Discussiones Mathematicae
Probability and Statistics 21 (2001) 81-88

STRONG LAW OF LARGE NUMBERS FOR ADDITIVE EXTREMUM ESTIMATORS

João Tiago Mexia and Pedro Corte Real

Universidade Nova de Lisboa, Departamento de Matemática
da Faculdade de Ciências e Tecnologia
Quinta da Torre, 2825–114 Monte de Caparica

e-mail: parcr@mail.fct.unl.pt

Abstract

Extremum estimators are obtained by maximizing or minimizing a function of the sample and of the parameters relatively to the parameters. When the function to maximize or minimize is the sum of subfunctions each depending on one observation, the extremum estimators are additive. Maximum likelihood estimators are extremum additive whenever the observations are independent. Another instance of additive extremum estimators are the least squares estimators for multiple regressions when the usual assumptions hold. A strong law of large numbers is derived for additive extremum estimators. This law requires only the existence of first order moments and may be of interest in connection with maximum likelihood estimators, since the usual assumption that the observations are identically distributed is discarded.

Keywords: Kolmogorov's strong law of large numbers, multiple regression, almost sure convergence, additive extremum estimators.

2000 Mathematics Subject Classification: 60F15, 62F10.

1. Introduction

Extremum estimators play nowadays a crucial role in statistical inference. These estimators are obtained by maximizing or minimizing a function of the sample; $\vec{y}^n$ and parameters relatively to the parameters. Thus maximum likelihood and least squares estimators are particular cases of extremum estimators. When the function we want to maximize or minimize is of the type $Q_n(\vec{\theta}^s|\vec{y}^n) = \sum_{i=1}^{n} h_i(\vec{\theta}^s|y_i)$ with $\{y_1, ..., y_n\}$ the observations, the extremum estimator we get is additive. Maximum likelihood estimators are additive whenever the observations are independent. Another instance of additive extremum estimators is given by least squares estimators for multiple regression. Under the usual assumptions these estimators are obtained minimizing

$$S_n(\vec{\beta}^k|\vec{y}^n) \;=\; \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{i,j}\beta_j\right)^2 = \sum_{i=1}^{n} h_i\left(\vec{\beta}^k|y_i\right).$$

Thus the class of additive extremum estimators contains sufficient important cases to justify its study. In what follows, we obtain a strong law of large numbers for such estimators under very mild assumptions. This law is an extension of the one we recently derived, see [4], for multiple regressions. Through the discussion of regularity conditions that hold for multiple regressions, we show that the extension we now present is a natural one, since these conditions ensure that the extended law holds too. Two final comments are as follows. First, we derive our results requiring only the existence of first order moments, both in the case of multiple regressions as in the case of additive extremum estimators. This is interesting since at least second order moments have been required, for instance see [1]and [2]. Second, our results may be of interest for maximum likelihood estimators, since the usual assumption, for instance see [6, page 233], that the observations are identically distributed is discarded.

2. Regularity conditions

We start by showing that certain results hold for multiple regression. These results will be used in formulating the conditions for the intended strong law that, thus, will be an extension of the one we derived for regressions.

The function $S_n(\vec{\beta}^k|\vec{y}^n)$ has the gradient and the hessian matrix given, respectively, by

$$\triangle S_n(\beta|\, y) = \triangle \sum_{i=1}^{n}\left(y_i - \sum_{j=1}^{k} x_{i,j}\beta_j\right)^2 \;= -2\sum_{i=1}^{n}\left(y_i - x_i^{'}\beta\right) x_i$$

$$HessS_n(\beta|\ y) = \sum_{i=1}^{n} Hess\left(y_i - \sum_{j=1}^{k} x_{i,j}\beta_j\right)^2 = 2\sum_{i=1}^{n} x_i x_i^{'} = 2X^{'}X$$

Moreover, with $\mu_{i0} = E[y_i] = x_i^{'}\beta_0$, where β_0 is the "true" coefficients vector and $i = 1, ..., n$, we have

$$E[\triangle h_i(\beta_0|y_i)] = \triangle h_i(\beta_0|\mu_{i0}) = 0$$

which is equivalent to

$$E[\triangle S_n(\beta_0|y)] = \triangle S_n(\beta_0|\mu_0) = 0.$$

From the previous results it is immediate that

$$E[\triangle S_{n+1}(\beta_0|\vec{y}^{n+1})] - E[\triangle S_n(\beta_0|\vec{y}^{n})] = E[\triangle h_{n+1}(\beta_0|\vec{y}^{n+1})] = 0$$

as well as,

$$\triangle S_{n+1}(\beta_0|\vec{\mu}_0^{n+1}) - \triangle S_n(\beta_0|\vec{\mu}_0^{n}) = \triangle h_{n+1}(\beta_0|\vec{\mu}_0^{n+1}) = 0$$

