ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
SCHOOL OF ECONOMIC SCIENCES
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF ECONOMICS

Department of Economics

Athens University of Economics and Business

WORKING PAPER no. 05-2022

Concentration Inequalities for Kernel Density Estimators under Uniform Mixing

Stelios Arvanitis

March 2022

Πατησίων 76, 104 34 Αθήνα. Τηλ.: 210 8203303-5. E-mail: econ@aueb.gr / www.aueb.gr
76, Patission Street, Athens 104 34 Greece. Tel.: (+30) 210 8203303-5

Concentration Inequalities for Kernel Density Estimators under Uniform Mixing

Stelios Arvanitis*

March 14, 2022

Abstract

We derive non-asymptotic concentration inequalities for the uniform deviation between a multivariate density function and its non-parametric kernel density estimator in stationary and uniform mixing time series framework. We derive analogous inequalities for their (first) Wasserstein distance, as well as for the deviations between integrals of bounded functions w.r.t. them. They can be used for the construction of confidence regions, the estimation of the finite sample probabilities of decision errors, etc. We employ the concentration results to the derivation of statistical guarantees and oracle inequalities in regularized prediction problems with expected costs exhibiting Lipschitz and strong convexity properties.

Keywords: Non-asymptotic concentration inequalities, Kernel Estimator, Density Estimation, Uniform Mixing, Statistical Gaurantees, Lipschitz costs, Strong Convexity, Support Vector Machines, Hinge Costs.

*Arvanitis is Associate Professor at the Department of Economics of Athens University of Economics and Business; Athens; Greece; email: stelios@aueb.gr. The hospitality of the Laboratory of Econometrics and the Laboratory of Economic Policy Studies (EMOP) of the department is gratefully aknowledged.

1 Introduction

In the present note we are occupied with the derivation of non-asymptotic concentration inequalities for the uniform deviation between a multivariate density function and its non-parametric kernel density estimator over the support of the former. We employ a time series setting consisting of multivariate stationary processes with summable uniform (phi-) mixing coefficients (see Davidson [2]). We rely heavily on the iid results of Vogel and Schettler [11], adjusting their proofs via the use of convenient Hoeffding-type and covariance inequalities for uniformly mixing processes. Regarding the two underlying probability measures, we readily derive analogous probability bounds for their (first) Wasserstein distance, as well as for the deviations between integrals of bounded functions w.r.t. them. Those hold uniformly w.r.t. the sample size. The inequalities can be used for the construction of confidence regions, the estimation of the finite sample probabilities of false inclusion of parameter values in level sets of moment conditions, etc. Extension of the results to other forms of temporal dependence, like absolute regularity or strong mixing, and subsequently, application to an empirically relevant wider range of time series models is delegated to further research.

As an example, we employ the concentration results to the derivation of statistical guarantees and oracle inequalities in regularized prediction problems with Lipschitz and strongly convex costs over function spaces. The inequalities imply a uniform non-asymptotic LLN for the deviation between the empirical (w.r.t. the kernel estimator) and the population cost differences. This coupled with the relevant sub-differential calculus in convex programming, implies a non asymptotic statistical guarantee for the L_2 deviation between the empirical and the population solutions as long as the regularization parameter is appropriately dominated. The framework is quite general and allows for dynamic parameter spaces and population solutions. It includes as special cases non-linear Support Vector Machines with Hinge costs.

For the remaining note: in Section 2 we derive and discuss the concentration inequalities, and in Section 3 we are occupied with the aforementioned regularized prediction problems. Section 4 contains the proofs.

2 Concentration Inequalities for Kernel Density Estimators

The researcher has at her disposal the time series sample of random vectors $(x_t)_{t=1,\ldots,T}$, and estimates f_x, the unknown density of x_0, via the kernel estimator $\frac{1}{Tb_T^n}\sum_{t=1}^{T}\mathcal{K}\left(\frac{x_t-\cdot}{b_T}\right)$ with $b_T > 0$ the bandwidth. We consider the problem of bounding, uniformly in T, the probability that the uniform deviation $\sup_{y\in\mathcal{X}}\left|\frac{1}{Tb_T^n}\sum_{t=1}^{T}\mathcal{K}\left(\frac{x_t-y}{b_T}\right)-f_x(y)\right|$ exceeds an asymptotically negligible deterministic sequence.

We begin the analysis with an assumption that specifies our probabilistic and statistical framework. The assumption imposes restrictions on the marginal distributions and the dynamics of the stochastic process involved in the density estimation. It also restricts the properties of the kernel technology employed:

Assumption 1. *(i) The $\mathbb{R}^n$-valued stochastic process $(x_t)_{t\in\mathbb{Z}}$ is strictly stationary and phi-mixing, with absolutely square root summable mixing coefficient sequence $(\phi_n)_{n\in\mathbb{N}}$. (ii) $\mathcal{K}$ is a positive symmetric bounded, Lipschitz continuous and compactly supported convolution kernel on $\mathbb{R}^n$ such that*
$\int_{\mathbb{R}^n}\mathcal{K}(u)\,du = 1$, $\int_{\mathbb{R}^n}\|u\|^2\mathcal{K}(u)\,du < +\infty$, and (iii) the distribution of x_0 has a compact support $\mathcal{X}$, and a continuous density $f_x(\cdot)$ that is twice differentiable with continuous second derivatives.

