

# A Note on the Uniform Closeness of Function Weighted Empirical Type Processes<sup>1</sup>

Mai Zhou

Department of Statistics, University of Kentucky  
Lexington, KY 40506-0027 USA

## Abstract

This article proves a lemma which shows that two empirical processes will be uniformly close if they are evaluated with a small time lag. A new feature of our lemma is that we allow the empirical processes to be functionally weighted (by  $f_i(t)$ , see Lemma 2). A third lemma for functionally weighted empirical process is also included.

*AMS 1991 Subject Classification:* Primary 62G10. Secondary 62P10, 62N05.

*Key Words and Phrases:* Weighted empirical process, regression.

## 1 Introduction

A lot of statistical estimators are defined as the maximizer or minimizer of a target function. Or, upon taking derivatives, the estimators are then defined as the root of certain equations. The explicit solution of the problem is often very difficult due to the complex nature of the target function. An approach that proven to be successful is to approximate uniformly in  $t$  the complicated function by some simple, usually linear or quadratic functions, at least locally in the neighborhood of the true parameter value. And then argue that the solution of the original equation can be approximated by the solution of the simpler equation, which is easier to solve. See e.g. Jureckova (1969) and Koul and Basawa (1984).

In this paper, we prove a lemma that is useful in those local uniform approximation. For more motivation and examples see Koul (198?), (1991). Lai and Ying (1990).

## 2 The Lemma

We prove a useful lemma on the uniform closeness of (functionally) weighted empirical distribution functions (lemma 2) in the case of non-identically distributed random variables. The following

---

<sup>1</sup>Abbreviated Title: *Closeness of Empirical processes.*

<sup>2</sup>Technical report #330, Department of Statistics, University of Kentucky.

lemma could be stated in triangular array with a subscript  $ni$  instead of  $i$ , but we choose not to use triangular array to avoid unnecessary notational complications. A typical case for the weighting constants below are  $|f_i| \sim \frac{1}{\sqrt{n}}$  and  $\xi_i = o(\frac{1}{\log n})$ .

**Lemma 1** *Let  $X_i, i = 1, \dots, n$  be independent random variables with  $P(X_i < t) = U_i(t)$  and  $f_i, \xi_i, i = 1, \dots, n$  be arbitrary constants. If the distributions  $U_i$  are uniformly Lipschitz:*

$$|U_i(x) - U_i(y)| \leq C|x - y|, \quad \forall i = 1, \dots, n \quad x, y \in \mathbb{R}^1 \quad (1)$$

then for those  $n$  and  $\epsilon$  such that  $C \sum_{i=1}^n f_i^2 |\xi_i| \leq \frac{\epsilon^2}{8}$  we have

$$P \left( \sup_t \left| \sum_{i=1}^n f_i \left[ I_{[X_i < t]} - U_i(t) - I_{[X_i < t + \xi_i]} + U_i(t + \xi_i) \right] \right| > \epsilon \right) \leq 16n \exp \left[ \frac{-\epsilon^2}{32(V + \epsilon/12 \max |f_i|)} \right] \quad (2)$$

where

$$V = \max_{1 \leq j \leq n} f_j^2 + C \sum_{i=1}^n f_i^2 |\xi_i|.$$

If furthermore

$$\max \left\{ \max_{1 \leq i \leq n} |f_i|, \sum_{i=1}^n f_i^2 |\xi_i| \right\} = o\left(\frac{1}{\log n}\right) \quad (3)$$

then

$$\sup_t \left| \sum_{i=1}^n f_i \left[ I_{[X_i < t]} - U_i(t) - I_{[X_i < t + \xi_i]} + U_i(t + \xi_i) \right] \right| = o(1) \quad a.s. \quad (4)$$

