Monografías del Seminario Matemático García de Galdeano **33**, 325–330 (2006)

Convergence of the approximations of an integral by sums of random variables

Henar Urmeneta and Víctor Hernández

Abstract. Numeric methods of approximation on an integral suggest, in a natural way, random methods of approximation by just evaluating approximation rules in random points.

In Einmahl-Van Zuijlen (1997) random approximations of the integral of a smooth function on $[0,1]$, based on the trapezoidal rule and Simpson's rule, are proposed. They obtain results for the convergence in probability, in the first case, and for the convergence in distribution, in the second case. The purpose of the present paper is to prove similar results for the almost sure convergence. We obtain the rate of almost sure convergence when smooth-conditions on f are holden.

Keywords: Numerical integration, order statistics, convergence of sums of random variables, convergence rates.

AMS classification: 60F05, 60F15, 62G30.

§1. Introduction

The approximation of integrals is an old problem of the numerical analysis. The classical methods of the numerical analysis are deterministic, fixed the number of points, the approximated value and the error bound are determinate. Also there exist random methods to approximate an integral, which consist in simulating the problem by random samples and justifying its by means of laws of large numbers. In both cases, deterministic and random methods, we would state the approximation together with a bound on the error in the approximation. The analysis of error is an intrinsic part of the application of any approximation method.

There are several methods based on multiple evaluations of the integrand function f. In these cases, the main idea is that the error can be proved to go to 0 as the number of function evaluations increase. Even more, we would examine how the error changes for an increasing number of function evaluation to analyze the convergence velocity.

Most of random methods estimate the integral I of a function f by means of sums like

$$I_n = \sum_{i=1}^{n} \omega_i f(X_i),$$

where X_i are random points where the function f has to be evaluated and w_i are real numbers, for $i = 1, \dots, n$.

The analysis of the error requires the study of convergence, in probability, in distribution and almost sure, of estimation error, $I_n - I$, for increasing n. The analysis of the convergence velocity require study rates of convergence, for instance, results as $I_n - I = O_{a.s.}(1/n^k)$ or $I_n - I = O_{p.}(1/n^k)$, where $\alpha_n = O_{a.s.(p.)}\beta_n$ means $|\alpha_n/\beta_n| \xrightarrow{a.s.(p.)} K$, with $K \in \mathbb{R}^+$.

Classical numerical rules of approximation of an integral suggest in a natural way random methods of approximation by just evaluating approximation rules in random points. The classical polynomial interpolation rules approximates the integral $\int_a^b f(x)\,dx$ by means of this other $\int_a^b P_n(x)\,dx$, where P_n is an interpolating polynomial of the function f in (a,b). These approximation takes the form

$$I_n = \sum_{i=1}^{n} \omega_i f(x_i),$$

where the x_i, with $i = 1,\dots,n$, are points in (a,b) chosen by means of a deterministic criteria. The idea consists in substituting in the rule each x_i by a random point X_i, for all $i = 1,\dots,n$.

§2. Approximations based on the trapezoidal rule

The approximation of $I = \int_0^1 f(x)dx$ by using trapezoidal rule is obtained dividing $[0,1]$ into $n+1$ parts by means of n equidistant points and summing the areas of the trapezoids which bases are each part and which heights are the images of the function f at the points that determine each part:

$$\frac{1}{2}\sum_{i=1}^{n+1} \frac{1}{n+1}\left(f\left(\frac{i-1}{n+1}\right) + f\left(\frac{i}{n+1}\right)\right).$$

If $f \in C^2[0,1]$, the approximation error in each subinterval $[a,b]$ is $1/12(b-a)^3 f''(c)$, where $c \in (a,b)$. Therefore, the total approximation error is

$$\frac{1}{12}\sum_{i=1}^{n+1}\left(\frac{1}{n+1}\right)^3 f''(\xi_i), \quad \text{with } \xi_i \in \left(\frac{i-1}{n+1}, \frac{i}{n+1}\right). \tag{1}$$

