Tatra Mt. Math. Publ. **17** (1999), 37–44

TATRA MOUNTAINS
Mathematical Publications

THE ASYMPTOTIC REGULARITY OF A GENERALIZATION OF A LINEAR REGRESSION MODEL WITH A NONMONOTONOUS LINK FUNCTION

RADOSLAV HARMAN

ABSTRACT. Consider the nonlinear regression model $E(y_i) = g(\mathbf{x}_i^T\theta)$, $i = 1, \dots, n$, where y_i is a random vector of observations, g is a differentiable, not necessarily monotonous real function, θ is a parameter from some parameter space $\Theta \subseteq \mathbb{R}^m$, and $\mathbf{x}_i$ are m-dimensional vectors. Next, consider a sequence of such models corresponding to adding observations $(n \to \infty)$, and assume that the $\mathbf{x}_i$ are random. In the article, we formulate conditions, under which the sequence of (random) models is asymptotically regular almost surely in the local and in the global sense. More precisely, we give conditions which ensure us that: **1.** For given θ^*, a natural number m exists with probability one, such that, starting from the mth observation, we shall get models that are regular in θ^*. **2.** A natural number m exists with probability one, such that from the mth observation, we shall get models that are regular on the whole of Θ.

1. Introduction

Let Θ be an open subset of $\mathbb{R}^m$, and let $\eta\colon \Theta \to \mathbb{R}^n$ be a differentiable function. We shall denote the Jacobian of η in $\theta \in \Theta$ as $\partial\eta(\theta)$. That is: $\partial\eta(\theta)$ is the $n \times m$ matrix, such that for $1 \le i \le n$ and $1 \le j \le m$: $\big(\partial\eta(\theta)\big)_{ij} = \frac{\partial\eta_i}{\partial\theta_j}(\theta)$.

Let us consider a nonlinear regression model

$$E\mathbf{y} = \eta(\theta), \tag{G}$$

where $\mathbf{y}$ is an n-dimensional random vector of observations. The model is said to be *locally regular* in $\theta \in \Theta$, if $\operatorname{rank}\big(\partial\eta(\theta)\big) = m$ and *globally regular* on Θ, if it is locally regular for all $\theta \in \Theta$. (For details of this notion see [1]). It is

2000 Mathematics Subject Classification: 62J02, 62K99, 68T05.
Keywords: statistical nonlinear regression, regularity, asymptotic properties, artificial neural networks.

Supported by Grant VEGA 1/4196/97.

very often necessary to consider a sequence $\eta_k\colon \Theta \to \mathbb{R}^k$, $k \in \mathbb{N}$, of regression models $E\boldsymbol{y}_k = \eta_k(\theta)$, where $\boldsymbol{y}_k$ can be interpreted as the k-dimensional random vector of the "first" k observations. We shall call this sequence *asymptotically locally regular* in $\theta \in \Theta$, if for some $n_0 \in \mathbb{N}$ and for all $n \geq n_0$ we have $\operatorname{rank}(\partial\eta_n(\theta)) = m$, and similarly, *asymptotically globally regular* on Θ, if for some $n_0 \in \mathbb{N}$, for all $\theta \in \Theta$ and $n \geq n_0$: $\operatorname{rank}(\partial\eta_n(\theta)) = m$.

In many cases, it is natural to assume that the rules by which the consequent models η_k are chosen have a probabilistic interpretation. Hence, for example we can speak of asymptotic global regularity *almost surely*. For example consider the linear regression line, where the ith observation is taken in a real number x_i for $i = 1, 2, \dots$ If we choose the points x_i independently and randomly from the uniform distribution on a finite interval, we can easily check that the corresponding sequence of regression models will be asymptotically globally regular almost surely.

It turns out to be difficult to deal with the general problem of asymptotical regularity because of its diversity. In this article, we focus on the regularity properties of the following generalization of the linear regression model.

