Tatra Mt. Math. Publ. **17** (1998), 219–226

TATRA MOUNTAINS
Mathematical Publications

GAUSSIAN APPROXIMATION FOR NONPARAMETRIC MODELS

Ion G. Grama

ABSTRACT. We show that the nonparametric model generated by independent observations X_i, $i = 1, \dots, n$, with densities $p(x, f(i/n))$, $i = 1, \dots, n$, where f is an "unknown" function $f \colon [0,1] \to \Theta$ in a smoothness class, can be approximated, in the sense of the Le Cam deficiency pseudodistance, by a nonparametric Gaussian shift model.

1. Introduction

Following Brown and Low [1], we say that two sequences of statistical experiments $\mathcal{E}^n$, $n = 1, 2, \dots$, and $\mathcal{G}^n$, $n = 1, 2, \dots$, are *asymptotically equivalent* if

$$\Delta(\mathcal{E}^n, \mathcal{G}^n) \to 0, \qquad \text{as} \quad n \to \infty,$$

where $\Delta(\mathcal{E}^n, \mathcal{G}^n)$ is the Le Cam deficiency pseudo-distance between statistical experiments $\mathcal{E}^n$ and $\mathcal{G}^n$. From the definition of the Le Cam distance (see Le Cam [5] or Le Cam and Yang [6]) it follows that asymptotically minimax risks over bounded loss functions in one sequence of models can be transferred to the asymptotically equivalent sequence of models, which means that one can compute the asymptotically minimax risk in non-Gaussian models by computing it in the accompanying Gaussian models. Thus asymptotic equivalence can be seen as an analog of the Hájek – Le Cam asymptotic minimax bound in the case of nonparametric models.

Such an asymptotic equivalence between the white noise model and its discrete time analog has been first established by Brown and Low [1]. For the density estimation, Nussbaum [7] found that the accompanying experiment is a white noise model, where the square root of the density is estimated. The more general case of generalized linear models was considered in Grama and

2000 Mathematics Subject Classification: Primary 62B15; Secondary 62G07, 62G20.
Keywords: asymptotic equivalence, deficiency distance, Gaussian approximation, nonparametric model.

The author was supported by the Alexander von Humboldt-Stiftung.

Nussbaum [3]. However, in all these models it is assumed, essentially, that the noises are from an exponential family. The purpose of the present work is to extend the area of applicability of the asymptotic equivalence to the case of regression models with arbitrary noises.

2. The results

2.1 The model. We start with a parametric model. Let Θ be an interval (possibly infinite) in the real line and

$$\mathcal{E} = \big(X, \mathscr{X}, \{P_\theta \colon \theta \in \Theta\}\big)$$

be a statistical experiment on the measurable space $(X, \mathscr{X})$ with the parameter set Θ and a dominating σ-finite measure μ. For any $\theta \in \Theta$, denote by

$$p(x, \theta) = P_\theta(dx)/\mu(dx)\,, \qquad x \in X\,,$$

the Radon-Nikodym derivative of P_θ w.r.t. μ. For the sake of brevity we set $p(\theta) = p(\cdot, \theta)$. We shall assume in the sequel that $p(\theta) > 0$, μ-a.s. on X, for any $\theta \in \Theta$, which implies that the measures P_θ, $\theta \in \Theta$ are equivalent: $P_\theta \sim P_u$, for $\theta,\ u \in \Theta$.

Now we shall introduce our nonparametric model. For any $\beta > 0$, let Σ^β be the Hölder ball of functions defined on $[0,1]$ and with values in Θ, i.e., the set of functions $f\colon [0,1] \to \Theta$, which satisfy Hölder's condition with exponent β and constant L:

$$\big|f^{(\beta_0)}(t) - f^{(\beta_0)}(s)\big| \le L|t-s|^{\beta_1}\,, \qquad t, s \in [0,1] \quad \text{and} \quad \sup_{t \in T} |f(t)| \le L\,,$$

where the integer β_0 and the real $0 < \beta_1 \le 1$ are so that $\beta = \beta_0 + \beta_1$. The nonparametrically driven model, which we are going to deal with, assumes that we observe a sequence of independent r.v.'s $X_1, \dots, X_n$, with values in the measurable space $(X, \mathscr{X})$, so that, for each $i = 1, \dots, n$, the observation X_i has the density $p\big(x, f(i/n)\big)$, where the function f is unknown and satisfies the smoothness condition $f \in \Sigma^\beta$. We shall make use of the notation $P_f^n = P_{f(1/n)} \times \cdots \times P_{f(n/n)}$, where P_θ is the probability measure in the parametric experiment $\mathcal{E}$ and $f \in \Sigma^\beta$.

