A Note on Posterior Consistency of Nonparametric Poisson Regression Models

Natesh S. Pillai* Robert L. Wolpert
Merlise A. Clyde
Department of Statistical Science
Duke University, Durham, NC 27708-0251

Revised: July 16, 2007

Summary

We introduce a new truncation approach to extend earlier methods for proving consistency in nonparametric Bayesian regression problems to non-compact state spaces. We illustrate the approach by proving posterior consistency for a nonparametric Poisson regression model. The key step is separating points in the parameter space by constructing hypothesis tests with suitably small error rates; we do this for individual pairs of points using our truncation approach, and then exploit the monotone likelihood-ratio property of the Poisson family to show that the tests have exponentially decaying errors of types I and II.

Key words: Bayesian; nonparametric Poisson regression; posterior consistency; stochastic ordering; monotone likelihood ratio.

1 Introduction

Posterior consistency for non-identically distributed random variables has received attention only recently in the literature. The general framework for proving such results was laid out by Amewou-Atisso et al. (2003), Ghosal & van der Vaart (2007a,b) and Choi & Schervish (2007). Choi & Schervish (2007) prove the posterior consistency for a regression model with Gaussian error terms. Recently posterior consistency was proved for nonparametric binary regression by Ghosal & Roy (2006), who conjectured that similar results might hold for Poisson regression.

In this note we verify that conjecture, proving posterior consistency for Poisson regression model under moderate assumptions. The standard proof technique requires one to verify

*Email: nsp2@stat.duke.edu

two conditions: positivity of prior mass for all information-metric (or equivalently Kullback-Leibler) neighborhoods of the true parameter, and existence of point-separating hypothesis tests with exponentially decaying error rates of types I and II. The first condition is prior-specific, and is relatively easy to verify. The existence of tests is independent of the prior and usually requires the calculation of the metric entropy of the parameter space. Since the Poisson distribution has support over the entire non-compact set of positive integers, we introduce a truncation approach, then utilize the Monotone Likelihood Ratio (MLR) property of the Poisson family to obtain the required exponential error rates.

2 Main Result

Let $\mathcal{X}$ be a compact subset of $\mathbb{R}^d$. For any vector $k = (k_1, k_2, \cdots, k_d) \in \mathbb{N}^d$, let D^k denote the differential operator,

$$D^k \equiv \frac{\partial^{k_+}}{\partial x_1^{k_1} \cdots \partial x_d^{k_d}},$$

where $k_+ \equiv \sum_{i=1}^d k_i$. Denote by $\lfloor \alpha \rfloor$ the greatest integer less than or equal to any $\alpha \in \mathbb{R}_+$. For a function $f : \mathcal{X} \mapsto \mathbb{R}$, set

$$\|f\|_\alpha \equiv \max_{\tilde{k} \le \lfloor \alpha \rfloor} \sup_{x \in \mathcal{X}} |D^k f(x)| + \max_{\tilde{k} \le \lfloor \alpha \rfloor} \sup_{x,y \in \mathcal{X}} \frac{|D^k f(x) - D^k f(y)|}{||x-y||^{\alpha - \lfloor \alpha \rfloor}}, \tag{1}$$

where the supremum is taken over all x, y in the interior of $\mathcal{X}$ with $x \neq y$. Let $C^\alpha(\mathcal{X})$ be the set of all continuous functions $f : \mathcal{X} \mapsto \mathbb{R}$ with $\|f\|_\alpha < \infty$ and, for $\gamma > 0$, set

$$\Theta_\gamma \equiv \{\theta \in C^\alpha(\mathcal{X}) : \quad (\forall x)\ \theta(x) \ge \gamma\}, \tag{2}$$

the set of C^α functions bounded below uniformly by $\gamma > 0$, with Borel σ-field $\mathcal{B}(\Theta_\gamma)$. We consider the following Poisson regression model:

$$\mathbb{P}(Y_i = y | X_i, \theta) = \exp\big(-\theta(X_i)\big) \frac{\theta(X_i)^y}{y!}, \qquad \theta \in \Theta_\gamma,\ y \in \mathbb{Z}_+,\ i \in \mathbb{N}. \tag{3}$$