PROOF: We first use the symmetrization lemma 2.8 of Pollard (1984) to get rid of  $U_i$  in the probability in (2). The condition required by the symmetrization lemma,  $\sup_t P(|\cdot| > \epsilon/2) < 1/2$  is satisfied here since  $C \sum_{i=1}^n f_i^2 |\xi_i| \leq \frac{\epsilon^2}{8}$  and Chebychev inequality. Thus, we have

$$LHS \text{ of } (2) \leq 2P \left( \sup_t \left| \sum_{i=1}^n f_i \left[ I_{[X_i < t]} - I_{[X_i^* < t]} - I_{[X_i < t + \xi_i]} + I_{[X_i^* < t + \xi_i]} \right] \right| > \epsilon/2 \right)$$

where  $X_i^*$  are independent copies of  $X_i$ . Notice the  $U_i$ 's disappeared because of symmetrization.

Next introduce auxiliary random variables,  $\sigma_i$ , independent of the  $X_i$  and  $X_i^*$ , and taking value  $\pm 1$  with probability  $1/2$ . The above probability can then be bounded by

$$4P \left( \sup_t \left| \sum_{i=1}^n f_i \sigma_i \left[ I_{[X_i < t]} - I_{[X_i < t + \xi_i]} \right] \right| > \epsilon/4 \right).$$

For a fixed  $\omega$ , the supremum over  $t$  inside the probability is achieved at one of the  $2n$  points:  $X_i(\omega)$ 's plus  $X_i(\omega) - \xi_i$ 's. Therefore, the supremum over  $t$  is reduced to a maximum over  $2n$  points.

The probability of the maximum then is bounded by the sum:

$$4 \sum_{t=X_j} P \left( \left| \sum_{i=1}^n f_i \sigma_i [I_{[X_i < t]} - I_{[X_i < t + \xi_i]}] \right| > \epsilon/4 \right) + 4 \sum_{t=X_j - \xi_j} P \left( \left| \sum_{i=1}^n f_i \sigma_i [I_{[X_i < t]} - I_{[X_i < t + \xi_i]}] \right| > \epsilon/4 \right)$$

Now let us find a bound for the probability terms above. Notice the term  $f_i \sigma_i [I_{[X_i < X_j]} - I_{[X_i < X_j + \xi_i]}]$  has mean zero because of  $\sigma_i$ . And it has variance  $f_i^2 E(I_{[X_i < X_j]} - I_{[X_i < X_j + \xi_i]})^2$ . The expectation of the square is at most one when  $i = j$  but in all other cases ( $i \neq j$ ) is equal to  $U_i(X_j) - U_i(X_j + \xi_i)$  which by our Lipschitz assumption is bounded by  $C|\xi_i|$ . Similar mean and variance calculations hold for the terms in the second sum where  $t = X_j - \xi_j$ . Therefore we can bound the variance of the summation term inside the probability by

$$\sum_{i=1}^n f_i^2 E(I_{[X_i < X_j]} - I_{[X_i < X_j + \xi_i]})^2 \leq \max f_j^2 + \sum_{i=1}^n f_i^2 C|\xi_i| = V$$

By Bernstein's inequality (cf. Pollard, 1984, pp. 193), each of the probability term above is then bounded by

$$2 \exp \left[ -\epsilon^2/32 \left( V + \frac{1}{3} \frac{\epsilon}{4} \max |f_i| \right) \right]$$

Since the above bound is independent of  $i$ , (2) will follow if we apply the bound to every term in the sum.

If (3) hold, we see that

$$32(V + \frac{\epsilon}{12} \max |f_i|) = o(1/\log n)$$

This makes the bound (2) so small that it sum to a finite number (when  $\epsilon$  is fixed). By Borel-Cantelli lemma, it implies (4).  $\diamond$

When the weighting sequence  $f_i$  are themselves functions of  $t$ ,  $f_i(t)$ , we have similar results. We need to impose the bounded variation requirement on the  $f_i(t)$ . Denote, for an arbitrary function  $g(t)$ ,  $\|g(t)\| = \sup_t |g(t)|$ .