2.2 Regularity assumptions. Assume that $\beta > 1/2$. In the sequel we shall impose, on the density $p(x, \theta)$ in the parametric experiment $\mathcal{E}$, the regularity assumptions (R1–R3), which are assumed to hold true with the same $\varepsilon > 0$.

(R1) The function $s(\theta)=\sqrt{p(\theta)}$ is smooth in the space $L^2(X,\mathscr{X},\mu)$: there is a real number $\delta\in\left(\frac{1}{2\beta},1\right)$ and a map $\overset{\bullet}{s}(\theta)\colon\Theta\to L^2(X,\mathcal{X},\mu)$ such that

$$\sup_{(\theta,u)}\frac{1}{|u-\theta|^{1+\delta}}\left(\int\limits_X\bigl(s(x,u)-s(x,\theta)-(u-\theta)\overset{\bullet}{s}(x,\theta)\bigr)^2\mu(dx)\right)^{1/2}<\infty\,,$$

where sup is taken over all pairs (θ,u) satisfying $\theta,u\in\Theta$, $|u-\theta|\le\varepsilon$.

It is well-known (see Strasser [8]) that the function $\overset{\bullet}{s}(\theta)$ in condition (R1) can be written as $\overset{\bullet}{s}(\theta)=\frac{1}{2}\overset{\bullet}{l}(\theta)\sqrt{p(\theta)}$, μ-a.s. on X, where $\overset{\bullet}{l}(\theta)\in L^2(X,\mathscr{X},\mu)$. Moreover, $\overset{\bullet}{l}(\theta)\in L^2(X,\mathscr{X},P_\theta)$ and $\mathrm{E}_\theta\overset{\bullet}{l}(\theta)=0$, $\theta\in\Theta$, where E_θ is the expectation under P_θ. The map $\overset{\bullet}{l}(\theta)$ is called *tangent vector* at θ. For any $\theta\in\Theta$, define an *extended* tangent vector $\overset{\bullet}{l}_\theta(u)$, $u\in\Theta$, by setting

$$\overset{\bullet}{l}_\theta(x,u)=\begin{cases}\overset{\bullet}{l}(x,\theta)\,, & \text{if } u=\theta\,,\\ \frac{2}{u-\theta}\left(\sqrt{\frac{p(x,u)}{p(x,\theta)}}-1\right), & \text{if } u\ne\theta\,.\end{cases}$$

(R2) There is a real number $\delta\in\left(\frac{2\beta+1}{2\beta-1},\infty\right)$ such that

$$\sup_{(\theta,u)}\int\limits_X\bigl|\overset{\bullet}{l}_\theta(x,u)\bigr|^{2\delta}p(x,\theta)\mu(dx)<\infty\,,$$

where sup is taken over all pairs (θ,u) satisfying $\theta,u\in\Theta$, $|u-\theta|\le\varepsilon$.

The Fisher information in the local experiment $\mathcal{E}$ is

$$I(\theta)=\int\limits_X\bigl(\overset{\bullet}{l}(x,\theta)\bigr)^2p(x,\theta)\mu(dx)\,,\qquad\theta\in\Theta\,.$$

(R3) There are two real numbers $I_{\min}$ and $I_{\max}$ such that

$$0<I_{\min}\le I(\theta)\le I_{\max}<\infty\,,\qquad\theta\in\Theta\,.$$

2.3 Local result. First we state a local Gaussian approximation. For any $f\in\Sigma^\beta$, denote by $\Sigma^\beta_f(r)$ the neighbourhood of f, shifted to the origin:

$$\Sigma^\beta_f(r)=\bigl\{h\colon|h|\le r\,,\ f+h\in\Sigma^\beta\bigr\}\,.$$

Set

$$\gamma_n=c(\beta)\left(\frac{\log n}{n}\right)^{\frac{2\beta}{2\beta+1}},$$

where $c(\beta)$ is a constant depending on β. By definition, the *local experiment*

$$\mathcal{E}_f^n = \left(X^n,\ \mathscr{X}^n,\ \{P_{f+h}^n \colon h \in \Sigma_f^\beta(\gamma_n)\}\right)$$

is generated by the sequence of independent r.v.'s $X_1, \dots, X_n$, with values in the measurable space $(X, \mathscr{X})$, so that, for each $i = 1, \dots, n$, the observation X_i has the density $p\left(x, g(i/n)\right)$, where $g = f + h$, $h \in \Sigma_f^\beta(r_n)$.