Let $g\colon \mathbb{R} \to \mathbb{R}$ be a fixed continuously differentiable function, and let $\mathbf{X} = [\boldsymbol{x}_1, \dots, \boldsymbol{x}_n]$ be a real matrix of the type $m \times n$. We consider a regression model of the form (G), where $\eta\colon \Theta \subseteq \mathbb{R}^m \to \mathbb{R}^n$ is defined via the function g and the "design matrix" $\mathbf{X}$ as $\eta(\theta) = \big(g(\boldsymbol{x}_1^T\theta), \dots, g(\boldsymbol{x}_n^T\theta)\big)^T$. Sometimes, we shall denote the function η as $\eta_{\mathbf{X}}$, in order to emphasize the dependence of η on the matrix $\mathbf{X}$. We shall refer to this model by the symbol (M).

Notice, that if we require g to be monotonic, we get the standard GLIM. (See [2]). The fact that we do not assume that g is monotonic, could be useful for example, for the theory of artificial neural networks. Here (M) corresponds in a certain sense to the most common model of a single neuron, and the function g can be interpreted as the so-called activation function of a neuron. Of course, we are always glad, if we are able to show some nice properties of a model without too many assumptions about the behavior of a *biological* element. See for example, [3] for a survey of statistical methods used in the theory of neural networks.

We can easily check that, in our case, the Jacobian has the following form:

$$\partial\eta(\theta) = \big(g'(\boldsymbol{x}_1^T\theta)\boldsymbol{x}_1, \dots, g'(\boldsymbol{x}_n^T\theta)\boldsymbol{x}_n\big)^T = \mathbf{D}(\theta)\mathbf{X}^T\,, \tag{J}$$

where $\mathbf{D}(\theta) = \mathbf{D}_{\mathbf{X}}(\theta) = \operatorname{diag}\big(g'(\boldsymbol{x}_1^T\theta), \dots, g'(\boldsymbol{x}_n^T\theta)\big)$ and $g'(\cdot) = \frac{dg}{dt}(\cdot)$. Throughout the article, we use the notation T_0 for the set of stationary points of the function g; $T_0 = \{t_0 \in \mathbb{R} : g'(t_0) = 0\}$. Since we suppose that g' is continuous, the set T_0 is measurable.

We denote by $\mathcal{M}_{p,q}$ the set of all real $p \times q$ matrices. By the Lebesgue measure λ on $\mathcal{M}_{p,q}$ we mean the measure naturally induced from the Lebesgue measure on $\mathbb{R}^{pq}$. We denote by $\mathcal{V}_m$ the set of all singular $m \times m$ matrices, and by $\mathcal{W}_m$ the set of matrices $m \times (m+1)$, $\mathbf{A} = (\boldsymbol{a}_1, \dots, \boldsymbol{a}_{m+1})$, for which the vectors $\boldsymbol{a}_1 - \boldsymbol{a}_{m+1}, \dots, \boldsymbol{a}_m - \boldsymbol{a}_{m+1}$ are linearly dependent. For the proofs, we need the following well-known lemma (See [4]).

Lemma 1. *a) $\mathcal{V}_m$ is a measurable subset of $\mathcal{M}_{m,m}$, and $\lambda(\mathcal{V}_m) = 0$,*
b) $\mathcal{W}_m$ is a measurable subset of $\mathcal{M}_{m,m+1}$ and $\lambda(\mathcal{W}_m) = 0$.

2. Local regularity of the model (M)

Proposition 2. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of m-dimensional random vectors, such that for all $n \in \mathbb{N}$ the random matrix $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ has probability distribution continuous with respect to λ, and let T_0 be at most countable. Let $\theta \in \Theta$, and suppose that either $0 \notin T_0$, or $\theta \neq 0$. Then for all $n \geq m$: $P\big[\mathrm{rank}(\partial\eta_{\mathbf{X}_n}(\theta)) = m\big] = 1$.*

P r o o f. Take any $\theta \in \Theta$. Since $\mathrm{rank}(\partial\eta_{\mathbf{X}_n}(\theta)) \geq \mathrm{rank}(\partial\eta_{\mathbf{X}_m}(\theta))$ for any $n \geq m$, it is sufficient to prove the proposition for $n = m$.