For $\theta \in \Theta_\gamma$ let $\mathbb{P}_\theta$ be the distribution of conditionally-independent $\{Y_i\}$ generated according to the model (3) with mean function θ. Let $\|\cdot\|_*$ denote the uniform norm in Θ_γ, and for any $\theta \in \Theta_\gamma$ and $\delta > 0$ define

$$B_\delta(\theta) \equiv \{\tilde{\theta} \in \Theta_\gamma : \|\tilde{\theta} - \theta\|_* < \delta\}.$$

For any increasing sequence of positive numbers $\{M_n\}$ (we will choose a specific sequence in Section 2.2 below), set

$$\Theta_\gamma^n \equiv \{\theta \in \Theta_\gamma : \|\theta\|_\alpha \le M_n\}. \tag{4}$$

Theorem 2.1. *Let Θ_γ be as in (2) and let Π be a (prior) probability measure on $\mathcal{B}(\Theta_\gamma)$ such that $\Pi\big(B_\delta(\theta)\big) > 0$ for each $\theta \in \Theta_\gamma$ and $\delta > 0$. Then for any $\epsilon > 0$, and any $\theta_0 \in \Theta_\gamma$,*

$$\lim_{n\to\infty} \Pi\left(\left\{\theta \in \Theta_\gamma^n : \int_{\mathcal{X}} |\theta(x) - \theta_0(x)|\, Q_n(dx) > \epsilon\right\} \middle| Y_1, Y_2, \cdots, Y_n\right) = 0$$

in $\mathbb{P}_{\theta_0}$ probability, where $Q_n(dx) \equiv \frac{1}{n}\sum_{i=1}^n \delta_{X_i}(dx)$ denotes the empirical distribution.

Proof. Fix $\epsilon > 0$ and $\theta_0 \in \Theta_\gamma$, and let

$$\begin{aligned}\Lambda_i(\theta_0, \theta) &\equiv \log \frac{\exp\big(-\theta_0(X_i)\big)\big(\theta_0(X_i)\big)^{Y_i}}{\exp\big(-\theta(X_i)\big)\big(\theta(X_i)\big)^{Y_i}} \\ &= \big[\theta(X_i) - \theta_0(X_i)\big] + Y_i \log \frac{\theta_0(X_i)}{\theta(X_i)}\end{aligned}$$

denote the log ratio of two Poisson densities with means $\theta_0(X_i)$ and $\theta(X_i)$, evaluated at Y_i. Set

$$K_i(\theta_0, \theta) \equiv \mathbb{E}_{\theta_0}\big(\Lambda_i(\theta_0, \theta)\big)$$

(the Kullback-Leibler divergence for observation i), and

$$V_i(\theta_0, \theta) \equiv \mathsf{Var}_{\theta_0}\big(\Lambda_i(\theta_0, \theta)\big).$$

Also set

$$U_n^\epsilon \equiv \left\{\theta \in \Theta_\gamma : \int |\theta(x) - \theta_0(x)|\ Q_n(dx) > \epsilon\right\}.$$

Our goal is to verify the following two conditions:

I) Positive prior probability: There exists a set $B \subset \Theta_\gamma$ with $\Pi(B) > 0$ such that

1. $(\forall \theta \in B)$ $\sum_{i=1}^\infty V_i(\theta_0, \theta)/i^2 < \infty$.
2. $(\forall i \in \mathbb{N})$ $\Pi\big(B \cap \{\theta : K_i(\theta_0, \theta) < \epsilon\}\big) > 0$.

II) Existence of tests: There exist non-negative bounded measurable functions $\{\Phi_n\}_{n=1}^\infty$ and positive constants $C_1, c_1 > 0$ such that

1. $\sum_{n=1}^\infty \mathbb{E}_{\theta_0}\Phi_n < \infty$.
2. $(\forall \theta \in U_n^\epsilon \cap \Theta_\gamma^n)$ $\mathbb{E}_\theta(1 - \Phi_n) \le C_1 e^{-c_1 n}$.