An example of a dynamic multivariate process that satisfies Assumption 1.(i) is given by the solution of the following stochastic recursion equation (SRE):

$$x_t = h(x_{t-1}) + z_t,$$

where $(z_t)_{t\in\mathbb{Z}}$ is an iid sequence of n-random vectors, the distribution of z_0 has a density, and

$h:\mathbb{R}^n\to\mathbb{R}^n$ is a contraction (w.r.t. some metric on $\mathbb{R}^n$) with compact range. Theorem 2.1.3 of Doukhan and Ghindes [3] implies then the required mixing property for the unique solution of the SRE. This incorporates the iid as a special case. If z_0 has also bounded support the compactness of the support of the distribution of x_0 from the first part of Assumption 1.(i) also holds. The remaining parts impose mostly usual conditions in non-parametric statistics (see El Machkouri, Fan, and Reding [7], and references therein). Boundedness of supports can be generalized as long as $\mathcal{K}$ has an integrable Fourier transform, and the Hessian of f_x is bounded in the Frobenius norm.

The following theorem summarizes the result. There, $\mathcal{W}\left(G,\,G^{\star}\right)$ denotes the first Wasserstein distance between the arbitrary distributions on $\mathcal{X}$, $G,\,G^{\star}$, defined as $\min_{\gamma\in\Gamma(G,\,G^{\star})}\int_{\mathcal{X}\times\mathcal{X}}d\left(z,z^{\star}\right)d\gamma\left(z,z^{\star}\right)$, where $\Gamma\left(G,\,G^{\star}\right)$ denotes the set of Borel probability distributions on $\mathcal{X}\times\mathcal{X}$ that have respective marginals $G,\,G^{\star}$, and d denotes the Euclidean distance (see Gao, Chen, and Kleywegt [4]). μ denotes the Lebesgue measure on $\mathbb{R}^n$ and diam $(\mathcal{X})$ denotes the Euclidean diameter of $\mathcal{X}$.

Theorem 1 (Concentration Inequalities)**.** *Suppose that Assumption 1 holds:*

A. *Uniformly in $T\geq 1$ and for any $k>0$,*

$$\mathbb{P}\left(\ \sup_{y\in\mathcal{X}}\left|\frac{1}{Tb_T^n}\sum_{t=1}^{T}\mathcal{K}\left(\frac{x_t-y}{b_T}\right)-f_x\left(y\right)\right|>\beta_{T,k}\ \right)\leq 2\exp\left(-\frac{k^2}{2C_3\left(1+2\sum_{n=1}^{\infty}\phi_n\right)^2}\right),\tag{1}$$

where, $\beta_{T,k}:=\frac{k}{\sqrt{T}b_T^n}+\frac{C_1}{(2\pi)^n\sqrt{T}\sqrt{1+2\sum_{n=1}^{\infty}\sqrt{\phi_n}b_T^n}}+\frac{1}{2}C_2b_T^2$, $C_1:=\int_{\mathbb{R}^n}\left|\int_{supp(\mathcal{K})}\mathcal{K}\left(u\right)\exp\left(iy^Tu\right)du\right|dy$, $C_2:=\sup_{i,j,\,y\in\mathcal{X}}\left|\frac{\partial^2 f_x(y)}{\partial x_i\partial x_j}\right|\int_{supp(\mathcal{K})}\left\|u\right\|^2\mathcal{K}\left(u\right)du$, and $C_3:=\sup_{u\in supp(\mathcal{K})}\mathcal{K}^2\left(u\right)$.

B. *Let F_T denote the cdf of $\frac{1}{Tb_T^n}\sum_{t=1}^{T}\mathcal{K}\left(\frac{x_t-\cdot}{b_T}\right)$, and analogously F the cdf of $f_x\left(\cdot\right)$. Then, uniformly in $T\geq 1$, and for any $k>0$,*

$$\mathbb{P}\left(\mathcal{W}\left(F_T,F\right)>\textit{diam}\left(\mathcal{X}\right)\mu\left(\mathcal{X}\right)\beta_{T,k}\right)\leq 2\exp\left(-\frac{k^2}{2C_3\left(1+2\sum_{n=1}^{\infty}\phi_n\right)^2}\right).\tag{2}$$

C. *Suppose that* $(\mathcal{F}, \|\cdot\|_{\mathcal{F}})$ *is an* R*-bounded subset of a semi-normed space of real functions defined on* $\mathcal{X}$*, and there exists some* $c^{\star} > 0$ *for which* $\|\cdot\|_{\infty} \leq c^{\star} \|\cdot\|_{\mathcal{F}}$*. Then, uniformly in* $T \geq 1$*, and for any* $k > 0$*,*

$$\mathbb{P}\left(\left|\mathbb{E}_{F_T}(f(y)) - \mathbb{E}_F(f(y))\right| > c^{\star} R\mu(\mathcal{X})\beta_{T,k}\right) \leq 2\exp\left(-\frac{k^2}{2C_3\left(1+2\sum_{n=1}^{\infty}\phi_n\right)^2}\right). \quad (3)$$