Theorem 2.1. *Let $\beta > 1/2$ and $I(\theta)$ be the Fisher information in the parametric experiment $\mathcal{E}$. Assume that the density $p(x, \theta)$ satisfies the regularity conditions* (R1–R3). *For any $f \in \Sigma^\beta$, let $\mathcal{G}_f^n$ be the local Gaussian experiment, generated by the observations*

$$Y_i^n = h(i/n) + I\left(f(i/n)\right)^{-1/2} \varepsilon_i, \qquad i = 1, \dots, n,$$

with $h \in \Sigma_f^\beta(r_n)$, where $\varepsilon_1, \dots, \varepsilon_n$ is a sequence of i.i.d. standard normal r.v.'s. Then, uniformly in $f \in \Sigma^\beta$, the sequence of local experiments $\mathcal{E}_f^n$, $n = 1, 2, \dots$, is asymptotically equivalent to the sequence of local Gaussian experiments $\mathcal{G}_f^n$, $n = 1, 2, \dots$,

$$\sup_{f \in \Sigma^\beta} \Delta\left(\mathcal{E}_f^n, \mathcal{G}_f^n\right) \to 0, \qquad \text{as} \quad n \to \infty.$$

2.4 Global result. By definition, the *global experiment*

$$\mathcal{E}^n = \left(X^n, \mathscr{X}^n, \{P_f^n \colon f \in \Sigma^\beta\}\right)$$

is generated by the sequence of independent r.v.'s $X_1, \dots, X_n$, with values in the measurable space $\left(X, \mathscr{X}\right)$, so that, for each $i = 1, \dots, n$, the observation X_i has the density $p\left(x, f(i/n)\right)$, where $f \in \Sigma^\beta$. We shall use the following assumptions:

(G1) For any $\beta > \frac{1}{2}$, there is an estimator $\widehat{f}_n \colon X^n \to \Sigma^\beta$, such that

$$\sup_{f \in \Sigma^\beta} P\left(\left\|\widehat{f}_n - f\right\|_\infty \ge c\overline{\gamma}_n\right) \to 0, \qquad n \to \infty.$$

for any sequence $\overline{\gamma}_n \le 0$ satisfying $\overline{\gamma}_n^{-1} = O(\gamma_n^{-1})$.

(G2) The Fisher information $I(\theta) \colon \Theta \to (0, \infty)$ satisfies Hölder's condition with exponent $\alpha \in \left(\frac{1}{2\beta}, 1\right)$.

The main result of the paper is the following theorem, which states a global Gaussian approximation for the sequence of experiments $\mathcal{E}^n$, $n = 1, 2, \dots$, in the sense of the Le Cam distance.

Theorem 2.2. *Let $\beta > 1/2$ and $I(\theta)$ be the Fisher information in the parametric experiment $\mathcal{E}$. Assume that the density $p(x, \theta)$ satisfies the regularity* *conditions* (R1–R3) *and that conditions* (G1–G2) *are satisfied. Let* $\mathcal{G}^n$ *be the Gaussian experiment generated by the observations*

$$Y_i^n = \Gamma\bigl(f(i/n)\bigr) + \varepsilon_i\,, \qquad i = 1, \dots, n\,,$$

with $f \in \Sigma^\beta$, *where* $\varepsilon_1, \dots, \varepsilon_n$ *is a sequence of i.i.d. standard normal r.v.'s and* $\Gamma(\theta)\colon \Theta \to R$ *is any function satisfying* $\Gamma'(\theta) = \sqrt{I(\theta)}$. *Then the sequence of experiments* $\mathcal{E}^n$, $n = 1, 2, \dots$, *is asymptotically equivalent to the sequence of Gaussian experiments* $\mathcal{G}^n$, $n = 1, 2, \dots$:

$$\Delta(\mathcal{E}^n, \mathcal{G}^n) \to 0\,, \qquad \text{as} \quad n \to \infty\,.$$

An example in Brown and Low [1] shows that asymptotic equivalence, in general, fails to hold true when $\beta < \frac{1}{2}$.

Assumption (G1) is of technical nature. It can be easily checked in the most particular cases of interest. We refer the reader to Section 3 for some examples.

