\in\mathbb{R}^J$ $\left(\text{where } \sum_j\theta_j=1\right)$ be the portfolio, and $R(\theta)=\sum_j R_j\theta_j$ be the return on portfolio θ. The budget constraint is then

$$c_1=R(\theta)(w-c_0). \tag{2}$$

The investor's objective is to maximize the expected utility (1) subject to the budget constraint (2).

Characterizing the solution to this problem is not difficult. Substituting (2) into (1), it suffices to solve

$$\max_{c,\theta}\frac{1}{1-\gamma}\left(c^{1-\gamma}+\beta \mathrm{E}[R(\theta)^{1-\gamma}](w-c)^{1-\gamma}\right),$$

which can be broken into

$$\frac{k}{1-\gamma}:=\max_{\theta}\frac{1}{1-\gamma}\mathrm{E}[R(\theta)^{1-\gamma}], \tag{3a}$$

$$U:=\max_{c}\frac{1}{1-\gamma}(c^{1-\gamma}+\beta k(w-c)^{1-\gamma}). \tag{3b}$$

Since the optimal consumption/saving problem (3b) can be solved by calculus, the original problem reduces to solving the optimal portfolio problem (3a).

Solving this problem numerically is not trivial, however, since for most probability distributions the expectation $\mathrm{E}[R(\theta)^{1-\gamma}]$ admits no closed-form expression. However, if the distribution of asset returns $\mathbf{R}=(R_1,\dots,R_J)$ is discrete, then $\mathrm{E}[R(\theta)^{1-\gamma}]$ is simply a sum and we can apply any optimization routine. Therefore, in practice it is important to approximate a given probability distribution by a discrete distribution.

A typical procedure for approximating a probability distribution by a discrete distribution is to partition the range of possible values (Tauchen, 1986) or the range of cumulative probabilities (Adda and Cooper, 2003) into intervals, and then assign the true probability to a representative point (usually the mid-point or the median) of each interval. The advantage of these methods is their simplicity

* Corresponding author. Tel.: +1 203 432 3576.
E-mail addresses: ketanaka@fun.ac.jp (K. Tanaka), alexisakira.toda@yale.edu (A.A. Toda).

1 Of course, the investor is long in asset j if $\theta_j>0$ and short if $\theta_j<0$.

0165-1765/$ – see front matter
doi:10.1016/j.econlet.2012.12.020

and that they work in any dimension. As shown by Miller and Rice (1983), however, this method underestimates the moments of the true distribution. Although the approximated moments approach the true moments as the number of points increases, we cannot always use a large number of points due to the computational cost with the problem we want to solve by discretization (in our case, (3a)). With a small number of points the discrepancy in moments can be substantial. As a remedy, Miller and Rice (1983) propose an approximation method based on Gaussian quadrature that match prescribed moments exactly, but their method works only in one dimension and with polynomial moments. DeVuyst and Preckel (2007) generalize the Gaussian quadrature method (which they call Gaussian cubature) to the multi-dimensional case, which reduces to solving a linear programming problem. However, their method is computationally intensive because the number of unknown variables is equal to that of points, which is typically large, and they do not prove that a solution exists.

In this paper we propose an approximation method based on Jaynes (1957)'s maximum entropy principle (MaxEnt) that matches prescribed moments exactly. Starting from any integration (quadrature) formula, we "fine-tune" the prior probabilities by minimizing the Kullback–Leibler information between the prior and posterior probabilities subject to the prescribed moment constraints. Since the dual problem of finding the probabilities is an unconstrained convex optimization problem with the number of unknown variables equal to the number of moments prescribed, it is computationally very simple. Furthermore, our method works on any discrete set (not necessarily a lattice) of any dimension with any prescribed moments (not necessarily polynomials).

2. Main results

2.1. Formulation of the problem

A researcher wants to approximate a probability density function f on $\mathbb{R}^K$ by probabilities $\{p(x)|x \in D\}$ on a finite discrete set $D \subset \mathbb{R}^K$. The density f may be known or unknown. For instance, f may represent a parametric distribution exogenously specified in the economic model, a nonparametric distribution estimated from data, or a prior distribution in the researcher's mind. In either case, we assume that some moments $\bar{T} = \int T(x)f(x)\mathrm{d}x$ are given, where $T : \mathbb{R}^K \to \mathbb{R}^L$ is a measurable function. For instance, if the first and second moments are given,

$$T(x) = (x_1, \dots, x_K, x_1^2 \dots, x_k x_l, \dots, x_K^2).$$

Since we are prescribed K expectations, K variances, and $\frac{K(K-1)}{2}$ covariances, in this case we have

$$L = K + K + \frac{K(K-1)}{2} = \frac{K(K+3)}{2}.$$

To match these moments with a discrete distribution, it suffices to assign probabilities $\{p(x)|x \in D\}$ such that

$$\sum_{x\in D} T(x)p(x) = \bar{T}.$$

This problem is ill-posed, for in general the number of unknowns ($p(x)$'s), namely $\#D$, is much larger than the number of equations (moments), $L + 1$.[2]