The results in (1)-(3) are non asymptotic. Whenever $b_T = o(1)$ while $\sqrt{T}b_T^n \to +\infty$, they can also provide estimates for the rates of convergence of the deviations considered. The derivation of (1) essentially follows from the proof of (the iid based) Theorem 1 of Vogel and Schettler [11], by taking into account Hoeffding type inequalities for functions of bounded differences (see Rio [8]), and covariance inequalities (see Davidson [2]), for uniform mixing processes. Then, (2)-(3) follow from the dual functional representation of $\mathcal{W}$ of Kantorovich [6], the uniform boundedness of the function space involved in **C**, and the compactness of $\mathcal{X}$. The bounding sequence $\beta_{T,k}$ depends on the integral of the Fourier transform of the kernel, the second moment of the kernel, the magnitude of the Hessian of f_x, the mixing coefficients, the sample size, and the bandwidth. In the last two cases, $\beta_{T,k}$ is complemented by the Lebesgue magnitude and the diameter of $\mathcal{X}$, and/or the uniform bound and the norm properties of the function space considered. The probability bound depends on the bound of the kernel and the mixing coefficients. It becomes tighter whenever $\mathcal{K}$ admits a low maximum, and/or the mixing coefficients are small and converge rapidly to zero.

A and **C** can be used among others in order to construct non-asymptotic conservative confidence sets. For example, (3) implies the of at least $1 - 2\exp\left(-\frac{k^2}{2C_3\left(1+2\sum_{n=1}^{\infty}\phi_n\right)^2}\right)$ probability, confidence interval $\left(\mathbb{E}_{F_T}(f(y)) \mp c^{\star}R\lambda(\mathcal{X})\beta_{T,k}\right)$ for $\mathbb{E}_F(f(y))$, uniformly on $\mathcal{F}$. The estimation of the supremum of the Hessian of C_2 could be facilitated by the derivation of the kernel estimator for the estimation of the derivatives of f_x-see for example Sheather [10]. The estimation of the mixing coefficient series can be facilitated by the results in Ahsen and Vidyasagar [1] and truncation. Alternatively, and given further assumptions on the rate of convergence of the mixing coefficients, the aforementioned series can be replaced by upper bounds at the cost of further conservatism. **B** can be used in order to bound from above the probability of including $\theta^\star \in \left\{\sup_{\mathcal{W}(F_T,G)\leq\lambda_T} \mathbb{E}_G\left(g\left(\theta^T y\right)\right) \leq 0\right\}$, while $\theta^\star \notin \left\{\mathbb{E}_F\left(g\left(\theta^T y\right)\right) \leq 0\right\}$, for $g : \mathbb{R} \to \mathbb{R}$, and $\theta^\star \in \Theta \subseteq \mathbb{R}^n$. If g is 1-Lipschitz, Θ is bounded in the Euclidean norm, and $\lambda_T \geq \text{diam}(\Theta)\text{diam}\left(\mathcal{X}\right)\mu\left(\mathcal{X}\right)\beta_{T,k}$, then the probability of falsely classifying $\theta^\star$ in the zero level set of $\mathbb{E}_F\left(g\left(\theta^T y\right)\right)$, via the use of the conservative statistical program $\sup_{\mathcal{W}(F_T,G)\leq\lambda_T} \mathbb{E}_G\left(g\left(\theta^T y\right)\right)$, is bounded above by the rhs of (2).

3 Regularized Prediction Problems with Lipschitz costs

In what follows $\left(\mathcal{F}, \|\cdot\|_{\mathcal{F}}\right)$ conforms to the function space in Theorem 1.**C**, $\mathcal{L}$ is a loss function on $\mathbb{R}^2$, and given the sample $(x_t)_{t\in\{1,\cdots,T\}}$ with $x_t := (y_t, \mathbf{X}_t)$ with y_t denoting the response variables, and $\mathbf{X}_t$ the predictors, and we consider the regularized prediction (conditional on z) empirical program:

$$f_T := \arg\min_{f\in\mathcal{F}} \mathbb{E}_{F_T}\left[\mathcal{L}\left(f\left(x\right), y\right)\right] + \lambda_T \|f\|_{\mathcal{F}}\,, \tag{4}$$

with $\lambda_T > 0$ a regularization parameter.

We employ the concentration inequalities of the previous section in order to obtain statistical guarantees for the L_2 distance between f_T, and the solution to the population analogue of (4): $\inf_{f\in\mathcal{F}} \mathbb{E}_F\left[\mathcal{L}\left(f\left(x\right), y\right)\right]$. This is summarized in the following result.

Theorem 2 (Statistical Guarantees). *Suppose that Assumption 1 and the conditions of* Theorem 1.***C*** *hold. Suppose furthermore that (SG.i) $\mathcal{F}$ is convex and uniformly R-bounded, and (SG.ii) for* *some $L, \kappa > 0$, uniformly in the second argument, $\mathcal{L}(\cdot,\cdot)$ is L-Lipschitz and $\mathbb{E}_F[\mathcal{L}]$ is κ-strongly convex. Let $f^\star$ be the unique solution of the population statistical program* $\inf_{\mathcal{F}} \mathbb{E}_F[\mathcal{L}(f(x), y)]$ *and suppose that it lies in the interior of $\mathcal{F}$. Then, and if $\beta_{T,k} > \frac{\lambda_T}{2c^\star \mu(\mathcal{X}) L}$, the following statistical guarantees hold:*

$$\kappa \|f_T - f^\star\|_2 \leq 4c^\star \mu(\mathcal{X}) LR\sqrt{\beta_{T,k}} + \sqrt{2R}\sqrt{4c^\star \mu(\mathcal{X}) L(\kappa + L)\beta_{T,k} + \kappa\lambda_T}, \tag{5}$$

with probability greater than or equal to $1 - 2\exp\left(-\frac{k^2}{2C_3\left(1+2\sum_{n=1}^{\infty}\phi_n\right)^2}\right)$.