The function $\Gamma(\theta)$ can be related to the so called *variance-stabilizing* transformation, which we proceed to introduce. Let $X_1, \dots, X_n$ be a sequence of real valued i.i.d. r.v.'s, which depends on the parameter $\theta \in \Theta$. Let θ be the common mean and $\sigma(\theta)$ be the common variance. By the central limit theorem,

$$\sqrt{n}\{S_n - \theta\} \xrightarrow{d} N\bigl(0, \sigma(\theta)\bigr)\,,$$

where $S_n = (X_1 + \dots + X_n)/n$. The variance-stabilizing transformation is defined to be a function F on the real line, which stabilize the variance in the limiting normal law, i.e., such that

$$\sqrt{n}\bigl\{F(S_n) - F(\theta)\bigr\} \xrightarrow{d} N(0, 1)\,.$$

In this case the function Γ just coincides with the function $F\colon \Gamma(\theta) = F(\theta)$, $\theta \in \Theta$. If the parametrization in the model $\mathcal{E}$ is such that $\mu'(\theta) = \sigma(\theta) = I(\theta)$, where $\mu(\theta) = \mathrm{E}_\theta X$ is the mean and $\sigma(\theta) = \mathrm{Var}_\theta X$ of an observation X in the experiment $\mathcal{E}$, then $\Gamma(\theta) = F\bigl(\mu(\theta)\bigr)$, $\theta \in \Theta$.

Theorem 2.2 is proved using the local result in Theorem 2.1 by means of a globalizing procedure. The main idea of the proof of Theorem 2.1 is to decompose the initial experiment $\mathcal{E}^n$ into a product of independent experiments and to show that each component can be "well approximated" by means of its Gaussian counterpart. For the last one, we develop a general approach according to which any experiment $\mathcal{E}^n$, satisfying a nonparametric analog of the local asymptotic quadraticity (LAQ) condition, can be constructed on the same measurable space with a Gaussian experiment $\mathcal{G}^n$, such that the Hellinger distance between the corresponding probability measures converges to 0 at a certain rate, as $n \to \infty$. The main tool in establishing this result is the functional analog of the Hungarian construction developed in Grama and Nussbaum [2]. Then we are able to check the nonparametric LAQ conditions for the model under consideration. Our approach is similar to that in Grama and Nussbaum [3] or Nussbaum [7].

3. Examples

3.1 Linear regression model. Consider the nonparametric regression model with a non-Gaussian noise

$$X_i = f(i/n) + \xi_i\,, \qquad i = 1, \ldots, n\,, \tag{3.1}$$

where $\xi_1, \ldots, \xi_n$ are i.i.d. r.v.'s of means 0 and finite variances, with density $p(x)$ on the real line, $f \in \Sigma^\beta$ and Σ^β is a Hölder ball on $[0,1]$ with the exponent $\beta > \frac{1}{2}$. This model is a particular case of the nonparametrically driven model, introduced in Section 2, when $p(x,\theta) = p(x-\theta)$ is the shifted density $p(x)$ and $\theta \in R$. It is easy to see that conditions (R1–R3) hold true, if we assume that the density $p(x)$ obeys the following items:

(L1) The function $s(x) = \sqrt{p(x)}$ satisfies Hölder's condition with an exponent $1+\alpha$, where $\alpha \in \left(\frac{1}{2\beta}, 1\right)$.

(L2) For some $\delta > \frac{2\beta+1}{2\beta-1}$ and $\varepsilon > 0$, we have

$$\sup_{|u| \le \varepsilon} \int_{-\infty}^{\infty} \left| s'(x+u)/s(x) \right|^{2\delta} s(x)\, dx < \infty\,.$$

(L3) The Fisher informational number is positive:

$$I = \int_{-\infty}^{\infty} \left(p'(x)^2/p(x)\right) dx > 0\,.$$

It is well-known that a preliminary estimator satisfying condition (G1) exists. Then, under conditions (L1–L3), the nonparametric linear regression model (3.1) is asymptotically equivalent to a linear regression with Gaussian noise, in which we observe

$$Y_i = f(i/n) + I^{-1/2}\varepsilon_i\,, \qquad i = 1, \ldots, n\,,$$

where $\varepsilon_1, \ldots, \varepsilon_n$ are i.i.d. standard normal r.v.'s.

3.2 Exponential family model. Other particular cases arise when the parametric experiment $\mathcal{E} = \left(X, \mathscr{X}, \{P_\theta \colon \theta \in \Theta\}\right)$ is an one-dimensional linearly indexed exponential family, where Θ is a possibly infinite interval on the real line. The parameter set Θ is assumed to be such that the regularity conditions (R1)–(R3) are satisfied. The parametrization in the examples below is such that an observation X in the experiment $\mathcal{E}$ has the mean $\mu(\theta) = \theta$, since this form of the parametrization makes it easier to compute the function $\Gamma(\theta)$. Note that a preliminary estimator satisfying condition (G1) in the exponential family model exists (see Grama and Nussbaum [3]).