Let $\mathcal{Q}$ be the set of matrices $\mathbf{X} \in \mathcal{M}_{m,m}$, for which $\mathrm{rank}(\partial\eta_{\mathbf{X}}(\theta)) < m$. Under the condition that either $0 \notin T_0$ or $\theta \neq 0$, we have that for all $t \in T_0$ and $i \in \{1, 2, \dots, m\}$, the set $\mathcal{T}_i(t) = \big\{[\mathbf{x}_1, \dots, \mathbf{x}_m] \in \mathcal{M}_{m,m} \colon \mathbf{x}_i^T\theta = t\big\}$ is λ-measurable, and $\lambda(\mathcal{T}_i(t)) = 0$. (It is a shifted *proper* subspace of $\mathcal{M}_{m,m}$.) By (J), if $\partial\eta_{\mathbf{X}_m}(\theta)$ is singular, then either $\mathbf{x}_1, \dots, \mathbf{x}_m$ are linearly dependent, or $\mathbf{x}_i^T\theta \in T_0$ for some $i \in \{1, \dots, m\}$. This fact can be expressed as: $\mathcal{Q} = \mathcal{V}_m \cup \left(\bigcup_{i=1}^{m} \bigcup_{t \in T_0} \mathcal{T}_i(t)\right)$. Hence, $\mathcal{Q}$ is a countable union of zero sets and consequently $\lambda(\mathcal{Q}) = 0$. □

A simple consequence of the previous proposition is that (under the stated conditions) the model (M) is asymptotically locally regular in θ. Note, that the assumption that $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ has continuous distribution excludes many interesting cases, for example the case of a simple regression line. Here the vectors $\mathbf{x}_i$ are of the form $(1, x_i)$, $x_i \in \mathbb{R}$, and hence $\mathbf{X}_n$ is *not* distributed continuously w.r.t. λ. This case is covered by the following general proposition. We shall say, that an m-dimensional random vector $\mathbf{x}$ has *linearly independent distribution*, if for any proper linear subspace L of $\mathbb{R}^m$: $P[\mathbf{x} \in L] = 0$.

Proposition 3. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of independent, identically distributed m-dimensional random vectors with linearly independent distribution. Let $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ for all $n \in \mathbb{N}$, and let $\theta \in \Theta$. Then the sequence of random variables: $\big(\mathrm{rank}(\partial\eta_{\mathbf{X}_n}(\theta))\big)_{n \in \mathbb{N}}$ is a homogeneous Markov chain with the*

$m \times m$-matrix $\mathbf{P}$ of transition probabilities given by:

$$(\mathbf{P})_{ij} = \begin{cases} p & \text{if } j = i+1,\ i \in \{0, \dots, m-1\}, \\ 1-p & \text{if } j = i,\ i \in \{0, \dots, m-1\}, \\ 1 & \text{if } i = j = m, \\ 0 & \text{else}, \end{cases}$$

where $p = P\left[\mathbf{x}_1^T \theta \notin T_0\right]$.

P r o o f. Let $\theta \in \Theta$. The vectors $\mathbf{x}_i$ are i.i.d. That is why the sequence $\left(\operatorname{rank}\left(\mathbf{D}_{\mathbf{X}_n}(\theta)\right)\right)_{n \in \mathbb{N}}$ is clearly a homogeneous Markov chain with transition probabilities $q_{ij} = p \cdot \delta_{i+1,j} + (1-p)\delta_{ij}$ for all $i, j \in \mathbb{N}_0$ ($\delta_{ij} = 1 \Leftrightarrow i = j$). Moreover, the fact that the distribution of $\mathbf{x}_i$ is linearly independent, together with (J) implies that, with probability one, $\operatorname{rank}\left(\partial\eta_{\mathbf{X}_n}(\theta)\right) = \operatorname{rank}\left(\mathbf{D}_{\mathbf{X}_n}(\theta)\mathbf{X}_n^T\right) = \min\left\{\operatorname{rank}\left(\mathbf{D}_{\mathbf{X}_n}(\theta)\right), m\right\}$. In other words, on a set of probability one, the process $\left(\operatorname{rank}\left(\partial\eta_{\mathbf{X}_n}(\theta)\right)\right)_{n \in \mathbb{N}}$ is derived from the Markov chain $\left(\operatorname{rank}\left(\mathbf{D}_{\mathbf{X}_n}(\theta)\right)\right)_{n \in \mathbb{N}}$ by amalgamating (unifying) states $m, m+1 \dots$ into the state m. ☐

COROLLARY 4. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of independent, identically distributed random vectors with linearly independent distribution, and let $p = P\left[\mathbf{x}_1^T \theta \notin T_0\right]$. Then for all $n \geq m$: $P\left[\operatorname{rank}\left(\partial\eta_{\mathbf{X}_n}(\theta)\right) = m\right] = \sum_{j=m}^{n} \binom{n}{j} p^j (1-p)^{n-j}$.*

As a consequence we get, that if $\mathbf{x}_1, \mathbf{x}_2, \dots$ is an i.i.d. sequence with linearly independent distribution, then the model (M) is asymptotically locally regular in θ if and only if $p > 0$.