The parameter space convexity and (small sample) boundedness in SG.i and the Lipschitz continuity property in SG.ii are not rare statistical applications. Strong convexity of the population criterion depends crucially on F and holds whenever $\mathbb{E}_F[\mathcal{L}]$ is convex and two times Frechet differentiable with second order derivative that has a bounded away from zero spectrum uniformly in $\mathcal{F}$. The statistical guarantees in (5) hold for any T. They allow for diverging R with $T \to \infty$, hence for cases where the parameter space $\mathcal{F}$ becomes asymptotically unbounded. They also allow for the population solution $f^\star$ to depend on T as well as on the strong convexity parameter κ to become asymptotically nullified. If $b_T = o(1)$ and $\sqrt{T} b_T^n \to +\infty$, they imply that $\|f_T - f^\star\|_2$ becomes asymptotically negligible w.h.p. as long as $\lambda_T < 2c^\star \mu(\mathcal{X}) L\beta_{T,k}$ for some $k \to \infty$, and $\frac{R}{\kappa}\left(b_T + \sqrt{\frac{k}{\sqrt{T} b_T^n}}\right) = o(1)$.

An example that adheres to the formulation above, is the one of Support Vector Machines with Hinge Costs (see Example 14.19 of Wainwright [12]). $\mathcal{F}$ is typically the R-ball of a Reproducing Kernel Hilbert Space comprised by discriminant real functions and centered at zero, and $\mathcal{L}(f(x), y) := (1 - yf(x))_+$. The latter is clearly 1-Lipschitz in its first argument, while κ-strong convexity holds as long as $\mathbb{E}_F[y^2\delta(1 - yf(x))]$, with δ denoting the Dirac delta function, is bounded away from zero uniformly on $\mathcal{F}$.

For a more general to (5) result, if (4) is substituted with $\min_{f \in \mathcal{G}_T} \mathbb{E}_{F_T}[\mathcal{L}(f(x), y)] + \lambda_T \|f\|_{\mathcal{F}}$, with convex $\mathcal{G}_T \subseteq \mathcal{F}$, and such that $g_T^\star := \arg\min_{f\in\mathcal{G}_T} \|f - f^\star\|_2$, and the sub-differential $\partial\mathbb{E}_F[\mathcal{L}(\cdot, y)]$ is L_∂-Lipschitz uniformly in y (see for example Ch. 9 of Rockafellar and Wetts [9]), then the following oracle inequality is similarly obtained (see the proof of Theorem 2):

$$\kappa \|f_T - f^\star\|_2 \leq \begin{array}{c} 4c^\star \mu(\mathcal{X}) LR\sqrt{\beta_{T,k}} + (1 + L + L_\partial)\|g_T^\star - f^\star\|_2 \\ + \sqrt{\begin{array}{c} \left((L + L_\partial)^2 \|g_T^\star - f^\star\|_2 + 8L^2 c^\star \mu(\mathcal{X}) R\sqrt{\beta_{T,k}}\right)\|g_T^\star - f^\star\|_2 \\ +2\left(4c^\star \mu(\mathcal{X}) L(\kappa + L)\beta_{T,k} + \kappa\lambda_T\right) R \end{array}} \end{array}, \quad (6)$$

whenever $\lambda_T < 2c^\star \mu(\mathcal{X}) L\beta_{T,k}$ holds, with probability greater than or equal to the probability bound in Theorem (2). This bounces back to (5) when $f^\star = g_T^\star$.

4 Proofs

Proof of Theorem 1. Consider (1). Due to the Hoeffding type inequality for phi-mixing processes (see Rio [8]), and working exactly as in the proof of Theorem 1 of Vogel and Schettler [11] (2013) we obtain that $\mathbb{P}\left(|J_T - \mathbb{E}(J_T)| > t \right) \leq 2\exp\left(-\frac{t^2 T b_T^{2n}}{2C_3\left(1+2\sum_{n=1}^\infty \phi_n\right)^2}\right)$, for any $t \geq 0$, where $J_T := \sup_{x\in\mathcal{X}} \left|\frac{1}{Tb_T^n}\sum_{t=1}^T \mathcal{K}\left(\frac{x_t - x}{b_T}\right) - f_x(x)\right|$. Working as in the proof of the first Lemma of Vogel and Settler (2013), and noting that due to the phi-mixing covariance inequality (see Corollary 14.5 of Davidson [2]), $\frac{1}{T^2}\text{Var}\left(\sum_{t=1}^T \exp\left(\mathrm{i}u^\intercal x_t\right)\right) \leq \frac{1}{T}\left(1 + 2\sum_{n=1}^\infty \sqrt{\phi_n}\right)$, we obtain that $\mathbb{E}\left(\sup_{x\in\mathcal{X}} \left|\frac{1}{Tb_T^n}\sum_{t=1}^T \mathcal{K}\left(\frac{x_t-x}{b_T}\right) - \mathbb{E}\left(\frac{1}{Tb_T^n}\sum_{t=1}^T \mathcal{K}\left(\frac{x_t-x}{b_T}\right)\right)\right|\right) \leq \frac{C_1}{(2\pi)^n\sqrt{T}\sqrt{1+2\sum_{n=1}^\infty\sqrt{\phi_n}}\, b_T^n}$. Finally, due to the second Lemma of Vogel and Schettler [11], we obtain the inequality $\sup_{x\in\mathcal{X}} \left|\mathbb{E}\left(\frac{1}{Tb_T^n}\sum_{t=1}^T \mathcal{K}\left(\frac{x_t-x}{b_T}\right)\right) - f_x(x)\right| \leq \frac{1}{2}C_2 b_T^2$. The result follows by choosing $t := \frac{k}{\sqrt{T}b_T^n}$ in the probability inequality above. (2) follows from Theorem 4 of Gibbs and Su [5], the compactness of $\mathcal{X}$ and (1). Analogously, (3) follows from the uniform boundedness of $\mathcal{F}$, the dominance of $\|\cdot\|_{\mathcal{F}}$, the compactness of $\mathcal{X}$ and (1). □