To establish our law for regressions we have assumed that $X^{'}X/n \to W$ with W a positive-definite matrix. So, we have, with $\rho(M)$ the spectral radius of the matrix M

$$\rho\left(\frac{1}{n}X^{'}X\right) \to \rho(W) > 0.$$

Let us rewrite $h_i(\beta|y_i)$ as $h_i(\beta, y_i)$ to include y_i in the variables. Then $\partial^2 h_i(\beta, y_i)/\partial\beta_j\partial y_i = -2x_{i,j}$. In [4] we have assumed that $|x_{i,j}| < a/2$, so that

$$\frac{\partial^2 h_i(\beta, y_i)}{\partial\beta_j\partial y_i} < a \quad with \quad i = 1, ..., n,\ j = 1, ..., k.$$

It is now easy to see that the following regularity conditions hold for multiple regressions:

$$y_i = \mu_{i0} + c_i e_i \tag{2.1}$$

with $|c_i| < c$, the e_i being $i.i.d.$ with null mean value, $i = 1, ..., n$.

$$E[\triangle h_i(\theta_0|y_i)] = \triangle h_i(\theta_0|\mu_{i0}) = 0, \quad i = 1, ..., n. \tag{2.2}$$

$$\rho\left(\frac{1}{n} Hess Q_n(\theta|y)\right) > b > 0 \tag{2.3}$$

where $\theta \in \Theta$ the parameter space, or at least in a subset Θ_0 of Θ where the true value θ_0 belongs.

$$\frac{\partial^2 h_i(\theta_0, y_i)}{\partial\theta_j \partial y_i} < a < \infty \text{ with } i = 1, ..., n, \ j = 1, ..., s. \tag{2.4}$$

These four conditions are the ones under which we derive our strong law of large numbers.

It is obvious, by the definition of $Q_n(\theta|y)$, that (2.2) is equivalent to:

$$E[\triangle Q_n(\theta_0|y)] = \triangle Q_n(\theta_0|\mu_0) = 0. \tag{2.5}$$

3. Convergence

Since extremum estimators $\tilde{\theta}_n$ are local extrema, they are solutions of systems of the form

$$\triangle Q_n(\theta|y) = 0$$

so that, we will have

$$0 = \triangle Q_n(\tilde{\theta}_n|y) = \triangle Q_n(\theta_0|y) + Hess Q_n(\hat{\theta}_n|y)(\tilde{\theta}_n - \theta_0)$$

with $\hat{\theta}_n = \lambda\theta_0 + (1 - \lambda)\tilde{\theta}_n$, where $\lambda \in [0, 1]$. When the hessian matrix is regular we have

$$\tilde{\theta}_n - \theta_0 = \left(\frac{1}{n} HessQ_n(\hat{\theta}_n|y)\right)^{-1} \left(-\frac{1}{n}\triangle Q_n(\theta_0|y)\right). \tag{3.6}$$

The components $q_{n,j}(\theta_0|y)$ of $\triangle Q_n(\theta_0|y)$, will be

$$q_{n,j}(\theta_0|y) = \sum_{i=1}^{n} \frac{\partial h_i(\theta_0|y_i)}{\partial\theta_j} = \sum_{i=1}^{n} \frac{\partial^2 h_i(\theta_0, v_i)}{\partial\theta_j \partial y_i}(y_i - \mu_{i0}), \quad j = 1, ..., s \tag{3.7}$$

with $||v_i - \mu_{i0}|| < ||y_i - \mu_{i0}||$, since according to (2.2), $\partial h_i(\theta_0|\mu_{i0})/\partial\theta_j = 0$, $i = 1, ..., n$, $j = 1, ..., s$.

Let us establish

Lemma 3.1. *If conditions* (2.1)–(2.4) *hold, with* $t(e_i) = e_i I_{|e_i|<i}$, $t(y_i) = y_i I_{|e_i|<i} + \mu_{i0} I_{|e_i|\geq i}$ *and* $z_i = v_i I_{|e_i|<i} + \mu_{i0} I_{|e_i|\geq i}$, $i = 1, ...n$. *Then,*

$$P\left[\liminf \frac{\partial^2 h_n(\theta_0, z_n)}{\partial\theta_j \partial y_n}(t(y_n) - \mu_{n0}) = \frac{\partial^2 h_n(\theta_0, v_n)}{\partial\theta_j \partial y_n}(y_n - \mu_{n0})\right] = 1 \tag{3.8}$$

$$E\left[\frac{\partial^2 h_n(\theta_0, z_n)}{\partial\theta_j \partial y_n}(t(y_n) - \mu_{n0})\right] = 0 \quad \textit{eventually} \tag{3.9}$$

$$\lim \sum_{i=1}^{n} \frac{1}{i^2} E\left[\left(\frac{\partial^2 h_i(\theta_0, z_i)}{\partial\theta_j \partial y_i}(t(y_i) - \mu_{i0})\right)^2\right] < \infty, \quad j = 1, ..., s. \tag{3.10}$$

Proof. First, we should note that $t(y_i) = \mu_{i0} + c_i e_i I_{|e_i|<i} = \mu_{i0} + c_i t(e_i)$.

