

Probability Theory

A functional Hungarian construction for the sequential empirical process

Michael Jähnisch^a, Michael Nussbaum^{b,1}

^a *SAP AG, Neurottstrasse 16, 69190 Walldorf, Germany*

^b *Department of Mathematics, Cornell University, Ithaca, NY 14853, USA*

Received 11 October 2005; accepted 18 October 2005

Presented by Michel Talagrand

Abstract

We establish a KMT coupling for the sequential empirical process and the Kiefer–Müller process. The processes are indexed by functions f from a Hölder class \mathcal{H} , but the supremum over $f \in \mathcal{H}$ is taken outside the probability. Compared to the coupling in sup-norm, this allows for larger functional classes \mathcal{H} . The result is useful for proving asymptotic equivalence of certain nonparametric statistical experiments. **To cite this article:** *M. Jähnisch, M. Nussbaum, C. R. Acad. Sci. Paris, Ser. I 341 (2005).*

© 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Résumé

Une construction hongroise fonctionnelle pour le processus empirique séquentiel. Nous établissons un couplage KMT pour le processus empirique séquentiel et le processus de Kiefer–Müller. Les processus sont indexés par des fonctions f appartenant à une classe de Hölder \mathcal{H} , mais le supremum est pris en dehors de la probabilité. Comparé au couplage en norme sup, ceci permet des classes de fonctions \mathcal{H} plus larges. Le résultat peut être utilisé pour démontrer l'équivalence asymptotique de certaines expériences statistiques non-paramétriques. **Pour citer cet article :** *M. Jähnisch, M. Nussbaum, C. R. Acad. Sci. Paris, Ser. I 341 (2005).*

© 2005 Académie des sciences. Published by Elsevier SAS. All rights reserved.

Let x_1, x_2, \dots be a sequence of independent random variables being uniformly distributed on the unit interval. The process

$$\widehat{G}_n(s, t) := \frac{1}{\sqrt{n}} \sum_{i=1}^{[ns]} (1_{[0,t]}(x_i) - t), \quad (s, t) \in [0, 1]^2$$

is called the sequential empirical process. It is well known that an invariance principle holds for this process. Let K be the Kiefer–Müller process on the unit square, defined as

$$K(s, t) = W(s, t) - tW(s, 1), \quad (s, t) \in [0, 1]^2,$$

E-mail addresses: michael.jaehnisch@sap.com (M. Jähnisch), nussbaum@math.cornell.edu (M. Nussbaum).

¹ Research supported by the National Science Foundation under grant DMS0306497.

where W is the Brownian sheet, i.e. the continuous centered Gaussian process with covariance function

$$EW(s_1, t_1)W(s_2, t_2) = (s_1 \wedge s_2)(t_1 \wedge t_2).$$

Then we have convergence in distribution

$$\widehat{G}_n \Longrightarrow K \quad \text{as } n \rightarrow \infty$$

(cp. [13], Theorem 1, p. 131).

According to the principle ‘nearby variables for nearby laws’ (cf. [3], Section 11.6), one can expect a strong coupling result to hold for these processes. Indeed the following result is due to Komlós, Major and Tusnády (cf. [10], Theorem 4, p. 114). For every $n \in \mathbb{N}$, there is a probability space on which there exist versions of the processes \widehat{G}_n and K such that for all $x \geq 0$ we have:

$$P\left(n^{1/2} \sup_{s \in \{i/n, i=1, \dots, n\}, t \in [0,1]} |\widehat{G}_n(s, t) - K(s, t)| \geq (C \log n + x) \log n\right) \leq L \exp(-\lambda x),$$

where C, L and λ are positive absolute constants.

A local refinement along with a specification of the constants has been obtained by Castelle and Laurent-Bonvalot [2]. These authors also present a complete proof of the above theorem based on a quantile inequality for hypergeometric distributions. This result is an analog of the quantile inequality for the symmetric binomial distribution known as Tusnády’s lemma which was proved in detail by Bretagnolle and Massart [1].