Proof of Theorem 2. Set $R^\star := 2c^\star \mu(\mathcal{X}) R$. We have first that for any $f \in \mathcal{F}$, due to the Lipschitz properties of $\mathcal{L}$, the $\|\cdot\|_{\mathcal{F}}$-boundedness of $\mathcal{F}$-which implies that $\mathcal{F} - f^{\star}$ is $2R$ bounded, and (3),

$$
\begin{aligned}
\mathbb{P}\left[\frac{\left|\mathbb{E}_{F_T}[\mathcal{L}(f(x),y)]-\mathbb{E}_{F_T}[\mathcal{L}(f^{\star}(x),y)]-\mathbb{E}_F[\mathcal{L}(f(x),y)]+\mathbb{E}_F[\mathcal{L}(f^{\star}(x),y)]\right|}{\|f-f^{\star}\|_2+\sqrt{\beta_{T,k}}} \geq 2LR^{\star}\sqrt{\beta_{T,k}}\right] \\
\leq \mathbb{P}\left[\frac{\left|\mathbb{E}_{F_T}[f(x)-f^{\star}(x)]-\mathbb{E}_F[f(x)-f^{\star}(x)]\right|}{\|f-f^{\star}\|_2+\sqrt{\beta_{T,k}}} \geq R^{\star}\sqrt{\beta_{T,k}}\right] \\
2\,\text{exp}\left(-\frac{k^2}{2C_3\left(1+2\sum_{n=1}^{\infty}\phi_n\right)^2}\right).
\end{aligned}
$$

Let us prove (6). Then (5) follows by assuming $f^{\star} = g_T^{\star}$, and noting that since $f^{\star}$ is interior, the normal cone of $\mathcal{F}$ at $f^{\star}$ is $\{0\}$, and the optimality of $f^{\star}$ implies that $\langle \partial \mathbb{E}_F\left[\mathcal{L}\left(f^{\star}\left(x\right), y\right)\right], f_T - f^{\star}\rangle_2 \in \{0\}$, where $\partial \mathbb{E}_F\left[\mathcal{L}\left(f^{\star}\left(x\right), y\right)\right]$ denotes the sub-differential of the population criterion at $f^{\star}$, and $\langle\cdot,\cdot\rangle_2$ the L_2 inner product. In this case the Lipschitz property of the sub-differential is redundant. Towards proving (6), remember that $g_T^{\star} := \text{arg min}_{f\in\mathcal{G}_T}\|f - f^{\star}\|_2$. Then consider the event

$$
\left\{\frac{\left|\mathbb{E}_{F_T}\left[\mathcal{L}\left(f\left(x\right), y\right)\right] - \mathbb{E}_{F_T}\left[\mathcal{L}\left(g_T^{\star}\left(x\right), y\right)\right] - \mathbb{E}_F\left[\mathcal{L}\left(f\left(x\right), y\right)\right] + \mathbb{E}_F\left[\mathcal{L}\left(g_T^{\star}\left(x\right), y\right)\right]\right|}{\|f - f^{\star}\|_2 + \sqrt{\beta_{T,k}^{\star}}} \leq 2LR^{\star}\sqrt{\beta_{T,k}}\right\},
$$

and notice that the probability of the above is bounded below by $1-2\,\text{exp}\left(-\frac{k^2}{2C_3\left(1+2\sum_{n=1}^{\infty}\phi_n\right)^2}\right)$, due to the previous result (which did not use the fact that $f^{\star}$ is interior). If the event holds then, and due to that $\mathbb{E}_{F_T}\left[\mathcal{L}\left(f\left(x\right), y\right)\right] - \mathbb{E}_{F_T}\left[\mathcal{L}\left(g_T^{\star}\left(x\right), y\right)\right] + \lambda_T\left(\|f_T\|_{\mathcal{F}} - \|g_T^{\star}\|_{\mathcal{F}}\right) \leq 0$, then it must be the case that

$$
\mathbb{E}_F\left[\mathcal{L}\left(f_T\left(x\right), y\right)\right] + \mathbb{E}_F\left[\mathcal{L}\left(g_T^{\star}\left(x\right), y\right)\right] + \lambda_T\left(\|f_T\|_{\mathcal{F}} - \|g_T^{\star}\|_{\mathcal{F}}\right) \leq 2LR^{\star}\sqrt{\beta_{T,k}}\left(\|f_T - g_T^{\star}\|_2 + \sqrt{\beta_{T,k}^{\star}}\right).
$$