Suppose we can only observe the values of f at the independent uniformly distributed random points $X_1, X_2, \dots, X_n$. We can apply the trapezoidal rule with $U_{n,1}, U_{n,2}, \dots, U_{n,n}$, the order statistics of $X_1, X_2, \dots, X_n$, instead of the equidistant points. Thus, by only using $(U_{n,i}, f(U_{n,i}))$, $i = 0,2,\dots,n+1$, where $U_{n,0} = 0$ and $U_{n,n+1} = 1$, it proposes the random approximation:

$$\widehat{I_n} = \frac{1}{2}\sum_{i=1}^{n+1} D_{n,i}\left(f(U_{n,i-1}) + f(U_{n,i})\right),$$

where $D_{n,i} = U_{n,i} - U_{n,i-1}$.

In [1], the following result about the convergence in probability of $\widehat{I_n}$ to I is proved.

Theorem 1. *If f''' is bounded in $[0,1]$, then*

$$n^2(\widehat{I_n} - I) \xrightarrow{p} 1/2(f'(1) - f'(0))$$

when $n \to \infty$.

Therefore, we have the following convergence rate for the convergence in probability:

$$\widehat{I}_n - I = O_{p.}(1/n^2).$$

We show in the following theorem the same rate for the almost sure convergence, relaxing the condition about the smoothness of f.

Theorem 2. *If $f \in C^2[0,1]$, then $\widehat{I}_n - I = O_{a.s.}(1/n^2)$.*

Proof. The following lemma is well-known and will be used in the proof (it can be found in, e.g., [4, p. 721]).

Lemma 3.

$$(D_{n,1}, D_{n,2}, \ldots, D_{n,n}) \stackrel{d}{=} \left(\frac{\alpha_1}{\sum_{i=1}^{n+1} \alpha_i}, \frac{\alpha_2}{\sum_{i=1}^{n+1} \alpha_i}, \ldots, \frac{\alpha_{n+1}}{\sum_{i=1}^{n+1} \alpha_i} \right),$$

where α_i are independent exponential random variables with mean $\lambda = 1$.

From (1) we have that

$$n^2 |\widehat{I}_n - I| = \left| \frac{n^2}{12} \sum_{i=1}^{n+1} D_{n,i}^3 f''(\tilde{U}_{n,i}) \right| \le \frac{n^2}{12} \sum_{i=1}^{n+1} D_{n,i}^3 |f''(\tilde{U}_{n,i})| \le M \frac{n^2}{12} \sum_{i=1}^{n+1} D_{n,i}^3,$$

where $\tilde{U}_{n,i} \in (U_{n,i-1}, U_{n,i})$ and M a positive constant.

Now, by Lemma 3 we have

$$\sum_{i=1}^{n+1} D_{n,i}^3 \stackrel{d}{=} \sum_{i=1}^{n+1} \frac{\alpha_i^3}{S_{n+1}^3},$$

where $\{\alpha_k\}_{k=1}^{\infty}$ is a sequence of independent exponential random variables with mean $\lambda = 1$ and $S_n = \sum_{k=1}^{n} \alpha_k$; therefore, we have that

$$n^2 \sum_{i=1}^{n+1} D_{n,i}^3 \stackrel{d}{=} \left(\frac{n}{S_{n+1}} \right)^3 \frac{\sum_{i=1}^{n+1} \alpha_i^3}{n}.$$

Since $E(\alpha_i) = 1$, where E denotes expectation, and $E(\alpha_i^3) = 6$, by strong Law of Kolmogorov it follows that

$$\frac{S_{n+1}}{n} = \frac{n+1}{n} \frac{S_{n+1}}{n+1} \xrightarrow{a.s.} E(\alpha_i) = 1$$

and

$$\frac{\sum_{i=1}^{n+1} \alpha_i^3}{n} = \frac{n+1}{n} \frac{\sum_{i=1}^{n+1} \alpha_i^3}{n+1} \xrightarrow{a.s.} E(\alpha_i^3) = 6.$$

