Economics Letters 120 (2013) 206–210

Contents lists available at SciVerse ScienceDirect

Economics Letters

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

Easy and flexible mixture distributions

Mogens Fosgerau [a,b,*], Stefan L. Mabit [a]

[a] Technical University of Denmark, Denmark

[b] Centre for Transport Studies, Sweden

HIGHLIGHTS

- We propose a method to generate flexible mixture distributions.
- The method is easy to implement.
- The method is useful for estimating models using simulation.
- We apply the method to estimate mixed logit models.
- We test it with good results in a simulation study and on real data.

ARTICLE INFO

Article history:
Received 14 March 2013
Accepted 30 March 2013
Available online 9 April 2013

JEL classification:
C14
C15
C25

Keywords:
Mixture distributions
Mixed logit
Simulation
Maximum simulated likelihood

ABSTRACT

We propose a method to generate flexible mixture distributions that are useful for estimating models such as the mixed logit model using simulation. The method is easy to implement, yet it can approximate essentially any mixture distribution. We test it with good results in a simulation study and on real data.

1. Introduction

This paper presents an easy yet powerful method for creating a mixture distribution for a random parameter in an econometric model that is estimated using simulation. The method is presented using maximum simulated likelihood estimation of the mixed logit model as an example, but can be applied in a wide range of circumstances. The advantages of the method are that essentially any distribution can be represented arbitrarily well, while implementation is very simple.

Consider a model that specifies the likelihood $P(y|x, \beta)$ of some outcome y conditional on variables x and an unobserved random parameter β having distribution F.[1] Assuming that x and β are independent, the likelihood $P(y|x)$ may be simulated given R independent draws β_r from F. This is the basis for estimation by simulation (Train, 2003; McFadden, 1989), which can be applied when the distribution F is considered as known.

Most applications of this method rely on the inversion method for generating draws from F: if u_r are draws from a standard uniform distribution, then $F^{-1}(u_r)$ are draws from F. In order to use this method, it is necessary to compute the inverse of F explicitly.[2]

There are many situations where it is not desirable to impose a specific functional form on F. Generally, this is the case whenever the choice of F has impact on the object of interest for the investigation but there is no a priori reason to choose a particular F. It is particularly undesirable to impose a specific form on F when F is the object of interest itself, e.g., when the purpose is to estimate a distribution of willingness-to-pay. Then it is preferable

* Correspondence to: DTU Transport, Bygningstorvet, Building 116B, 2800 Kgs. Lyngby, Denmark. Tel.: +45 45256521.

E-mail addresses: mf@transport.dtu.dk (M. Fosgerau), smab@transport.dtu.dk (S.L. Mabit).

[1] There will generally be other parameters to be estimated in the likelihood. They are suppressed in the notation here as the focus lies elsewhere.

[2] Devroye (1986) provides a comprehensive treatment of techniques for random variable generation.

0165-1765/$ – see front matter © 2013 Elsevier B.V. All rights reserved.
http://dx.doi.org/10.1016/j.econlet.2013.03.050

if the shape of F can be estimated. This can be accomplished by the method of sieves (see e.g. Chen, 2007; Gallant and Nychka, 1987), also known as series estimators. It is however necessary to guarantee that the approximation of F is actually a CDF and then it must be inverted in order to generate random draws from F using the inversion method.

Another idea is to approximate F^{-1} directly. Then inversion is unnecessary. It is however still necessary to ensure that F^{-1} is monotone, which might involve somewhat complicated restrictions on the deep parameters of F^{-1} in a series approximation.

The key insight of this paper is that approximating F or F^{-1} is actually an unnecessary complication for the present purpose. All that is required for simulating the likelihood is draws β_r from some distribution F that depends on some deep parameters to be estimated. The simulated likelihood is simply

$$\frac{1}{R}\sum_r P\left(y|x,\beta_r\right). \tag{1}$$

It is not necessary that the draws β_r are monotone functions of standard uniform draws. It is not even necessary to know explicitly the distribution of the draws β_r in order to compute (1); the ability to generate draws from the distribution is sufficient. Being able to obtain the draws, it is always possible to estimate their distribution.