The κ-strong convexity of the population criterion then implies that

$$
\begin{aligned}
\langle \partial \mathbb{E}_F\left[\mathcal{L}\left(g_T^{\star}\left(x\right), y\right)\right], f_T - g_T^{\star}\rangle_2 + \tfrac{\kappa}{2}\|f_T - g_T^{\star}\|_2^2 + \lambda_T\left(\|f_T\|_{\mathcal{F}} - \|g_T^{\star}\|_{\mathcal{F}}\right) \\
\leq 2LR^{\star}\sqrt{\beta_{T,k}}\left(\|f_T - g_T^{\star}\|_2 + \sqrt{\beta_{T,k}}\right).
\end{aligned}
$$

Now, notice that due to the optimality of $g_T^{\star}$, $\partial \mathbb{E}_F\left[\mathcal{L}\left(g_T^{\star}\left(x\right), y\right)\right]$ must lie inside the normal cone of $\mathcal{G}_T$ at $g_T^\star$. This and the fact that f_T satisfies the empirical optimality conditions imply that

$$\langle \partial \mathbb{E}_F \left[\mathcal{L} \left(g_T^\star \left(x \right), y \right) \right], f_T - g_T^\star \rangle_2 \leq 0.$$

The lhs of the previous display is greater than or equal to

$$\langle \partial \mathbb{E}_F \left[\mathcal{L} \left(g_T^\star \left(x \right), y \right) \right] - \partial \mathbb{E}_F \left[\mathcal{L} \left(f^\star \left(x \right), y \right) \right], f_T - g_T^\star \rangle_2 + \langle \partial \mathbb{E}_F \left[\mathcal{L} \left(f^\star \left(x \right), y \right) \right], f_T - g_T^\star \rangle_2$$

which (due to the sub-differential inclusion condition) is greater than or equal to

$$\begin{gathered} \langle \partial \mathbb{E}_F \left[\mathcal{L} \left(g_T^\star \left(x \right), y \right) \right] - \partial \mathbb{E}_F \left[\mathcal{L} \left(f^\star \left(x \right), y \right) \right], f_T - g_T^\star \rangle_2 \\ + \langle \partial \mathbb{E}_F \left[\mathcal{L} \left(f^\star \left(x \right), y \right) \right], f_T - g_T^\star \rangle_2 + \langle \partial \mathbb{E}_F \left[\mathcal{L} \left(f^\star \left(x \right), y \right) \right], f^\star - g_T^\star \rangle_2, \end{gathered}$$

which in turn, due to Cauchy-Schwarz inequality and the Lipschitz property of the sub-differential, is greater than or equal to

$$-L_\partial \left\| g_T^\star - f^\star \right\|_2 \left\| f_T - g_T^\star \right\|_2 - L \left(\left\| f_T - g_T^\star \right\|_2 + \left\| f^\star - g_T^\star \right\|_2 \right).$$

The previous then imply that

$$\begin{gathered} -L_\partial \left\| g_T^\star - f^\star \right\|_2 \left\| f_T - g_T^\star \right\|_2 - L \left(\left\| f_T - g_T^\star \right\|_2 + \left\| f^\star - g_T^\star \right\|_2 \right) + \tfrac{\kappa}{2} \left\| f_T - g_T^\star \right\|_2^2 + \lambda_T \left(\left\| f_T \right\|_{\mathcal{F}} - \left\| g_T^\star \right\|_{\mathcal{F}} \right) \\ \leq 2LR^\star \sqrt{\beta_{T,k}} \left(\left\| f_T - g_T^\star \right\|_2 + \sqrt{\beta_{T,k}} \right) \Rightarrow \\ \tfrac{\kappa}{2} \left\| f_T - g^\star \right\|_2^2 - \left(L \left(2R^\star \sqrt{\beta_{T,k}} + \left\| f^\star - g_T^\star \right\|_2 \right) + L_\partial \left\| g_T^\star - f^\star \right\|_2 \right) \left\| f_T - g_T^\star \right\|_2 \\ + \left(\lambda_T \left(\left\| f_T \right\|_{\mathcal{F}} - \left\| g_T^\star \right\|_{\mathcal{F}} \right) - 2LR^\star \beta_{T,k} \right) \leq 0. \end{gathered}$$

The condition that ensures that the quadratic polynomial in the lhs of the previous display has two distinct roots, one negative and one positive, is $\beta_{T,k} > \frac{\lambda_T}{2c^\star \mu(\mathcal{X}) L}$, since this and the fact that $\left\| f_T \right\|_{\mathcal{F}} - \left\| g_T^\star \right\|_{\mathcal{F}} \leq 2R$, imply that $\lambda_T \left(\left\| f_T \right\|_{\mathcal{F}} - \left\| g_T^\star \right\|_{\mathcal{F}} \right) - 2LR^\star \beta_{T,k} < 0$. Comparing with the positive root we obtain that