3. Global regularity

Firstly, as in the previous paragraph, let us focus on the situation when $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ has continuous distribution on $\mathcal{M}_{m,n}$ w.r.t. λ. It turns out, that the global regularity behavior of (M) is very simple in the case $T_0 = \emptyset$:

PROPOSITION 5. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of random vectors, such that for all $n \in \mathbb{N}$ the matrix $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ has distribution on $\mathcal{M}_{m,n}$ continuous with respect to λ, and let $T_0 = \emptyset$. Then for all $n \geq m$: $P\left[\forall \theta \in \mathbb{R}^m\colon \operatorname{rank}\left(\partial\eta_{\mathbf{X}_n}(\theta)\right) = m\right] = 1$.*

P r o o f. As in the Proposition 2, it is sufficient to consider the case $n = m$. If $T_0 = \emptyset$, then the diagonal matrix $\mathbf{D}_{\mathbf{X}_m}(\theta)$ is always regular, therefore $\operatorname{rank}\left(\partial\eta_{\mathbf{X}_m}(\theta)\right) = \operatorname{rank}(\mathbf{X}_m)$. Hence, the rank of $\partial\eta_{\mathbf{X}_m}(\theta)$ is the same for all $\theta \in \Theta$. The proof then follows from Proposition 2. ☐

Hence, if $T_0 = \emptyset$ and $n \geq m$, the model (M) is almost surely globally regular on the whole of $\mathbb{R}^m$. In the case of nonempty T_0, less than $2m$ observations (i.e., $n < 2m$) is too few:

Proposition 6. *Let $\mathbf{x}_1, \mathbf{x}_2, \ldots$ be any sequence of m-dimensional (nonrandom) vectors and $\mathbf{X}_n = [\mathbf{x}_1, \ldots, \mathbf{x}_n]$ for $n \in \mathbb{N}$. Let us suppose that $T_0 \neq \emptyset$. Then for all $n < 2m$ we can find $\theta^* \in \mathbb{R}^m$, such that $\operatorname{rank}(\partial\eta_{\mathbf{X}_n}(\theta^*)) < m$.*

P r o o f. If $\operatorname{rank}(\mathbf{X}_n) = m$, then $\mathbf{X}_n$ contains m linearly independent columns, say $\mathbf{x}_1, \ldots, \mathbf{x}_m$. Let $t_0 \in T_0$ and let θ^* be the solution of the (regular) linear system $\mathbf{x}_1^T\theta = t_0, \ldots, \mathbf{x}_m^T\theta = t_0$. Then from (J) we get: $\operatorname{rank}(\partial\eta_{\mathbf{X}_n}(\theta)) \leq \operatorname{rank}(\mathbf{D}_{\mathbf{X}_n}(\theta)) \leq n - m < m$. The case of singular $\mathbf{X}_n$ is trivial. $\square$

Moreover, the situation is complicated by the fact that (under the assumptions of the previous proposition) there are singular points θ^*, such that for any open neighbourhood U of θ^* there is a *regular* parameter $\theta \in U$, except of course for the case when $\operatorname{rank}(\partial\eta_{\mathbf{X}_n}(\theta)) = 0$ for all θ. In fact, if $\operatorname{rank}(\mathbf{X}_n) = m$, then the set of singular points has Lebesgue measure zero.

On the other hand, if $n \geq 2m$, and T_0 is countable, then "most" of models of the type (M) are globally regular, as is precisely stated in the next proposition.