Therefore

$$n^2 \sum_{i=1}^{n+1} D_{n,i}^3 \xrightarrow{a.s.} 6.$$

Thus, we have

$$\widehat{I}_n - I = O_{a.s.}(1/n^2). \quad \square$$

§3. Approximations based on the Simpson's rule

A much better estimator is obtained by applying a 3-points formula, i.e., in each subinterval (a,b) is given $c \in (a,b)$, $\int_a^b f(x)\,dx$ is approximated by $\int_a^b P_n(x)\,dx$, where P_n is an interpolating polynomial of degree 2 of the function f at the points a, b and c. If the points a, b and c are equidistant, this approximation is know as Simpson's rule and in this case it has that

$$\int_a^b f(x)dx \simeq \int_a^b P_n(x)dx = \frac{1}{6}(b-a)\left(2-\frac{b-c}{c-a}\right) f(a) + \frac{1}{6}\frac{(b-a)^3}{(c-a)(b-c)} f(c) + \frac{1}{6}(b-a)\left(2-\frac{c-a}{b-c}\right) f(b).$$

If $f \in C^3[0,1]$ the approximation error can be written as follows:

$$\int_a^b f(x)dx - \int_a^b P_n(x)dx = -\frac{1}{6}\int_a^b (x-a)(x-b)(x-c) f^3(\eta)dx,$$

where $\eta = \eta(x) \in (a,b)$ (see, e.g., [3, p. 304]).

In [1], an estimator of $I = \int_0^1 f(x)dx$ based on Simpson's rule is proposed. It takes independent uniformly distributed random points $X_1, X_2, \ldots, X_n$, where, for convenience, n is taken to be odd. They apply the Simpson's rule again with $U_{n,1}, U_{n,2}, \ldots, U_{n,n}$, the order statistics of $X_1, X_2, \ldots, X_n$, instead of the equidistant points. Therefore it proposes the random approximation:

$$\widetilde{I}_n = \sum_{i=1}^{\frac{n+1}{2}} \left\{ \frac{1}{6}(D_{n,2i-1}+D_{n,2i})\left(2-\frac{D_{n,2i}}{D_{n,2i-1}}\right) f(U_{n,2i-2}) + \frac{1}{6}\frac{(D_{n,2i-1}+D_{n,2i})^3}{D_{n,2i-1}D_{n,2i}} f(U_{n,2i-1}) + \frac{1}{6}(D_{n,2i-1}+D_{n,2i})\left(2-\frac{D_{n,2i-1}}{D_{n,2i}}\right) f(U_{n,2i}) \right\}.$$

If $f \in C^3[0,1]$, the total error can be written as follows:

$$\widetilde{I}_n - I = \sum_{i=1}^{\frac{n+1}{2}} \frac{1}{6}\int_{U_{n,2i-2}}^{U_{n,2i}} (x-U_{n,2i-2})(x-U_{n,2i-1})(x-U_{n,2i}) f^3(\eta_{n,i})\,dx, \tag{2}$$

with $\eta_{n,i} = \eta_{n,i}(x) \in (U_{n,2i-2}, U_{n,2i})$.

In [1], the following result about the convergence in distribution of $\widetilde{I}_n$ to I is proved.

Theorem 4. *If f^{v} is bounded in $[0,1]$, then*

$$n^{7/2}(I_n - I) \xrightarrow{d} \sqrt{\frac{35}{3}\int_0^1 (f^{3)}(x))^2 dx}\; Z$$

where Z is a standard normal random variable.

In the following theorem we obtain a rate for the almost sure convergence for $\widetilde{I}_n$.

Theorem 5. *If* $f \in C^3[0,1]$, *then* $\widetilde{I}_n - I = O_{a.s.}(1/n^3)$.