$$\begin{aligned}
\kappa \left\| f_T - g_T^\star \right\|_2 &\le L \left(2R^\star \sqrt{\beta_{T,k}} + \left\| f^\star - g_T^\star \right\|_2 \right) + L_\partial \left\| g_T^\star - f^\star \right\|_2 \\
&+ \sqrt{4LR^\star \left(\kappa + L \right) \beta_{T,k} + \left(L + L_\partial \right)^2 \left\| g_T^\star - f^\star \right\|_2^2 + 4L^2 R \sqrt{\beta_{T,k}^\star} \left\| f^\star - g_T^\star \right\|_2 - 2\kappa\lambda_T \left(\left\| f_T \right\|_{\mathcal{F}} - \left\| g_T^\star \right\|_{\mathcal{F}} \right)} \\
&\le 2LR^\star \sqrt{\beta_{T,k}} + \left(L + L_\partial \right) \left\| g_T^\star - f^\star \right\|_2 \\
&+ \sqrt{4LR^\star \left(\kappa + L \right) \beta_{T,k} + \left(L + L_\partial \right)^2 \left\| g_T^\star - f^\star \right\|_2^2 + 4L^2 R^\star \sqrt{\beta_{T,k}} \left\| f^\star - g_T^\star \right\|_2 + 2\kappa\lambda_T \left\| g_T^\star \right\|_{\mathcal{F}}}
\end{aligned},$$

from which the oracle inequality (6) follows by noting that $\left\| g_T^\star \right\|_{\mathcal{F}} \le R$. □

References

[1] M. E. Ahsen and Vidyasagar M., 2013, "On the computation of mixing coefficients between discrete-valued random variables," 9th Asian Control Conference (ASCC), pp. 1-5, doi: 10.1109/ASCC.2013.6606096.

[2] Davidson, J., 1994, Stochastic limit theory: An introduction for econometricians, OUP Oxford.

[3] Doukhan, P., and Ghindès, M., 1983, Estimation de la transition de probabilité d'une chaîne de Markov Doëblin-récurrente. Étude du cas du processus autorégressif général d'ordre 1, *Stochastic processes and their applications*, 15(3), 271-293.

[4] Gao, R., Chen, X. and Kleywegt, A.J., 2017, Distributional robustness and regularization in statistical learning, *arXiv preprint arXiv:1712.06050*.

[5] Gibbs, A.L. and Su, F.E., 2002, On choosing and bounding probability metrics. International statistical review, 70(3), pp.419-435.

[6] Kantorovich, L.V., 1960, Mathematical methods of organizing and planning production, Management science, 6(4), pp.366-422.

[7] El Machkouri, M., X. Fan, and L. Reding, 2020, On the Nadaraya–Watson kernel regression

estimator for irregularly spaced spatial data, Journal of Statistical Planning and Inference, 205, 92-114.

[8] Rio, E., 2000, Inégalités de Hoeffding pour les fonctions lipschitziennes de suites dépendantes, Comptes Rendus de l'Académie des Sciences-Series I-Mathematics, 330(10), pp.905-908.

[9] Rockafellar, R.T. and Wets, R.J.B., 2009, Variational analysis (Vol. 317), Springer Science & Business Media.

[10] Sheather, S.J., 2004, Density estimation, Statistical science, pp.588-597.

[11] Vogel, S. and Schettler, A., 2013, A uniform concentration-of-measure inequality for multivariate kernel density estimators. Techn. Univ., Inst. für Mathematik.

[12] Wainwright, M.J., 2019, High-dimensional statistics: A non-asymptotic viewpoint (Vol. 48), Cambridge University Press.

ΟΙΚΟΝΟΜΙΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ
ATHENS UNIVERSITY OF ECONOMICS AND BUSINESS
ΣΧΟΛΗ ΟΙΚΟΝΟΜΙΚΩΝ ΕΠΙΣΤΗΜΩΝ
SCHOOL OF ECONOMIC SCIENCES
ΤΜΗΜΑ ΟΙΚΟΝΟΜΙΚΗΣ ΕΠΙΣΤΗΜΗΣ
DEPARTMENT OF ECONOMICS

Department of Economics
Athens University of Economics and Business

List of Recent Working Papers

2020

01-20 Incubated Early Stage Startuppers in Athens and their Initiatives during the Crisis (2010-2016), Ioannis Besis and Ioanna Sapfo Pepelasis
02-20 Alternative Monetary Policy Rules in an Imperfectly Competitive Dynamic Stochastic General Equilibrium Model, George Alogoskoufis and Stelios Giannoulakis
03-20 Quantifying Qualitative Survey Data: New Insights on the (Ir)Rationality of Firms' Forecasts, Alexandros Botsis, Christoph Görtz and Plutarchos Sakellaris
04-20 Can Country-Specific Interest Rate Factors Explain the Forward Premium Anomaly?, Efthymios Argyropoulos, Nikolaos Elias, Dimitris Smyrnakis and Elias Tzavalis
05-20 Signaling product quality through prices in an oligopoly model with costly consumer search, Minghua Cheny, Konstantinos Serfes and Eleftherios Zacharias
06-20 Thinking ahead of the next big crash: Clues from Athens in classical times, George C. Bitros
07-20 Financial crises, firm-level shocks, and large downturns: Evidence from Greece, Stelios Giannoulakis and Plutarchos Sakellaris
08-20 Notes on the Demand Side Secular Stagnation, George D. Demopoulos and Nicholas A. Yannacopoulos
09-20 Institutions and Macroeconomic Performance: Core vs Periphery Countries in the Eurozone, Tryfonas Christou, Apostolis Philippopoulos and Vanghelis Vassilatos
10-20 Macroeconomic Policy Lessons for Greece, George Economides, Dimitris Papageorgiou and Apostolis Philippopoulos
11-20 Energy Transitions and the role of the EU ETS: The case of Greece, Andriana Vlachou and Georgios Pantelias
12-20 Measuring the Default Risk of Small Business Loans: Improved Credit Risk Prediction using Deep Learning, Yiannis Dendramis, Elias Tzavalis and Aikaterini Cheimarioti
14-20 Saving Democracy by Switching to Corporate-like Governance, George C. Bitros
15-20 The impact of the lockdown on the Greek economy and the role of the Recovery Fund, George Economides, Apostolis Philippopoulos and Vanghelis Vassilatos