Proposition 7. *Let $n \geq 2m$. Let T_0 be at most countable and $0 \notin T_0$. Next, let $\mathcal{Q}$ be the set of design matrices $\mathbf{X} \in \mathcal{M}_{m\times n}$ for which $\exists\theta \in \mathbb{R}^m$, such that $\operatorname{rank}(\partial\eta_{\mathbf{X}_n}(\theta)) < m$. Then $\mathcal{Q}$ is a λ-measurable subset of $\mathcal{M}_{m\times n}$, and $\lambda(\mathcal{Q}) = 0$.*

P r o o f. Let $\mathbf{X} = [\mathbf{x}_1, \ldots, \mathbf{x}_n] \in \mathcal{Q}$. Consider the following two statements:
(a): Some m vectors from among $\mathbf{x}_1, \ldots, \mathbf{x}_n$ are linearly dependent, and
(b): For some $\theta \in \Theta$, more that $n - m$ diagonal elements of $\mathbf{D}(\theta)$ are zeroes.

It can be seen, that the logical conjunction of the converse of (a) and (b) implies $\mathbf{X} \notin \mathcal{Q}$. Consequently, at least one of the two statements is true. Hence, if we denote by $\mathcal{U}$ the set of matrices $\mathbf{X}$ satisfying (a) and by $\mathcal{K}$ the set of matrices satisfying (b), then $\mathcal{Q} \subseteq \mathcal{U} \cup \mathcal{K}$. We shall prove that both $\mathcal{U}$ and $\mathcal{K}$ are measurable sets and $\lambda(\mathcal{U}) = \lambda(\mathcal{K}) = 0$.

Let ζ be a subset of $\{1, \ldots, n\}$ consisting of m elements and let $\mathcal{U}_\zeta$ be the set of matrices $m \times n$ $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$, such that $\{\mathbf{x}_i : i \in \zeta\}$ is a set of linearly dependent vectors. Let $\zeta = \{1, \ldots, m\}$ for example. Then $\mathcal{U}_\zeta = [\mathcal{V}_m;\ \mathcal{M}_{m,n-m}]$ and consequently from the Lemma 1 we get $\lambda(\mathcal{U}_\zeta) = 0$. Since it is obvious that the same holds for all $\binom{n}{m}$ sets ζ, we can conclude that $\lambda(\mathcal{U}) = \lambda\left(\bigcup_\zeta \mathcal{U}_\zeta\right) = 0$.

Now let ξ be a subset of $\{1, \ldots, n\}$ consisting of $m+1$ elements and let $\xi_1 < \xi_2 < \cdots < \xi_{m+1}$ be the elements of ξ. For any such ξ and for any $\boldsymbol{t} = (t_1, \ldots, t_{m+1}) \in T_0^{m+1}$ let $\mathcal{K}_{\xi,\boldsymbol{t}} \subseteq \mathcal{M}_{m,n}$ be the set of matrices $[\mathbf{x}_1, \ldots, \mathbf{x}_n]$, for which there exists $\theta \in \mathbb{R}^m$, such that $\mathbf{x}_{\xi_1}^T\theta = t_1$, $\mathbf{x}_{\xi_2}^T\theta = t_2, \dots, \mathbf{x}_{\xi_{m+1}}^T\theta = t_{m+1}$. Let us observe, that $\theta \neq 0$, since otherwise clearly $0 \in T_0$, which we exclude.

Next, assume that $\xi = \{1, \dots, m+1\}$ and take a fixed $\boldsymbol{t} \in T_0^{m+1}$. We consider the mapping $l \colon \mathcal{M}_{m,n} \to \mathcal{M}_{m,n}$ defined by $[\mathbf{x}_1, \dots, \mathbf{x}_n] \to^l \left[\frac{\mathbf{x}_1}{t_1}, \dots, \frac{\mathbf{x}_{m+1}}{t_{m+1}}, \mathbf{x}_{m+2}, \dots, \mathbf{x}_n\right]$. (Again, we use the fact that $0 \notin T_0$.) Next, take any $\mathbf{X} = [\mathbf{x}_1, \dots, \mathbf{x}_n] \in \mathcal{K}_{\xi,\boldsymbol{t}}$. Then for all $i \in \{1, 2, \dots, m\}$ and for some nonzero θ: $\theta^T\left(\frac{\mathbf{x}_i}{t_i}\right) - \theta^T\left(\frac{\mathbf{x}_{m+1}}{t_{m+1}}\right) = 1 - 1 = 0$, which obviously implies that the matrix consisting of the first $m+1$ vectors of $l(\mathbf{X})$ is an element of the set $\mathcal{W}_m$. Hence $l(\mathcal{K}_{\xi,\boldsymbol{t}}) \subseteq [\mathcal{W}_m; \mathcal{M}_{m,n-m-1}]$, which together with the Lemma 1 implies $\lambda\big(l(\mathcal{K}_{\xi,\boldsymbol{t}})\big) = 0$. Consequently, as l is a regular linear transformation, $\lambda(\mathcal{K}_{\xi,\boldsymbol{t}}) = 0$.