**Lemma 2** *Let  $X_i$ ,  $i = 1, \dots, n$  be independent random variables with  $P(X_i < t) = U_i(t)$  and  $\xi_i$   $i = 1, \dots, n$  be arbitrary constants. Suppose the distributions  $U_i$  are uniformly Lipschitz:*

$$|U_i(x) - U_i(y)| \leq C|x - y|, \quad \forall i = 1, \dots, n \quad x, y \in R^1. \quad (5)$$

*If  $f_i(t)$  are functions of bounded variation of  $t$  with total variation  $V_{-\infty}^{\infty} f_i(t) \leq K < \infty$ , with  $K$  independent of  $i$ , then for those  $\epsilon > 0$  and  $n$  such that  $\|C \sum_{i=1}^n f_i^2(t) |\xi_i|\| \leq \frac{\epsilon^2}{2}$  we have*

$$P \left( \left\| \sum_{i=1}^n f_i(t) [I_{[X_i < t]} - U_i(t) - I_{[X_i < t + \xi_i]} + U_i(t + \xi_i)] \right\| > \epsilon \right) \leq 16C_{\epsilon}(n) \exp \left( \frac{-\epsilon^2}{128(V + \epsilon/24 \max \|f_i\|)} \right) \quad (6)$$

where  $C_\epsilon(n) = \frac{16K}{\epsilon}n^2 + 2n$  and

$$V = \max_j \|f_j^2(t)\| + C \sum_{i=1}^n \|f_i^2(t)\| |\xi_i|.$$

If furthermore,

$$\max \left\{ \max_{1 \leq i \leq n} \|f_i(t)\|, \sum_{i=1}^n \|f_i^2(t)\| |\xi_i| \right\} = o\left(\frac{1}{\log n}\right) \quad (7)$$

then

$$\sup_t \left| \sum_{i=1}^n f_i(t) \left[ I_{[X_i < t]} - U_i(t) - I_{[X_i < t + \xi_i]} + U_i(t + \xi_i) \right] \right| = o(1) \quad a.s. \quad (8)$$

PROOF: The same symmetrization argument in the above proof of Lemma 1 still work here and leads to

$$P \left( \left\| \sum_{i=1}^n f_i(t) \left[ I_{[X_i < t]} - U_i(t) - I_{[X_i < t + \xi_i]} + U(t + \xi_i) \right] \right\| > \epsilon \right) \leq 4P \left( \left\| \sum_{i=1}^n f_i(t) \sigma_i \left[ I_{[X_i < t]} - I_{[X_i < t + \xi_i]} \right] \right\| > \frac{\epsilon}{4} \right). \quad (9)$$

Since the  $f_i$  are now functions of  $t$ , it is not enough to merely check the  $2n$  points for the supremum over  $t$  as before.

But we can check  $\frac{16K}{\epsilon}n^2$  additional points of  $t$  to make sure that we reach within  $\frac{\epsilon}{8}$  of the supremum. The reason is as follows. Since  $f_1(t)$  is of bounded variation, for any  $\epsilon > 0$ , we can choose no more than  $8\frac{2K}{\epsilon}n$  points on the line such that  $f_1(t)$  varies by no more than  $\frac{\epsilon}{8n}$  in any of the intervals between consecutive points. This can most easily be seen by writing  $f_1(t)$  as the difference of two increasing functions. Do the same thing with the other  $f_i(t)$  to get a total of no more than  $n \times \frac{16K}{\epsilon}n = \frac{16K}{\epsilon}n^2$  points on the real line. Within each of the two consecutive points, each and every  $f_i(t)$  do not vary by more than  $\frac{\epsilon}{8n}$  and thus  $\sum_{i=1}^n f_i(t)$  does not vary by more than  $\sum_{i=1}^n \frac{\epsilon}{8n} = \frac{\epsilon}{8}$  within any interval. Therefore for each fixed  $\omega$ , by checking those points in addition to the  $2n$  points as in lemma one, we will definitely find a maximum that is at most  $\epsilon/8$  shy of the supremum.