2021

01-21 Historical Cycles of the Economy of Modern Greece From 1821 to the Present, George Alogoskoufis
02-21 Greece Before and After the Euro: Macroeconomics, Politics and the Quest for Reforms, George Alogoskoufis
03-21 Commodity money and the price level, George C. Bitros

04-21 **Destabilizing asymmetries in central banking: With some enlightenment from money in classical Athens, George C. Bitros**
05-21 **Exploring the Long-Term Impact of Maximum Markup Deregulation, Athanasios Dimas and Christos Genakos**
06-21 **A regularization approach for estimation and variable selection in high dimensional regression models, Y. Dendramis, L. Giraitis, G. Kapetanios**
07-21 **Tax Competition in the Presence of Environmental Spillovers, Fabio Antoniou, Panos Hatzipanayotou, Michael S. Michael, Nikos Tsakiris**
08-21 **Firm Dynamics by Age and Size Classes and the Choice of Size Measure, Stelios Giannoulakis and Plutarchos Sakellaris**
09-21 **Measuring the Systemic Importance of Banks, Georgios Moratis, Plutarchos Sakellaris**
10-21 **Firms' Financing Dynamics Around Lumpy Capacity Adjustments, Christoph Görtz, Plutarchos Sakellaris, John D. Tsoukalas**
11-21 **On the provision of excludable public goods General taxes or user prices? George Economides and Apostolis Philippopoulos**
12-21 **Asymmetries of Financial Openness in an Optimal Growth Model, George Alogoskoufis**
13-21 **Evaluating the impact of labour market reforms in Greece during 2010-2018 , Georgios Gatopoulos, Alexandros Louka, Ioannis Polycarpou, Nikolaos Vettas**
14-21 **From the Athenian silver to the bitcoin standard: Private money in a state-enforced free banking model, George C. Bitros**
15-21 **Ordering Arbitrage Portfolios and Finding Arbitrage Opportunities. Stelios Arvanitis and Thierry Post**
16-21 **Inconsistency for the Gaussian QMLE in GARCH-type models with infinite variance, Stelios Arvanitis and Alexandros Louka**
17-21 **Competition and Pass-Through: Evidence from Isolated Markets, Christos Genakos and Mario Pagliero**
18-21 **Exploring Okun's Law Asymmetry: An Endogenous Threshold LSTR Approach, Dimitris Christopoulos, Peter McAdam and Elias Tzavalis**
19-21 **Limit Theory for Martingale Transforms with Heavy-Tailed Multiplicative Noise, Stelios Arvanitis and Alexandros Louka**
20-21 **Optimal taxation with positional considerations, Ourania Karakosta and Eleftherios Zacharias**
21-21 **The ECB's policy, the Recovery Fund and the importance of trust: The case of Greece, Vasiliki Dimakopoulou, George Economides and Apostolis Philippopoulos**

2022

01-22 **Is Ireland the most intangible intensive economy in Europe? A growth accounting perspective, Ilias Kostarakos, KieranMcQuinn and Petros Varthalitis**
02-22 **Common bank supervision and profitability convergence in the EU, Ioanna Avgeri, Yiannis Dendramis and Helen Louri**
03-22 **Missing Values in Panel Data Unit Root Tests, Yiannis Karavias, Elias Tzavalis and Haotian Zhang**
04-22 **Ordering Arbitrage Portfolios and Finding Arbitrage Opportunities, Stelios Arvanitis and Thierry Post**

Department of Economics
Athens University of Economics and Business

The Department is the oldest Department of Economics in Greece with a pioneering role in organising postgraduate studies in Economics since 1978. Its priority has always been to bring together highly qualified academics and top quality students. Faculty members specialize in a wide range of topics in economics, with teaching and research experience in world-class universities and publications in top academic journals.

The Department constantly strives to maintain its high level of research and teaching standards. It covers a wide range of economic studies in micro-and macroeconomic analysis, banking and finance, public and monetary economics, international and rural economics, labour economics, industrial organization and strategy, economics of the environment and natural resources, economic history and relevant quantitative tools of mathematics, statistics and econometrics.

Its undergraduate program attracts high quality students who, after successful completion of their studies, have excellent prospects for employment in the private and public sector, including areas such as business, banking, finance and advisory services. Also, graduates of the program have solid foundations in economics and related tools and are regularly admitted to top graduate programs internationally. Three specializations are offered:1. Economic Theory and Policy, 2. Business Economics and Finance and 3. International and European Economics. The postgraduate programs of the Department (M.Sc and Ph.D) are highly regarded and attract a large number of quality candidates every year.

For more information:

https://www.dept.aueb.gr/en/econ/