# A functional Hungarian construction for the sequential empirical process

Michael Jaehnisch and Michael Nussbaum[1]

*SAP AG and Cornell University*

**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 avoids the loss of approximation rate which occurs for large functional classes $\mathcal{H}$. The result is useful for proving asymptotic equivalence of certain nonparametric statistical experiments.

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

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

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

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

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

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

(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 Komlos, Major and Tusnady, (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:

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

---

[1]Supported in part by NSF Grant DMS-03-06497

where $C$, $L$ and $\lambda$ 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 presented 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 Tusnady'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:

$$\widehat{G}_n(f) := \frac{1}{\sqrt{n}} \sum_{i=1}^{n} \left( f(i/n, x_i) - \int_0^1 f(i/n, v) dv \right). \tag{0.1}$$

Furthermore, let $B_i, i = 1, \ldots, n$ 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 $\{i/n, i = 1, \ldots, n\}$ can be represented as

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

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 that 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

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

**Theorem 1** *For every $n \ge 2$, 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 $x \ge 0$:*

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

*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 $\Delta$-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 $d\widetilde{P}_f^n/d\widetilde{P}_{f_0}^n$ and $d\widetilde{Q}_f^n/d\widetilde{Q}_{f_0}^n$ 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 $\Delta$-distance between the experiments as follows:

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

More details can be found in [11]. Theorem 1 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 that is 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 Theorem 1 for classes of functions defined on the unit interval. This result was instrumental for proving asymptotic equivalence of experiments in [11]. Theorem 1 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] 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. Note that in this framework, for a general function class $\mathcal{H}$, the rate in probability of the term

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

will no longer be almost $n^{-1/2}$ as for each individual $f$ in Theorem 1, but will be a slower rate depending on the entropy of $\mathcal{H}$. This is evident from [9] and also from analogous results in [12]. Furthermore, it should be noted that the Hölder exponent 1 on the unit square figuring in Theorem 1 is an analog of the well-known smoothness bound 1/2 for results of this type on the unit interval.

Another coupling result of the kind established here was obtained by Grama and Nussbaum [4]. These authors consider the partial sum process indexed by functions $h$ on the unit interval

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

where $y_i$ are independent centered random variables. Note that setting $f(s,t) = h(s)t$ in (0.1) 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].

Theorem 1 is based on a coupling for the sequential empirical and the discretized Kiefer-Müller process for finite sets of functions. It is in fact Theorem 2 below which is used in [7] for the computation of the $\Delta$-distance and which is comparable to Theorem 3.5 of Koltchinskii [9]. This author also first assumed a finite set of functions as index set and then developed an extension to larger classes.

For $f \in L^2([0,1],\lambda)$, we consider a smoothness measure $R_M^2(f)$ related to the $L^2$-modulus of continuity on the unit interval and the unit square; for details cf. [7]. Analogously to Theorem 3.5 in [9] the result below for a finite set set of functions $\mathcal{F}$ involves both the cardinality of $\mathcal{F}$ and the smoothness of functions in $\mathcal{F}$.

**Theorem 2** *There is a probability space and for all* $n \in \mathbb{N}$ *there exist versions of the processes* $\widehat{G}_n$ *and* $K_n$ *on that space such that for all* $x \geq 0$, $y \geq 0$ *and* $\mathcal{F} \subset L^2([0,1]^2, \lambda^2)$ *where* $\|f\|_\infty \leq 1$ *for all* $f \in \mathcal{F}$ *such that* $\#\mathcal{F} < \infty$ *holds, we have:*

$$P(n^{\frac{1}{2}} \|\widehat{G}_n(f) - n^{-\frac{1}{2}} K_n(f)\|_{\mathcal{F}} \geq (\log n)^2 (Ax + Bx^{\frac{1}{2}}(y^{\frac{1}{2}} + C)(\log n)^3 R_M(\mathcal{F})))$$
$$\leq D[\#\mathcal{F} \exp(-Gx) + n \exp(-Gy)]$$

*(A,B,C,D and G are positive, absolute constants and* $R_M(\mathcal{F}) = \|R_M(f)\|_{\mathcal{F}}$ *and* $\|.\|_{\mathcal{F}} = \max_{f \in \mathcal{F}} |.|$*).*

## References


[1] J. Bretagnolle and P. Massart, Hungarian constructions from the non-asymptotic viewpoint. *Ann. Probab.* **17** (1989) 239-256.

[2] N. Castelle and 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 and 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 and M. Nussbaum, Asymptotic equivalence for nonparametric generalized linear models. *Prob. Theor. Rel. Fields* **111** (1998) 167-214.

[6] I. Grama and M. Nussbaum, Asymptotic equivalence for nonparametric regression. *Math. Meth. 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 and M. Nussbaum, Asymptotic equivalence for a model of independent non identically distributed observations. *Statistics and Decisions* **21** (3) 2003 197-218.

[9] V. Koltchinskii, Komlos-Major-Tusnady approximation for the general empirical process and Haar expansions of classes of functions. *J. Theoretical Probability* **7** (1) (194) 73-118.

[10] J. Komlos, P. Major and G. Tusnady, An approximation of partial sums of independent r.v.'s and the sample df. I *Z. Wahrscheinlichkeitstheorie 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. *Prob. Theor. Rel. Fields* **98** (1994) 21-45.

[13] G. Shorack and J. Wellner, *Empirical Processes with Applications to Statistics.* Wiley, New York (1986).