Thus

$$4P \left( \sup_t \left| \sum_{i=1}^n f_i(t) \sigma_i \left[ I_{[X_i < t]} - I_{[X_i < t + \xi_i]} \right] \right| > \frac{\epsilon}{4} \right) \leq 4P \left( \max_{t_j} \left| \sum_{i=1}^n f_i(t_j) \sigma_i \left[ I_{[X_i < t_j]} - I_{[X_i < t_j + \xi_i]} \right] \right| > \frac{\epsilon}{8} \right) \leq 4 \sum_{t=t_j} P \left( \left| \sum_{i=1}^n f_i(t_j) \sigma_i \left[ I_{[X_i < t_j]} - I_{[X_i < t_j + \xi_i]} \right] \right| > \frac{\epsilon}{8} \right) \quad (10)$$

where  $2n$  of the points  $t_j$  are random as in Lemma 1 ( $X_j(\omega)$  and  $X_j(\omega) - \xi_j$ 's) while the rest of the points are nonrandom, depending only on the function  $f_i$ . The variance bound  $V$  is still valid for those nonrandom choice of  $t$ 's. Same application of the Bernstein's inequality (cf. Pollard, 1984, pp. 193) will finish the proof.  $\diamond$

If we further suppose that  $V + \max \|f_i\| = O(1/\sqrt{n})$  then the result of (4) can be strengthened to give an order. For instance, (4) =  $o(\log n/n^{1/4})$  a.s..

The following lemma deals with the case where weights are functions but involves only one empirical process (no time lag). It can be proved similarly to Lemma 2. We used this Lemma in the study of the stratified Cox model with number of strata go to infinity.

**Lemma 3** *Let  $X_i, i = 1, \dots, n$  be independent random variables and  $f_i(t), i = 1, \dots, n$  be nonrandom functions of bounded variation. Assume the total variation for each  $f_i(t)$  is bounded by  $K$ :  $V_{-\infty}^{\infty} f_i(t) \leq K$ .*

*Then for those  $n$  and  $\epsilon > 0$  such that  $\|\sum_{i=1}^n f_i^2(t)\| \leq \epsilon^2/2$ , we have*

$$P\left(\left\|\sum f_i(t)[I_{[X_i < t]} - P(X_i < t)]\right\| > \epsilon\right) \leq 16C_\epsilon(n) \exp\left[-\frac{\epsilon^2}{128(V_n + \epsilon \max_i \|f_i(t)\|/24)}\right]$$

where  $C_\epsilon(n) = \frac{16K}{\epsilon}n^2 + 2n$  and  $V_n = \sum_{i=1}^n \|f_i^2(t)\|$ .

*Furthermore if  $\max\{V_n, \max \|f_i(t)\|\} = o(1/\log n)$  then*

$$\sup_t \left| \sum_{i=1}^n f_i(t)[I_{[X_i < t]} - P(X_i < t)] \right| = o(1) \text{ a.s.}$$

A typical case where this lemma is useful is when  $\|f_i(t)\| = O(n^{-s})$  for  $s > 1/2$ .

## References

- [1] JURECKOVA, J. (1969). Asymptotic linearity of a rank statistic in regression parameter. *Ann. Math. Statist.* **40**, 1889-1900.
- [2] KOUL, H. AND BASAWA, (1984).
- [3] KOUL, H. (1991). A weak convergence result useful in robust autoregression. *J. Statist. Planning and Inf.* **29** 291-308.
- [4] LAI, T. L. & YING, Z. (1988). Stochastic integrals of empirical-type processes with applications to censored regression. *J. Mult. Analysis* **27**, 334-358.
- [5] POLLARD, D. (1984). *Convergence of Stochastic Processes*. Springer, New York.
- [6] SHORACK, G. R. AND WELLNER, J. (1986). *Empirical Processes with Applications to Statistics*. Wiley, New York.
- [7] SRINIVASAN, C. & ZHOU, M. (1992). Minimum distance regression estimators for censored survival data. University of Kentucky Technical report.