Spectral density model. Assume that we are given a sequence of normal observations $X_1, \dots, X_n$ with means 0 and standard deviations $f(i/n)$, where the function $f(t)$, $t \in [0,1]$, satisfies Hölder's condition with exponent $\beta > \frac{1}{2}$ and is such that $c_1 \le f(t) \le c_2$ for some positive absolute constants c_1 and c_2. In this model the density of the observations has the form $p(x,\theta) = \frac{1}{\sqrt{2\pi\theta}} \exp\left(-\frac{x^2}{2\theta^2}\right)$, $x \in \mathbb{R}$ and the Fisher information is $I(\theta) = 2\theta^{-2}$. This gives us $\Gamma(\theta) = \sqrt{2} \log \theta$. Then the model is asymptotically equivalent to the Gaussian model, in which we observe

$$Y_i = \sqrt{2} \log f(i/n) + \varepsilon_i\,, \qquad i = 1, \dots, n\,,$$

where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. standard normal r.v.'s.

Density model. Assume that we are given a sequence of Poisson observations $X_1, \dots, X_n$ with "unknown" intensities $f(i/n)$, where the function $f(t)$, $t \in [0,1]$ satisfies Hölder's condition with exponent $\beta > \frac{1}{2}$ and is such that $c_1 \le f(t) \le c_2$ for some positive absolute constants c_1 and c_2. In this model $p(x,\theta) = \theta^x \exp(-\theta)$, $x \in X = \{0, 1, \dots\}$, and $I(\theta) = \theta^{-1}$. This gives us $\Gamma(\theta) = 2\sqrt{\theta}$. Then the model is asymptotically equivalent to the Gaussian model, in which we observe

$$Y_i = 2\sqrt{f(i/n)} + \varepsilon_i\,, \qquad i = 1, \dots, n\,,$$

where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. standard normal r.v.'s. We called this model the density model, since the density model reduces to the last one by using Katz poissonization technique (see Nussbaum [7]).

Binary response model. Assume that we are given a sequence of Bernoulli observations $X_1, \dots, X_n$ taking values 0 and 1 with "unknown" probabilities of success $f(i/n)$, where the function $f(t)$, $t \in [0,1]$, satisfies Hölder's condition with exponent $\beta > \frac{1}{2}$ and is such that $c_1 \le f(t) \le c_2$ for some absolute constants $c_1 > 0$ and $c_2 < 1$. In this model, $p(x,\theta) = \theta^x (1-\theta)^{1-x}$, $x \in X = \{0,1\}$, and $I(\theta) = \frac{1}{\theta(1-\theta)}$. This gives us $\Gamma(\theta) = 2 \arcsin \sqrt{\theta}$. Then this model is asymptotically equivalent to the Gaussian model, in which we observe

$$Y_i = 2 \arcsin \sqrt{f(i/n)} + \varepsilon_i\,, \quad i = 1, \dots, n\,,$$

where $\varepsilon_1, \dots, \varepsilon_n$ are i.i.d. standard normal r.v.'s.

REFERENCES

[1] BROWN, L. D.—LOW, M.: *Asymptotic equivalence of nonparametric regression and white noise*, Ann. Statist. **24** (1992), 2384–2398.

[2] GRAMA, I. G.—NUSSBAUM, M.: *A nonstandard Hungarian construction for partial sums*, Preprint No. 324, Weierstrass-Institute, Berlin, 1997.

[3] GRAMA, I. G.—NUSSBAUM, M.: *Asymptotic equivalence for nonparametric generalized linear models*, Probability Theory and Related Fields. Preprint No. 289, Weierstrass-Institute, Berlin, 1986.

[4] IBRAGIMOV, I. A.—HAS'MINSKII, R. H.: *Statistical Estimation. Asymptotic Theory*, Springer-Verlag, New York, 1981.

[5] LE CAM, L.: *Asymptotic Methods in Statistical Decision Theory*, Springer-Verlag, New-York, 1986.

[6] LE CAM, L.—YANG, G.: *Asymptotics in Statistics: Some Basic Concepts*, Springer-Verlag, New York, 1990.

[7] NUSSBAUM, M.: *Asymptotic equivalence of density estimation and white noise*, Ann. Statist. **24** (1996), 2399–2430.

[8] STRASSER, H.: *Mathematical Theory of Statistics*, Walter de Gruyter, Berlin–New York, 1985.

Received April 9, 1998

Institute of Mathematics
Academy of Sciences
Academiei str. 5
277028 Kishinev
MOLDOVA

E-mail: ion.grama@math.uni-giessen.de