Similarly to [2], our starting point in this note is the quantile inequality for the hypergeometric distribution but we are aiming at a different kind of coupling result. Firstly, we consider functional versions of the processes, i.e. for functions $f \in \mathcal{H}$, where \mathcal{H} is the class of real valued functions on the unit square, let:

$$\widehat{G}_n(f) := \frac{1}{\sqrt{n}} \sum_{i=1}^n \left(f(i/n, x_i) - \int_0^1 f(i/n, v) dv \right). \quad (1)$$

Furthermore, let $B_i, i = 1, \dots, n$, be a set of independent standard Brownian bridges and define

$$K_n(f) := \sum_{i=1}^n \int_0^1 f(i/n, v) dB_i(v).$$

Since the Kiefer–Müller process $K(s, t)$ with the first argument restricted to $\{i/n, i = 1, \dots, n\}$ can be represented as

$$K(s, t) = n^{-1/2} \sum_{i=1}^{ns} B_i(t), \quad s \in \{i/n, i = 1, \dots, n\}, t \in [0, 1],$$

it is clear that the process $n^{-1/2} K_n(f), f \in \mathcal{H}$ is a discretized functional version of K . We will call $K_n(f)$ the discretized Kiefer–Müller process, bearing in mind that rigorously the term is appropriate for $n^{-1/2} K_n(f)$. The process $\widehat{G}_n(f)$ indexed by $f \in \mathcal{H}$ will be called the sequential empirical process.

Secondly, the supremum will not be taken inside the probability, as in the classical KMT result (cited above) and refinements by Castelle and Laurent-Bonvalot [2], but it will be taken outside, with the same type of exponential upper bound. A detailed discussion of the statistical motivation for such a result can be found in [4]; it will also be briefly touched upon below.

Let $\mathcal{H}(1, L)$ be the class of functions on the unit square having Lipschitz norm less or equal to L and being uniformly bounded. More precisely

$$\mathcal{H}(1, L) = \{f : [0, 1]^2 \mapsto \mathbb{R} : |f(x) - f(y)| \leq L \|x - y\|, |f(x)| \leq L, x, y \in [0, 1]^2\}.$$

Theorem 1. *For every $n \geq 2$, there is a probability space on which there exist versions of the sequential empirical process and the discretized Kiefer–Müller process, such that for every $L > 0$ and all $x \geq 0$:*

$$\sup_{f \in \mathcal{H}(1, L)} P(|\widehat{G}_n(f) - n^{-1/2} K_n(f)| \geq n^{-1/2} \log^7(n) x) \leq C_1 \exp(-C_2 x),$$

where C_1 and C_2 are positive constants, depending only on L .

The proof is in the thesis [7]. This result has an application in the theory of statistical experiments, where asymptotic equivalence is understood in the sense of Le Cam's Δ -distance. It is preferable for such applications that the supremum be taken outside the probability. Indeed, consider two sequences of experiments E_n and G_n given by families of probability measures P_f^n and Q_f^n defined possibly on different sample spaces but indexed both by $f \in \Sigma$. Assume that for some $f_0 \in \Sigma$, all measures P_f^n are absolutely continuous with respect to $P_{f_0}^n$, and the same for Q_f^n and $Q_{f_0}^n$, respectively. If there exist versions $d\tilde{P}_f^n/d\tilde{P}_{f_0}^n$ and $d\tilde{Q}_f^n/d\tilde{Q}_{f_0}^n$ of the likelihood processes of the experiments on some common probability space $(\Omega_n, \mathcal{F}_n, \mathbb{P}_n)$, then one can estimate the Δ -distance between the experiments as follows:

$$\Delta^2(E_n, G_n) \leq \sqrt{2} \sup_{f \in \Sigma} E_{\mathbb{P}_n} \left(\sqrt{d\tilde{P}_f^n/d\tilde{P}_{f_0}^n} - \sqrt{d\tilde{Q}_f^n/d\tilde{Q}_{f_0}^n} \right)^2.$$

More details can be found in [11]. The above theorem is used to obtain such a coupling for an experiment given by independent non-identically distributed observations and an accompanying Gaussian one. The corresponding result on asymptotic equivalence of experiments can be found in the companion paper to the present note [8] and the thesis [7].