By Theorem 1 of (Choi & Schervish, 2007), these conditions imply that

$$\lim_{n\to\infty} \Pi\big(U_n^\epsilon \cap \Theta_\gamma^n | Y_1, Y_2, \cdots, Y_n\big) = 0, \quad \text{in } \mathbb{P}_{\theta_0} \text{ probability.}$$

2.1 Positive prior probability

First we consider Condition I. Fix any $\delta_0 > 0$ and set $B \equiv B_{\delta_0}(\theta_0)$. Then for any $\theta \in B$,

$$V_i(\theta_0, \theta) = \theta_0(X_i) \left[\log \frac{\theta_0(X_i)}{\theta(X_i)}\right]^2 \leq \|\theta_0\|_* \left[\log \frac{\|\theta_0\|_* + \delta_0}{\gamma}\right]^2,$$

uniformly in $i \in \mathbb{N}$, so $\sum_{i=1}^{\infty} V_i(\theta_0, \theta)/i^2 < \infty$, for all $\theta \in B$ and I.1 holds.

Notice that

$$\begin{aligned} K_i(\theta_0, \theta) &= \theta(X_i) - \theta_0(X_i) + \theta_0(X_i) \log \frac{\theta_0(X_i)}{\theta(X_i)} \\ &= \theta(X_i)\left(1 - r + r \log r\right), \end{aligned}$$

where $r \equiv \frac{\theta_0(X_i)}{\theta(X_i)}$. The inequality $\log r \leq r - 1$ implies that

$$\begin{aligned} K_i(\theta_0, \theta) &\leq \theta(X_i)\big(1 - r + r(r-1)\big) = \theta(X_i)(r-1)^2 \\ &= \frac{\big(\theta(X_i) - \theta_0(X_i)\big)^2}{\theta(X_i)} \leq \frac{1}{\gamma}\|\theta - \theta_0\|_*^2. \end{aligned}$$

Since $\Pi\big(B_\delta(\theta_0)\big) > 0$ for each $\delta > 0$ by hypothesis, and in particular for $\delta = \big(\delta_0 \wedge \sqrt{\gamma\epsilon}\big)$, condition I.2 is verified since $B_\delta(\theta_0) \subset B \cap \{\theta : K_i(\theta_0, \theta) < \epsilon\}$.

2.2 Existence of tests

Now we turn to the second condition. First we construct tests for the simple *vs.* simple hypotheses:

$$H_0 : Y_i \sim \mathsf{Po}\big(\theta_0(X_i)\big) \;\textit{vs.}\; H_1 : Y_i \sim \mathsf{Po}\big(\theta_1(X_i)\big).$$

Below we construct test functions for testing a more general class of hypotheses, including those involving stochastically ordered positive random variables. For an integrable positive random variable Y with mean μ and an integer $n \in \mathbb{N}$, denote by $\mu^{(n)}$ the mean of the truncated random variable

$$\mu^{(n)} = \mathbb{E}\big[Y \wedge n\big]$$

where $\mathbf{1}_{\{A\}}$ represents the indicator function for the event A. For a Poisson-distributed random variable this is $\mu^{(n)} = n - [\Gamma(n+1, \mu) - \mu\Gamma(n, \mu)]/\Gamma(n) \leq (\mu \wedge n)$, where $\Gamma(a, x)$ denotes the incomplete Gamma function (Abramowitz & Stegun, 1974, § 6.5.3). In particular, we define the "truncated mean" of any $\theta \in \Theta_\gamma$ to be

$$\theta^{(n)}(X_i) \equiv \mathbb{E}_\theta\big[Y_i \wedge n\big].$$

Lemma 2.2. *Let $\{Y_i\}$ be independent positive random variables with means $\{\mu_i\}$, and let $\{X_i\} \subset \mathcal{X}$. Consider the following hypothesis testing problem:*

