近三年统计本科生将毕业论文成果发表国际期刊论文

Statistical Papers
https://doi.org/10.1007/s00362-021-01236-1

REGULAR ARTICLE

Properties and generation of representative points of the exponential distribution

Long-Hao Xu[1] · Kai-Tai Fang[1,2] · Ping He[1]

Received: 27 August 2020 / Revised: 27 March 2021 / Accepted: 19 April 2021

Abstract
It is known that the exponential distribution has many nice properties. Graf and Luschgy (2000) pointed out that the mean squared error of the set of representative points of the exponential distribution is fully determined by the smallest representative point. In this paper we concern with the representative points of the exponential distribution and find a number of new interesting properties. A new algorithm is proposed to effectively generate representative points of the exponential distribution. In addition, the performance of representative points of the exponential distribution is evaluated.

Keywords Discrete approximation · Exponential distribution · Mean squared error · Principal points · Representative points

1 Introduction

The problem of selecting a given number of representative points (RPs for short) which retain as much information of the population as possible arises in many situations. It can also be considered as a problem of approximating a continuous distribution by a discrete distribution. Let X be a univariate random variable with probability density function (p.d.f.) $p(x)$ and cumulative distribution function (c.d.f.) $F(x)$. We want to find a discrete random variable Y shown as below to represent the continuous random

✉ Ping He
heping@uic.edu.cn

Long-Hao Xu
893405062@qq.com

Kai-Tai Fang
ktfang@uic.edu.cn

1 Division of Science and Technology, BNU-HKBU United International College, Zhuhai 519085, China

2 The Key Lab of Random Complex Structures and Data Analysis, The Chinese Academy of Sciences, Beijing 100864, China

Published online: 10 May 2021

COMMUNICATIONS IN STATISTICS - SIMULATION AND COMPUTATION®
https://doi.org/10.1080/03610918.2021.1938121

A novel algorithm for generating minimum energy points from identically charged particles in 1D, 2D and 3D unit hypercubes

A. M. Elsawah[a,b], Li Meng Hua[a,c], and Kai-Tai Fang[a,d]

[a]Division of Science and Technology, Beijing Normal University–Hong Kong Baptist University United International College, Zhuhai, China; [b]Department of Mathematics, Faculty of Science, Zagazig University, Zagazig, Egypt; [c]Chinasoft International Company, Shenzhen, China; [d]The Key Lab of Random Complex Structures and Data Analysis, The Chinese Academy of Sciences, Beijing, China

ABSTRACT

Generating minimum energy points (MEPs) is an optimal solution of many real-world problems, such as the selection of best locations for hospitals inside a city that reduce the overcrowding and competition and avoid the less-populated regions. The key idea is considering these locations as charged particles with the same sign (i.e., repel each other) inside a box and distribute these points by minimizing the total electric potential energy (TEPE) among them. The practice demonstrated that most of the existing techniques for generating MEPs are complex, especially for non-mathematicians. Therefore, the greedy algorithm (GreA) is the classical widely used algorithm for its simplicity even though a satisfactory result is not guaranteed. This paper gives a novel algorithm for generating MEPs from identically charged particles in 1D, 2D and 3D unit hypercubes. The results show that the new algorithm distributes the points far away from each other to reduce the TEPE of the generated MEPs more effectively than the GreA. The new algorithm is a significant improvement of the GreA to overcome its unsatisfactory results. Therefore, the new algorithm in its current form or after some improvements is highly recommended to be used instead of the GreA for many different applications.