Proof of (3.8). With H a random variable having the same distribution as the e_i, $i = 1, ..., n$.

$$\sum_{i=1}^{\infty} P\left[\frac{\partial^2 h_i(\theta_0, z_i)}{\partial\theta_j \partial y_i}(t(y_i) - \mu_{i0}) \neq \frac{\partial^2 h_i(\theta_0, v_i)}{\partial\theta_j \partial y_i}(y_i - \mu_{i0})\right] \leq \sum_{i=1}^{\infty} P[|e_i| > i]$$

$$= E\left[\sum_{i=1}^{\infty} I_{|H|>i}\right] = E\left[\sum_{1\leq i<|H|} 1\right] \leq E[|H|] < \infty.$$

Hence, the first part of the thesis follows from the First Borel-Cantelli lemma for $j = 1, ..., s$, see [5, page 27].

Proof of (3.9): The proof is immediate, since, by (3.8)

$$\frac{\partial^2 h_n(\theta_0, z_n)}{\partial\theta_j \partial y_n}(t(y_n) - \mu_{n0}) \overset{a.s.}{\to} \frac{\partial^2 h_n(\theta_0, v_n)}{\partial\theta_j \partial y_n}(y_n - \mu_{n0})$$

and by (3.7) we know that $E[(\partial^2 h_n(\theta_0, v_n)/\partial\theta_j \partial y_n)(y_n - \mu_{n0})] = 0$, hence

$$E\left[\frac{\partial^2 h_n(\theta_0, z_n)}{\partial\theta_j \partial y_n}(t(y_n) - \mu_{n0})\right] \to 0.$$

Proof of (3.10): Lastly, according to (2.4)

$$\left(\frac{\partial^2 h_i(\theta_0, z_i)}{\partial\theta_j \partial y_i}(t(y_i) - \mu_{i0})\right)^2 \leq a^2(t(y_i) - \mu_{i0}))^2$$

and by [5, page 119]

$$\sum_{i=1}^{\infty} \frac{1}{i^2} E\left[\left(\frac{\partial^2 h_i(\theta_0, z_i)}{\partial\theta_j \partial y_i}(t(y_i) - \mu_{i0})\right)^2\right] \leq a^2 \sum_{i=1}^{\infty} \frac{1}{i^2} E[(c_i t(e_i))^2]$$

$$\leq a^2 c^2 \sum_{i=1}^{\infty} \frac{1}{i^2} E[t(e_i)^2] \leq a^2 c^2 2E[|H|] < \infty.$$

We can now establish the *Strong Law of Large Numbers for Additive Extremum Estimators*

Proposition 3.2. *Under the stated conditions in Lemma* (3.1), $\tilde{\theta}_n \stackrel{a.s.}{\to} \theta_0$.

Proof. According to (3.6) and (2.3) we have only to show that $\triangle Q_n(\tilde{\theta}_n|y)/n \stackrel{a.s.}{\to} 0$, which, according to (3.7) and (3.8), is equivalent to

$$\lim \frac{1}{n} \sum_{i=1}^{n} \frac{\partial^2 h_i(\theta_0, z_i)}{\partial\theta_j \partial y_i}(t(y_i) - \mu_{i0}) \stackrel{a.s.}{\to} 0, \quad j = 1, ..., s.$$

Now, this last convergence follows immediately, from (3.8) and (3.9), due to the Kolmogorov's Strong Law of Large Numbers, for independent sequences of random variables, see, for instance, [3, page 57].

References

[1] N. Bac Van, *Strong convergence of least squares estimates in polynomial regression with random explanatory variables,* Acta Mathematica Vietnamica **23** (2) (1998), 195–205.

[2] N. Bac Van, *Strong convergence of least squares estimates in polynomial regression with random explanatory variables*, Acta Mathematica Vietnamica **19** (1) (1994), 111–137.

[3] J. Galambos, *Advanced Probability Theory*, Marcel Dekker 1988.

[4] J.T. Mexia and P.C. Real, *Extension of Kolmogorov's strong law to multiple regression*, 23rd European Meeting of Statisticians, Funchal (Madeira Island), August 13–18 2001, Revista de Estatística, 2 Quadrimestre de 2001, page 277–278.

[5] D. Williams, *Probability with Martingales*, Cambridge Mathematical Textbooks 1991.

[6] S. Zacks, *The Theory of Statistical Inference*, John Wiley 1971.

Received 10 June 2001
Revised 26 August 2001