$$H_0 : \mu_i = \theta_0(X_i) \;\textit{vs.}\; H_1 : \mu_i = \theta_1(X_i)$$

for fixed elements $\theta_0, \theta_1 \in \Theta_\gamma$ that satisfy the inequality

$$\theta_1(x) > \theta_0(x) + \epsilon$$

for some number $\epsilon > 0$ and all $x \in \mathcal{X}$. Consider the sequence of indicator random variables,

$$\Phi_n \equiv \mathbf{1}_{\{A_n\}}, \qquad A_n \equiv \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \left[(Y_i \wedge n) - \theta_0^{(n)}(X_i) \right] > \epsilon \right\}.$$

If the distributions of $\{Y_i\}$ are from a location family with location parameters μ_i, or they are absolutely continuous with respect to some reference measure and have density functions $\{p_{\mu_i}(y)\}$ with a monotone likelihood ratio (MLR), then for all sufficiently large n,

$$\mathbb{E}_{\theta_0}(\Phi_n) \leq \exp(-n\epsilon^2/2) \tag{5a}$$

$$\mathbb{E}_{\theta_1}(1 - \Phi_n) \leq \exp(-n\epsilon^2/2). \tag{5b}$$

Proof. The proof is deferred to the Appendix. ■

Remark 2.3. *In the above lemma if we had, $\theta_0(x) > \theta_1(x) + \epsilon$, with all the other conditions being the same, then the test*

$$\Phi_n = \mathbf{1}_{\{B_n\}}, \; B_n \equiv \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i) \right] < \epsilon \right\}$$

will satisfy (5), for all sufficiently large n (see Appendix for the details of the proof).

Since the Poisson family has a monotone likelihood ratio, the above Lemma certainly applies to the case of testing $H_0 : Y_i \sim \mathsf{Po}\big(\theta_0(X_i)\big)$ *vs.* $H_1 : Y_i \sim \mathsf{Po}\big(\theta_j(X_i)\big)$ for any fixed $\theta_j \in \Theta_\gamma$. The rest of our proof is based on a standard argument using covering numbers (Ghosal & Roy, 2006). First notice that for any $\theta_j \in \Theta_\gamma$ with $\int |\theta_j(x) - \theta_0(x)| Q_n(dx) > \epsilon$,

$$Q_n\{x : |\theta_j(x) - \theta_0(x)| > \epsilon/2\} \geq \frac{\epsilon}{2(\|\theta_0\|_* + \|\theta_j\|_*)}.$$

This can be seen from:

$$\begin{aligned} \epsilon &< \int |\theta_j(x) - \theta_0(x)| \, Q_n(dx) \\ &< \big(\|\theta_0\|_* + \|\theta_j\|_*\big) \, Q_n\{x : |\theta_j(x) - \theta_0(x)| > \epsilon/2\} + \epsilon/2. \end{aligned}$$

Hence the cardinality $\#\{i : |\theta_j(X_i) - \theta_0(X_i)| > \epsilon/2\} \geq \tilde{C} n$ for the constant $\tilde{C} \equiv \epsilon(2\|\theta_0\|_* + 2\|\theta_j\|_*)^{-1}$. This implies that one of the two sets $\{i : \theta_j(X_i) > \theta_0(X_i) + \epsilon/2\}$ or $\{i : \theta_0(X_i) > \theta_j(X_i) + \epsilon/2\}$ will have at least $\tilde{C} n/2$ points. For definiteness we will assume it is the former and set

$$S_{\theta_j}^{+} \equiv \{x : \theta_j(x) > \theta_0(x) + \epsilon/2\} \tag{6}$$

(otherwise see Remark 2.3, and make the corresponding change in equation (7) below). Now for testing θ_0 *vs.* θ_j such that $\int |\theta_j(x) - \theta_0(x)|\, Q_n(dx) > \epsilon$, we construct the test:

$$\Psi_{nj} = \mathbf{1}_{\{E_{nj}\}},\ E_{nj} \equiv \left\{ \frac{1}{n^2} \sum_{i=1}^{n} \big[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\big] \mathbf{1}_{\{X_i \in S^+_{\theta_j}\}} > \epsilon/2 \right\}. \tag{7}$$