Since, clearly, we can argue similarly for all such sets ξ, we conclude that $\lambda(\mathcal{K}) = \lambda\left(\bigcup_{\xi} \bigcup_{\boldsymbol{t} \in T_0^{m+1}} \mathcal{K}_{\xi,\boldsymbol{t}}\right) = 0$.

Therefore, $\lambda(\mathcal{U} \cup \mathcal{K}) = 0$ and from the completeness of the Lebesgue measure we see that $\mathcal{Q}$ is a measurable set and $\lambda(\mathcal{Q}) = 0$. □

Immediately we get:

Corollary 8. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of random vectors such that for all $n \in \mathbb{N}$ the matrix $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ has a continuous distribution on $\mathcal{M}_{m,n}$ and let $n \geq 2m$. Let T_0 be at most countable and $0 \notin T_0$. Then $P\big[\forall \theta \in \Theta \colon \operatorname{rank}\big(\partial\eta_{\mathbf{X}}(\theta)\big) = m\big] = 1$. (The model (M) is almost surely globally regular on $\mathbb{R}^m$.)*

In the final part of the paper, we deal with the case when $\mathbf{x}_1, \mathbf{x}_2, \dots$ is a sequence of i.i.d. vectors with linearly independent distribution. Note, that in contrast to the Proposition 7 we do not exclude the models, for which T_0 is uncountable.

To shorten the formulation of the following proposition, let us define the set $\Gamma(\mathbf{x}) = \{\theta \in \Theta \colon \mathbf{x}^T\theta \in T_0\}$ for $\mathbf{x} \in \mathbb{R}^m$.

Proposition 9. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of independent and identically distributed m-dimensional random vectors with linearly independent distribution. Let $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ for $n \in \mathbb{N}$. Then the model (M) is asymptotically globally regular on Θ with probability 1, if and only if $P\left[\bigcap_{i=1}^{k} \Gamma(\mathbf{x}_i) = \emptyset\right] > 0$ for some $k \in \mathbb{N}$.*

P r o o f. With probability one, for $n \geq m$, any set of m columns of $\mathbf{X}_n$ creates a system of linearly independent vectors. Hence, since $\partial\eta_{\mathbf{X}_n}(\theta) = \mathbf{D}_{\mathbf{X}_n}(\theta)\mathbf{X}_n^T$, the model (M) is almost surely asymptotically globally regular, iff with probability one there is $n_0 \in \mathbb{N}$, such that for all $n \geq n_0$ and $\theta \in \Theta$, the number of nonzero elements from among $g'(\mathbf{x}_1^T\theta), \dots, g'(\mathbf{x}_n^T\theta)$ is not less that m. As the vectors $\mathbf{x}_i$ are i.i.d., this condition is satisfied if and only if with probability 1 for some $l \in \mathbb{N}$ and for all $\theta \in \Theta$ at least *one* number from among $g'(\mathbf{x}_1^T\theta), \dots, g'(\mathbf{x}_l^T\theta)$ is nonzero. In other words, the model (M) is globally asymptotically regular iff $P\Big[\exists l \in \mathbb{N}\colon \bigcap_{i=1}^{l} \Gamma(\mathbf{x}_i) = \emptyset\Big] = 1$. We can easily prove that this condition is satisfied iff for some $k \in \mathbb{N}$: $P\left[\bigcap_{i=1}^{k} \Gamma(\mathbf{x}_i) = \emptyset\right] = p > 0$. Indeed, let this condition hold and for $j \in \mathbb{N}$ create the event: $A_j = \left[\bigcap_{i=jk+1}^{jk+k} \Gamma(\mathbf{x}_i) = \emptyset\right]$. Supposing that the $\mathbf{x}_i$ are i.i.d., we have that for all $r \in \mathbb{N}$: $P\left[\exists l \in \mathbb{N}\colon \bigcap_{i=1}^{l} \Gamma(\mathbf{x}_i) = \emptyset\right] \geq P\left[\bigcap_{i=1}^{rk} \Gamma(\mathbf{x}_i) = \emptyset\right] \geq P\left[\bigcup_{j=0}^{r-1} A_j\right] = 1 - (1-p)^r$, since A_j are mutually independent and $P[A_j] = p$. $\square$