To circumvent this difficulty, we apply Jaynes (1957)'s maximum entropy principle (MaxEnt). Jaynes proposed that when we want to assign probabilities $\mathbf{p} = (p_1, \dots, p_N)$, given some 'background information' (such as moment constraints), we should choose the least informative distribution by maximizing the Shannon (1948) entropy

$$H(\mathbf{p}) = -\sum_{n=1}^{N} p_n \log p_n. \tag{4}$$

MaxEnt has been subsequently generalized and axiomatized in such a way to minimize the Kullback–Leibler information of $\mathbf{p} = (p_1, \dots, p_N)$ given the prior distribution[3] $\mathbf{q} = (q_1, \dots, q_N)$,

$$H(\mathbf{p}|\mathbf{q}) = \sum_{n=1}^{N} p_n \log \frac{p_n}{q_n}, \tag{5}$$

subject to the constraints specified in the underlying problem (Caticha and Giffin, 2006). The Shannon entropy (4) can be interpreted as the (negative of) Kullback–Leibler information (5) corresponding to the uniform prior $\mathbf{q} = (1/N, \dots, 1/N)$ modulo an additive constant. MaxEnt methods can be justified in a number of ways[4] and has been successfully applied in many fields including economics and finance.[5]

The problem can now be formalized as follows. The researcher is given a discrete set $D \subset \mathbb{R}^K$, a prior distribution $\{q(x)|x \in D\}$, a function determining the moments $T : \mathbb{R}^K \to \mathbb{R}^L$, and moments $\bar{T} \in \mathbb{R}^L$. The objective is to obtain the least informative posterior (in the sense of the Kullback–Leibler information) $\{p(x)|x \in D\}$ that matches the prescribed moments, that is,

$$\min_{\{p(x)\}} \sum_{x\in D} p(x) \log \frac{p(x)}{q(x)}$$
$$\text{subject to } \sum_{x\in D} T(x)p(x) = \bar{T}, \sum_{x\in D} p(x) = 1, p(x) \geq 0. \tag{P}$$

Returning to the original problem of approximating a density f, if the true density is totally unknown, there is no reason to discriminate one point $x \in D$ over another, so it is natural to choose the uniform prior $q(x) = 1/N$. If the true density f is known (either exogenously given in the model or nonparametrically estimated from data), we can take D to be a lattice on $\mathbb{R}^K$ and take the prior $q(x) = f(x)/\sum_{x\in D} f(x)$, proportional to the given density. If the researcher already has an integration (quadrature) formula

$$\int g(x)f(x)\mathrm{d}x \approx \sum_{x\in D} w(x)g(x)f(x), \tag{6}$$

where $\{w(x)\}_{x\in D}$ are weights and g is the integrand (here we regard g as a function whose expectation with respect to the density f we want to compute), then one can use $q(x) = w(x)f(x)$. Note that the "proportional" case is a special case by letting the weighting function $w(x)$ be the constant $1/\sum_{x\in D} f(x)$.

As is common in maximum entropy and Bayesian inference, we do not address how to choose the prior $q(x)$ but take the prior $q(x)$ as given. Instead we fine-tune the probabilities $q(x)$ and obtain the optimal (least informative) posterior $p(x)$ by solving the problem (P).

2.2. Solution

The following theorem shows how to solve problem (P).

2 The "+1" comes from accounting the probabilities $\sum_{x\in D} p(x) = 1$.

3 The maximum entropy literature typically refers to the starting distribution $\mathbf{q}$ as "prior". This usage (although it may appear unfamiliar) does not contradict with that of Bayesian inference, for Caticha and Giffin (2006) proved that Bayes's rule is implied by minimizing the K–L information.

4 In this paper we use MaxEnt merely as a tool. Interested readers can refer to Jaynes (2003) for interpretations, Van Campenhout and Cover (1981) for the relation to Bayesian inference, and Shore and Johnson (1980), Caticha and Giffin (2006), and Knuth and Skilling (2010) for axiomatic approaches.

5 See Shore and Johnson (1980) for a short review of applications and Buchen and Kelly (1996), Wu (2003), Veldkamp (2011), and Cabrales et al. (forthcoming) for applications in economics.