Let $N \equiv N\big(\epsilon/2, \Theta^n_\gamma, \|\cdot\|_*\big)$ denote the covering number of the set Θ^n_γ under the uniform norm, *i.e.*, the smallest number N such that

$$\Theta^n_\gamma \subset \bigcup_{j=1}^{N} B_{\epsilon/2}(\theta_j)$$

for some $\{\theta_1, \theta_2, \cdots, \theta_N\} \subset \Theta^n_\gamma$. Then by Theorem 2.7.1 of van der Vaart & Wellner (1996),

$$\log N \le K_1 M_n^{d/\alpha} \epsilon^{-1/\alpha} \tag{8}$$

for some constant $K_1 < \infty$. For each $\theta \in \Theta^n_\gamma$ there is some $j \le N$ such that $\|\theta - \theta_j\|_* \le \epsilon/2$. Recall the tests $\{\Psi_{nj}\}_{j=1}^N$, from (7). By Lemma 2.2 (with $\epsilon/2$ replacing ϵ),

$$\begin{aligned} \mathbb{E}_{\theta_0}(\Psi_{nj}) &\le \exp(-n\epsilon^2/8), \\ \mathbb{E}_{\theta_j}(1 - \Psi_{nj}) &\le \exp(-n\epsilon^2/8), \end{aligned} \qquad 1 \le j \le N.$$

Now set $\Psi_n \equiv \max\{\Psi_{n,\theta_j} :\ 1 \le j \le N\}$. Then,

$$\mathbb{E}_{\theta_0}(\Psi_n) \le \sum_{j=1}^{N} \mathbb{E}_{\theta_j}(\Psi_{nj}) \le \exp(\log N - n\epsilon^2/8). \tag{9}$$

If we set $M_n \equiv K_2 n^{\alpha/d}$ (see equation (4)) for sufficiently small K_2, then equations (8, 9) imply

$$\mathbb{E}_{\theta_0}(\Psi_n) \le \exp(-K_3 n)$$

for some constant $K_3 > 0$. In particular, $\sum_{n=1}^{\infty} \mathbb{E}_{\theta_0}(\Psi_n) < \infty$, verifying II.1. Now we turn to II.2.

For any $\theta \in U_n^\epsilon \cap \Theta^n_\gamma$, there is at least one $j \in \{1, 2, \cdots, N\}$ such that $\|\theta - \theta_j\|_* < \epsilon/2$. For $X_i \in S^+_{\theta_j}$ (see equation (6)), by the triangle inequality, $\big[\theta(X_i) - \theta_0(X_i)\big] \ge \epsilon/2$. Hence

$$\mathbb{E}_\theta(1 - \Psi_n) \le \mathbb{E}_\theta(1 - \Psi_{nj}) \le \exp(-n\epsilon^2/8).$$

Since this holds uniformly for all $\theta \in U_n^\epsilon \cap \Theta^n_\gamma$,

$$\sup_{\theta \in U_n^\epsilon \cap \Theta^n_\gamma} \mathbb{E}_\theta(1 - \Psi_n) \le \exp(-n\epsilon^2/8)$$

verifying II.2. Hence we have proved the theorem. ∎

Remark 2.4. *If the prior* Π *also satisfies the condition* $\Pi(\Theta_\gamma^{nc}) \le C_2 e^{-c_2 n}$ *for constants* $C_2, c_2 > 0$*, then the conclusion of Theorem 2.1 can be strengthened to:*

$$\lim_{n\to\infty} \Pi\left\{ \left[\theta \in \Theta_\gamma : \int_{\mathcal{X}} |\theta(x) - \theta_0(x)|\, dQ_n(x) > \epsilon \right] \Big| Y_1, Y_2, \cdots, Y_n \right\} = 0$$

in $\mathbb{P}_{\theta_0}$ *probability.*

Remark 2.5. *The predictors* $\{X_i\}$*, instead of being fixed, can be sampled according to a distribution* D *on* $\mathcal{X}$*. With similar arguments as above, we can show that consistency holds in* L_1 *norm, instead of the empirical* L_1 *norm. See (Ghosal & Roy, 2006) for a result of this type.*

