Available online at www.sciencedirect.com
ScienceDirect
Computational Statistics & Data Analysis 51 (2007) 4717–4730
COMPUTATIONAL STATISTICS & DATA ANALYSIS
www.elsevier.com/locate/csda

Random Riemann Sum estimator versus Monte Carlo

Henar Urmeneta*, Víctor Hernández

Dpto. de Estadística e Investigación Operativa, Universidad Nacional de Educación a Distancia, Madrid, Spain

Received 10 May 2005; received in revised form 4 April 2006; accepted 11 September 2006
Available online 25 October 2006

Abstract

Two estimators of the expectation of a function, the classical based in Monte Carlo sampling method and one based in Random Riemann Sums, are compared. It presents the differences on bias, variance, convergence and mainly convergence rates. Two ways of sampling to obtain a Random Riemann Sum estimator are given. The first one provides a sequence of estimations whose terms are independent, this fact produces a loss of order one in the convergence rate for the strong law compared with Monte Carlo sampling method. The second one is considered in order to improve these results.

PACS: 60F15; 60G50; 28A25

Keywords: Random Riemann Sum; Triangular array; Regular cover; Rates of convergence in the law of large numbers

1. Introduction

The classical estimator of the integral $I = \int g \, \mathrm{d}F$, where F is the distribution function of the random vector X, which takes values on R^k, and $g : R^k \to R$ is a finite-valued measurable function, is the average $\widehat{I}_n = (1/n)\sum_{k=1}^{n} g(X_k)$, with $\{X_n,\ n \geqslant 1\}$ a sequence of independent copies of X. This estimator, $\widehat{I}_n$, is called Monte Carlo estimator.

For certain functions g this estimator does not seem to be "intelligent"; when g takes very large values at the points of a region B such that $P(X \in B)$ is very small, we can expect that only few sampled points are in B and the estimation can be very smaller than the value of I. In these conditions, although the law of large numbers assures that $\widehat{I}_n \xrightarrow{\text{a.s.}} I$, as $n \to \infty$, it seems reasonable to expect an estimation error too big, at least for small samples.

Pruss (1996) considered another estimator of I. The idea is to divide R^k into n parts, $I_{n1}, I_{n2}, \ldots, I_{nn}$, so that $P(X \in I_{nk}) = 1/n$, $k = 1, \ldots, n$, and take a random point in each element of the partition. Let $X_{n1}, X_{n2}, \ldots, X_{nn}$ be such random points, the Pruss estimator is $\widetilde{I}_n = (1/n)\sum_{k=1}^{n} g(X_{nk})$. The advantage of this estimator over Monte Carlo's one is that the sampled points always lie well spread according to the distribution of X.

Every estimator value, $(1/n)\sum_{k=1}^{n} g(x_{nk})$, obtained by the Pruss method can be interpreted as an Riemann sum where the range of the variable is divided into the n parts with the same probability respecting of F, not necessarily with area equal to $1/n$, and the point corresponding to each part, x_{nk}, is chosen randomly. For this reason $\widetilde{I}_n$ is called Random Riemann Sum (RRS) of the function g.

* Corresponding author. Dpto. Estadística e Investigación Operativa, Universidad Pública de Navarra, Campus Arrosadía, 31006-Pamplona, Spain. Tel.: +34 94 8169194; fax:+34 94 8169204.
E-mail addresses: henar@unavarra.es (H. Urmeneta), victorher@ccia.uned.es (V. Hernández).

0167-9473/$ - see front matter © 2006 Elsevier B.V. All rights reserved.
doi:10.1016/j.csda.2006.09.041

The aim of our paper is to compare both estimators about several aspects: bias, variance and asymptotic behavior (convergence and convergence rates). It shows two empirical examples where the RRS estimator is much better than Monte carlo one in the short-term.

The paper is organized as follows: in the second section a brief summary of the results of the Monte Carlo estimator is presented. The third section is devoted to study the RRS estimator. We present the framework to analyze the characteristics of the RRS estimator, triangular rowwise independent arrays of which rows are a regular cover, and results about bias, variance, convergence and convergence rates. Also, we present the comparison with the results of Monte Carlo estimator, some differences are showed by several simulated examples. In the fourth section we present another way, based in Kieffer and Stanojević (1982), to obtain a RRS estimator in order to improve the strong law results. We prove results to compare it with the obtained in the previous sections. Finally, in the fifth section the conclusions are presented.

2. Monte Carlo estimator

Given $I = \int g \, \mathrm{d}F$, where F is the distribution function of the random vector X defined on the probability space $(\Omega, \mathscr{A}, P)$ in R^k and $g : R^k \longrightarrow R$ is a finite-valued measurable function, the average $\widehat{I}_n = (1/n)\sum_{k=1}^{n} g(X_k)$, with $\{X_n,\ n \geqslant 1\}$ a sequence of independent copies of X, is a classical estimator of I called Monte Carlo estimator.

It is well-known that $\widehat{I}_n$ is an unbiased estimator with variance

$$V(\widehat{I}_n) = \frac{1}{n} E g^2(X) - \frac{1}{n} E^2 g(X). \tag{1}$$

The strongest convergence result is the Kolmogorov theorem (see, for instance, Shiryaev, 1996, Theorem IV.3.3) which characterizes the almost sure convergence by a moment condition:

Theorem 1. *$E|g(X)| < \infty$ is a necessary and sufficient condition to*

$$\widehat{I}_n \xrightarrow{\text{a.s.}} I.$$

We note that the way of obtaining the Monte Carlo estimator with n observations takes into account the $n-1$ points previously sampled, to obtain $\widehat{I}_{n-1}$, and the new sampled observation, $g(X_n)$. Thus, the equation $n\widehat{I}_n = (n-1)\widehat{I}_{n-1} + g(X_n)$ describes the relation between the elements of the sequence $\{\widehat{I}_n\}$, this is a crucial fact to prove the strong law results.

Baum and Katz (1965) characterized, by means of moment conditions, rates of convergence for Marcinkiewicz–Zygmund weak and strong laws (Baum and Katz, 1965, Theorems 1–3). These results provide the following convergence rates for the weak law:

Theorem 2. *If $p \geqslant -1$ then, $E|g(X)|^{p+2} < \infty$ is a necessary and sufficient condition to*

$$\sum_{n=1}^{\infty} n^p P(|\widehat{I}_n - I| > \varepsilon) < \infty \quad \textit{for all } \varepsilon > 0.$$

And the followings ones for the strong Law:

Theorem 3. *If $p > -1$ then, $E|g(X)|^{p+2} < \infty$ is a necessary and sufficient condition to*

$$\sum_{n=1}^{\infty} n^p P\left(\sup_{k \geqslant n} |\widehat{I}_k - I| > \varepsilon\right) < \infty \quad \textit{for all } \varepsilon > 0.$$

Theorem 4. *The condition $E|g(X)| \log^+ |X| < \infty$, where $\log^+ x = \max\{0, \log x\}$, is a necessary and sufficient condition to*

$$\sum_{n=1}^{\infty} \frac{1}{n} P\left(\sup_{k \geqslant n} |\widehat{I}_k - I| > \varepsilon\right) < \infty \quad \textit{for all } \varepsilon > 0.$$