The type of coupling result discussed here was first studied by Koltchinskii [9]. This author considered functional versions of the empirical process and the Brownian bridge and established an analog of the above theorem for classes of functions defined on the unit interval. This result was instrumental for proving asymptotic equivalence of experiments in [11]. Our theorem can be seen as an extension where the uniform empirical process is replaced by the uniform sequential empirical process and functions are defined on the unit square. In [9] and [12] the coupling was then carried on to a supremum over functional classes inside the probability, in the spirit of the original KMT result, but this development remains outside the scope of the present paper. We note only that the function class $\mathcal{H}(1, L)$ on the unit square is well known not to be Donsker, and hence there is no coupling (or sequence of couplings) in which the term

$$\sup_{f \in \mathcal{H}(1, L)} |\hat{G}_n(f) - n^{-1/2} K_n(f)|$$

tends to zero in probability.

As to the type of coupling established in the theorem above, we note that a related result has been obtained by Grama and Nussbaum [4]. These authors consider the partial sum process indexed by functions h on the unit interval

$$\hat{S}_n(h) := \frac{1}{\sqrt{n}} \sum_{i=1}^n h(i/n) y_i,$$

where y_i are independent centered random variables. Note that setting $f(s, t) = h(s)t$ in (1) we obtain the partial sum process for centered uniforms $y_i = x_i - 1/2$ but the paper [4] covers more general y_i . The corresponding results on asymptotic equivalence of experiments can be found in [5] and [6].

References

- [1] J. Bretagnolle, P. Massart, Hungarian constructions from the non-asymptotic viewpoint, *Ann. Probab.* 17 (1989) 239–256.
- [2] N. Castelle, F. Laurent-Bonvalot, Strong approximation of bivariate uniform empirical processes, *Ann. Inst. H. Poincaré Probab. Statist.* 34 (1998) 425–480.
- [3] R. Dudley, *Real Analysis and Probability*, Wadsworth Brooks/Cole, Pacific Grove, CA, 1989.
- [4] I. Grama, M. Nussbaum, A functional Hungarian construction for sums of independent random variables, *Ann. Inst. H. Poincaré Probab. Statist.* 38 (2002) 923–957.
- [5] I. Grama, M. Nussbaum, Asymptotic equivalence for nonparametric generalized linear models, *Probab. Theory Related Fields* 111 (1998) 167–214.
- [6] I. Grama, M. Nussbaum, Asymptotic equivalence for nonparametric regression, *Math. Methods Statist.* 11 (1) (2002) 1–36.
- [7] M. Jähnisch, *Asymptotische Äquivalenz für ein Modell unabhängiger nicht identisch verteilter Daten*, Doctoral Thesis, Humboldt Universität zu Berlin, 1999.
- [8] M. Jähnisch, M. Nussbaum, Asymptotic equivalence for a model of independent non identically distributed observations, *Statistics and Decisions* 21 (3) (2003) 197–218.
- [9] V. Koltchinskii, Komlós–Major–Tusnády approximation for the general empirical process and Haar expansions of classes of functions, *J. Theoret. Probab.* 7 (1) (1994) 73–118.
- [10] J. Komlós, P. Major, G. Tusnády, An approximation of partial sums of independent r.v.'s and the sample df. I, *Z. Wahr. Verw. Gebiete* 32 (1975) 111–131.
- [11] M. Nussbaum, Asymptotic equivalence of density estimation and Gaussian white noise, *Ann. Statist.* 24 (1996) 2399–2430.
- [12] E. Rio, Local invariance principles and their application to density estimation, *Probab. Theory Related Fields* 98 (1994) 21–45.
- [13] G. Shorack, J. Wellner, *Empirical Processes with Applications to Statistics*, Wiley, New York, 1986.