ARTICLE HISTORY
Received 4 June 2020
Accepted 29 May 2021

KEYWORDS
DEoptim; Electric field lines; Electric potential energy; Equipotential circles; GenSA; Greedy algorithm; Minimum energy points; PSO; Representative points

MATHEMATICAL SUBJECT CLASSIFICATION
1.00940; 1.00950; 1.00990

1. Introduction

Consider the following real-life problem of selecting the best locations to open gas stations in a new city. How to find the best locations that will avoid the less-populated regions and minimize the competition among these gas stations? For solving this significant real-life problem, consider the new city (experimental region) as a box and the locations of the gas stations are points inside this box. To effectively distribute these points inside this box, consider that these points are charged particles with the same sign and the charge represents the experimental response (inversely proportional to the population density). Therefore, these points will repel each other and try to be as far away as possible from each other. The distributed points will take positions inside the box so as to minimize the total potential energy in electrodynamics (Thomson's theorem, see, e.g., Zhou 1999). These points are called minimum energy points (MEPs). MEPs are

CONTACT A. M. Elsawah a_elsawah85@yahoo.com or amelsawah@uic.edu.cn or a.elsawah@zu.edu.eg Division of Science and Technology, Beijing Normal University–Hong Kong Baptist University United International College, Zhuhai 519085, China.

Supplemental data for this article is available online at https://doi.org/10.1080/03610918.2021.1938121

COMMUNICATIONS IN STATISTICS—THEORY AND METHODS
https://doi.org/10.1080/03610926.2021.1970772

Taylor & Francis
Taylor & Francis Group

Check for updates

Limiting behavior of the gap between the largest two representative points of statistical distributions

Long-Hao Xu[a], Kai-Tai Fang[a,b], and Jianxin Pan[c]

[a]Division of Science and Technology, BNU-HKBU United International College, Zhuhai, China; [b]The Key Lab of Random Complex Structures and Data Analysis, The Chinese Academy of Sciences, Beijing, China; [c]Department of Mathematics, The University of Manchester, Manchester, UK

ABSTRACT

This paper explores the properties of the gap of representative points (RPs) in the sense of minimum mean square error for various univariate statistical distributions. We illustrate the relationship between RPs and doubly truncated mean residual life (DMRL) as well as mean residual life (MRL), which are widely used in survival analysis. The limiting behavior of the gap between the largest two RPs is discussed. In addition, an upper bound of the optimal MSE is given when the univariate random variable *X* has a domain on finite interval. In simulation studies, the performance of RPs for various distributions is assessed in terms of moment estimation and resampling technique. A brief discussion about the relationship between the tail of the distribution and the gap of RPs is also given.

ARTICLE HISTORY
Received 6 January 2021
Accepted 15 August 2021

KEYWORDS
Discrete approximation; mean residual life; mean square error; principal points; representative points; survival analysis

2020 MATHEMATICAL SUBJECT CLASSIFICATION
62E17

1. Introduction

It is often to request to find a discrete distribution to approximate a given continuous distribution with retaining information as much as possible. Let X be a continuous random variable with probability density function $f(x)$ and $\mathbb{E}(X^2) < \infty$. One wants to use a discrete random variable Y shown in Equation (1) to represent X, where $\mathbb{P}(Y = y_i) = p_i > 0, i = 1, \ldots, n$ and $y_1 < \ldots < y_n$.

$$\begin{array}{c|ccc} Y & y_1 & \cdots & y_n \\ \hline p & p_1 & \cdots & p_n \end{array} \tag{1}$$

The concept of representative points in the sense of minimum mean square error (MSE RPs or RPs for short) has been widely used to solve the problem (Cox 1957; Fang and He 1982; Flury 1990). The representative points are also called principal points (Flury 1990, 1993; Tarpey 1995; Tarpey, Li and Flury 1995) and quantizer (Max 1960; Lloyd 1982; Graf and Luschgy 2007) in the literature. There are many applications of RPs in signal compression (Max 1960; Lloyd 1982), cluster analysis (Anderberg 1973), statistical simulation (Fang, Zhou, and Wang 2014; Lemaire, Montes, and Pagès 2020), and numerical integration (Pagès 1998; Pagès and Printems 2003; Pagès 2015).

CONTACT Jianxin Pan Jianxin.Pan@manchester.ac.uk Department of Mathematics, The University of Manchester, Oxford Road, Manchester M13 9PL, UK.