In this paper we take draws u_r from some distribution and transform them using a power series

$$f\left(u|\alpha\right) = \sum_{k=0}^{K} \alpha_k u^k \tag{2}$$

to compute random draws $\beta_r = f\left(u_r|\alpha\right)$ that depend on deep parameters $\alpha = (\alpha_0, \ldots, \alpha_K)$ to be estimated. The random draws are inserted into (1) and the resulting expression is very easy to implement in software. For instance, if the model contains a term βx, then that is replaced by $\sum_{k=0}^{K} \alpha_k \left(x u_r^k\right)$. This is a convenient form, since it is linear in deep parameters α that are multiplied by easily computed variables $x u_r^k$. In most cases the distribution of $f\left(u|\alpha\right)$ is not easily derived analytically. The distribution is by construction, however, very easy to simulate, which is all that is really needed.

A predecessor of our method is Fleishman (1978), who considers the problem of generating random variables with prespecified moments. He generates a random variable as a third-order polynomial in a standard normal random variable and provides formulae for the coefficients of the polynomial such that specific values of the first four moments are matched by such a variable. The present case is similar, except we are not concerned with matching given moments, but estimate coefficients in order to match a given dataset and may use polynomials of any degree. We present results using both uniform and normal draws.

The following Section 2 presents some properties of the proposed method. It will also be argued that essentially any distribution can be approximated arbitrarily well by (2) by choosing a sufficiently large number of parameters K. This section also discusses extension to multivariate random parameter distributions. Section 3 provides simulation results that illustrate the ability of the method to recover various true distributions from binary discrete choice panel data. Section 4 presents an application to real data and Section 5 concludes.

2. Some properties of the method

Let $\alpha = (\alpha_0, \ldots, a_K) \in \mathbb{R}^K$ be a parameter vector and let u be a random variable. Then $\beta = f\left(u|\alpha\right) = \sum_{k=0}^{K} \alpha_k u^k$ is a random variable and it is convenient for use as a random parameter. The following proposition summarises a few properties of β.

Proposition 1. *Let u follow a uniform distribution. Then the random parameter β has compact support ranging between α_0 and $\sum_{k=0}^{K} \alpha_k$, either of which may be greatest; the mean is*

$$E\beta = \sum_{k=0}^{K} \frac{\alpha_k}{1+k},$$

and the m'th raw moment $(m > 1)$ is

$$E\left(\beta^m\right) = \sum_{k_1=0,\ldots,k_m=0}^{K,\ldots,K} \frac{\prod_{i=1}^{m} \alpha_{k_i}}{1+\sum_{i=1}^{m} k_i}.$$

The variance of β is

$$\begin{aligned} V\left(\beta\right) &= E\left(\beta^2\right) - \left(E\beta\right)^2 \\ &= \sum_{k=0,j=0}^{K,K} \frac{kj\alpha_k\alpha_j}{\left(1+k+j\right)\left(1+k\right)\left(1+j\right)}. \end{aligned}$$

Proof. Immediate. □

Remark 1. It is straightforward (but quite tedious) to show that with uniform u and $K = 2$, then it is possible to attain any skewness while maintaining that $E\beta = 0$ and $E\left(\beta^2\right) = 1$.

Remark 2. If the first K moments are to be matched, it may be necessary to include more than K terms. The necessity of this has been shown for a third-order polynomial in a standard normal random variable (Headrick, 2002).

Remark 3. By the Weierstrass approximation theorem, the set of functions $\left\{f\left(\cdot|\alpha\right)|\alpha \in \mathbb{R}^{(\mathbb{N})}\right\}$ uniformly approximates any continuous function on the unit interval. This comprises all inverse CDF of distributions that have densities.

Remark 4. Consistency of series estimators has been established for a range of cases (see e.g. Geman and Hwang, 1982; Chen, 2007; Bierens, 2008; Fosgerau and Nielsen, 2010), but not formally for the present. Consistency of the proposed estimator seems highly likely, meaning that the estimated distribution of β will become arbitrarily close to the true distribution given a large enough dataset and a correspondingly large value of K. For a fixed K, the standard results regarding consistency of maximum simulated likelihood apply (Newey and McFadden, 1994; Hajivassiliou and Ruud, 1993).

Remark 5. Given R i.i.d. draws β_r from some distribution, its CDF F can be estimated by

$$F\left(t\right) = E\left(1\left\{\beta \le t\right\}\right) \simeq \frac{1}{R}\sum_r 1\left\{\beta_r \le t\right\}.$$

As β_r are the results of simulation, we are free to choose R and hence it can be chosen to achieve any desired degree of precision of the estimate of F.

2.1. Multivariate distributions

The method can be extended to allow for a multivariate random parameter. The extension is straightforward if the random parameters are independent, so in the following we allow them to be dependent.

One way to go is to combine the proposed method with a copula. Let c be the density of a bivariate copula function, i.e. a density on the unit cube with uniform marginal distributions. A range of