Proof. From (2), we have that

$$\begin{aligned} n^3|\widetilde{I}_n - I| &= n^3\left|\sum_{i=1}^{\frac{n+1}{2}} \frac{1}{6}\int_{U_{n,2i-2}}^{U_{n,2i}} (x-U_{n,2i-2})(x-U_{n,2i-1})(x-U_{n,2i}) f^3(\eta_{n,i})\,dx\right| \\ &\le n^3 \sum_{i=1}^{\frac{n+1}{2}} \frac{1}{6}\int_{U_{n,2i-2}}^{U_{n,2i}} (x-U_{n,2i-2})|x-U_{n,2i-1}|(U_{n,2i}-x)|\,f^3(\eta_{n,i})|\,dx \\ &\le \frac{M}{6}\sum_{i=1}^{\frac{n+1}{2}} (D_{n,2i-1}+D_{n,2i})^4, \end{aligned}$$

where $\{\alpha_k\}_{k=1}^{\infty}$ is a sequence of independent exponential random variables with mean $\lambda = 1$ and $S_n = \sum_{k=1}^{n} \alpha_k$.

From Lemma 3 it follows that

$$(D_{n,2i-1}+D_{n,2i})^4 \overset{d}{=} \frac{(\alpha_{2i-1}+\alpha_{2i})^4}{S_{n+1}^4}.$$

Therefore, we have that

$$n^3 \sum_{i=1}^{\frac{n+1}{2}} (D_{n,2i-1}+D_{n,2i})^4 \overset{d}{=} \left(\frac{n}{S_{n+1}}\right)^4 \frac{\sum_{i=1}^{\frac{n+1}{2}} (\alpha_{2i-1}+\alpha_{2i})^4}{n}.$$

The sequence $\{\alpha_{2i-1}+\alpha_{2i}\}$, with $i = 1, 2, \ldots$, is a sequence of independent and identically distributed random variables with finite mean. Then, by Strong Law of Kolmogorov, it follows that

$$\left(\frac{n}{S_{n+1}}\right)^4 \frac{\sum_{i=1}^{\frac{n+1}{2}} (\alpha_{2i-1}+\alpha_{2i})^4}{n} \xrightarrow{a.s.} K. \quad \square$$

Remark 1. This method seems very relevant because it allows us estimate the mean of a function f suppose that U_i represents some uncontrollable random quantity, like temperature, humidity or light intensity with a known distribution function G having density g, not necessarily the Uniform distribution. Thus, if we want estimate $I = \int f(x)\,dG(x)$ and we can only measure f at points which are given randomly with G distribution, we can replace $(U_{n,i}, f(U_{n,i}))$ in the estimator $\widehat{I}_n$ by $\big(G(U_{n,i}), f\big(G^{-1}(G(U_{n,i}))\big)\big) = (G(U_{n,i}), f(U_{n,i}))$.

§4. Final comments

These methods are related to the well-known Monte Carlo approximation. We have studied some alternative methods from those using random Riemann sums in Hernández and Urmeneta [2]. In this work we showed an unbiased estimator of $I = \int f dG$ which has less variance than that of Monte Carlo. Although the variance is lower it is not enough to improve the asymptotic behaviour of the error probability; however, simulated examples suggest that for certain functions f the Random Riemann Sum estimator can have much less error probability in the short-term.

Acknowledgements

We acknowledge the comments of an anonymous reviewer in the first version of this paper that helped to clarify the contents.

References

[1] EINMHAL, J., AND VAN ZUIJLEN, M.. On the aproximation of an integral by sum of random variables. *Journal of Applied Mathematics and Stochastics Analysis 11*, 2 (1998), 107–114.

[2] HERNÁNDEZ, V., AND URMENETA, H. Random Riemann sum estimator versus Monte Carlo. *Computational Statistics and Data Analysis.* (In second revisión, 2006)

[3] ISAACSON, E., AND KELLER, H. B. *Analysis of Numerical Methods*. Wiley, 1996.

[4] SHORACK, G. R., AND WELLNER, J. A. *Empirical Processes with Applications to Statistics*. Wiley, 1986.

Henar Urmeneta
Department of Statistics and Operations Research.
Public University of Navarra.
Campus Arrosadía s/n, 31006 Pamplona , Spain. henar@unavarra.es

Víctor Hernández
Department of Statistics and Operations Research.
National Distance University of Spain.
C/ Senda del Rey 9, 28040 Madrid, Spain. victorher@ccia.uned.es