Corollary 10. *Let $\mathbf{x}_1, \mathbf{x}_2, \dots$ be a sequence of independent, continuously and identically distributed m-dimensional random vectors with* $\operatorname{supp} P^{\mathbf{x}_1} = \mathbb{R}^m$. *(E.g., let $\mathbf{x}_i$ have regular normal distribution.) Let $\mathbf{X}_n = [\mathbf{x}_1, \dots, \mathbf{x}_n]$ for $n \in \mathbb{N}$. Then any of the assumptions: a. T_0 is bounded and $\Theta = \mathbb{R}^m$, or b. $T_0 \neq \mathbb{R}$ and Θ is bounded, implies asymptotic global regularity of (M).*

P r o o f. At the outset, we note that the random vectors $\mathbf{x}_i$ have linearly independent distribution, since they are distributed continuously.

a) Specifically, the probability that the vectors $\mathbf{x}_1, \mathbf{x}_2$ will be collinear is zero. Therefore, as T_0 is a subset of a finite interval, the probability that $\Gamma(\mathbf{x}_1) \cap \Gamma(\mathbf{x}_2)$ is unbounded, is zero. As $\operatorname{supp} P^{\mathbf{x}_3} = \mathbb{R}^m$, the probability that $\Gamma(\mathbf{x}_3)$ does not intersect *any* bounded set, is nonzero. Consequently, $P\big[\Gamma(\mathbf{x}_1) \cap \Gamma(\mathbf{x}_2) \cap \Gamma(\mathbf{x}_3) = \emptyset\big] > 0$. The proof of a) then follows from the Proposition 9.

b) As g' is continuous, $T_0 \subseteq (-\infty, a\rangle \cup \langle b, \infty)$ for some $a < b$. Supposing that Θ, we can cover Θ with a finite number of spheres $S_1, \dots, S_k$, all of which have the diameter $\frac{b-a}{2}$. It can be easily shown, that since $\operatorname{supp} P^{\mathbf{x}_i} = \mathbb{R}^m$, we have $P\big[\Gamma(\mathbf{x}_i) \cap S_i = \emptyset\big] > 0$. Therefore, from the independence of $\mathbf{x}_i$ we get $P\big[\forall i = 1, \dots, k\colon \Gamma(\mathbf{x}_i) \cap S_i = \emptyset\big] > 0$. Since $\bigcap_{i=1}^{k} \Gamma(\mathbf{x}_i) \neq \emptyset$ implies: $\exists i \in \{1, \dots, k\}$ such that $\Gamma(\mathbf{x}_i) \cap S_i \neq \emptyset$, we conclude that $\left[\bigcap_{i=1}^{k} \Gamma(\mathbf{x}_i) = \emptyset\right] \supseteq \big[\forall i = 1, \dots, k\colon \Gamma(\mathbf{x}_i) \cap S_i = \emptyset\big]$. Whence, $P\left[\bigcap_{i=1}^{k} \Gamma(\mathbf{x}_i) = \emptyset\right] > 0$. $\square$

REFERENCES

[1] PÁZMAN, A.: *Nonlinear Statistical Models*, Kluwer, Dodrecht, 1993.

[2] MC CULLAGH, P.—NEDLER, J. A.: *Generalized Linear Models*, Chapman and Hall, London, 1983.

[3] CHENG, B.—TITTERINGTON, D. M.: *Neural Networks: A Review from a Statistical Perspective*, Statistical Science, Vol 9., No. 1, 1994.

[4] GOLUBICKIJ, M.—GIJEMIN, V.: *Stable Mappings and Their Singularities*, Springer-Verlag, New York, 1973.

Received May 7, 1998

Department of Probability and Statistics
Faculty of Mathematics and Physics
Comenius University
Mlynská dolina
SK–842 15 Bratislava
SLOVAKIA
E-mail: harman@fmph.uniba.sk