3 An Example

Set $\mathcal{X} \equiv [-1,1]$. Fix $\alpha, \gamma > 0$ and set $\Theta_\gamma \equiv \{\theta \in C^\alpha(\mathcal{X}) : \theta(x) \ge \gamma\}$. Let $\Phi \equiv \{\phi_k(x),\ k \in \mathbb{N},\ x \in \mathcal{X}\}$ be the Legendre polynomials $\phi_k(x) \equiv P_k(x) = \frac{1}{2^k k!}\frac{d^k}{dx^k}\left[(x^2-1)^k\right]$ (see Abramowitz & Stegun, 1974, §8); note these are continuous on $\mathcal{X}$ and, since they are a complete orthonormal set in $L_2(\mathcal{X}, dx)$, their linear span is dense in Θ_γ in the uniform norm. Fix $\lambda > 0$ and $\sigma^2 > 0$, and consider the prior distribution $\Pi(d\theta)$ on Θ_γ defined implicitly by:

$$J \sim \mathsf{Po}(\lambda), \quad \{\beta_j\}_{j=1}^J \overset{iid}{\sim} \mathsf{No}(0, \sigma^2)$$
$$\theta(x) \equiv \sum_{j=1}^J \beta_j \phi_j(x).$$

Lemma 3.1. *The prior* $\Pi(d\theta)$ *defined above satisfies* $\Pi\big(B_\delta(\theta_0)\big) > 0$ *for any* $\theta_0 \in \Theta_\gamma$ *and* $\delta > 0$*. Hence the conclusion of Theorem 2.1 holds true.*

Proof. Fix $\delta > 0$ and $\theta_0 \in \Theta_\gamma$. By the density in Θ_γ of the linear span of Φ, there exist a $J^* \in \mathbb{N}$ and $\{\beta_j^*\}_{j=1}^{J^*} \in \mathbb{R}$ such that

$$\|\theta_0(\mathcal{X}) - \sum_{j=1}^{J^*} \beta_j^* \phi_j(x)\|_* < \delta/2. \tag{10}$$

Set $\kappa \equiv \max_{j \le J^*} \|\phi_j\|_*$. By the triangle inequality and (10),

$$B_\delta(\theta_0) \supset B^* \equiv \left\{ \sum_{j=1}^J \beta_j \phi_j(x) : \qquad J = J^*, \quad \sum_{j=1}^{J^*} |\beta_j - \beta_j^*| \le \delta/2\kappa \right\}.$$

The $\Pi(d\theta)$ probability of $B_\delta(\theta_0)$ must exceed

$$\Pi\big(B^*\big) = e^{-\lambda}\left(\lambda/\sqrt{2\pi\sigma^2}\right)^{J^*} \int_B \exp(-|\beta|^2/2\sigma^2) d\beta > 0$$

where $B \equiv \left\{ \{\beta_j\}_{j=1}^{J^*} : \sum_{j=1}^{J^*} |\beta_j - \beta_j^*| \le \delta/2\kappa \right\} \subset \mathbb{R}^{J^*}$. Hence $\Pi\big(B_\delta(\theta_0)\big) > 0$ and the lemma is proved. ∎

Remark 3.2. *This example shows illustrates a simple sufficient condition for consistency to hold in Poisson regression models based on series expansions, such as those in (3):*

1. *The linear span of the basis functions is dense in* Θ_γ *with the uniform (or any stronger) topology (so, for example, the Legendre polynomials could be replaced by splines, wavelets, radial basis functions, etc.);*

2. *The support of the prior distribution for the basis coefficients is all of* $\mathbb{R}$ *(actually, it's enough for the additive group generated by the support to be dense in* $\mathbb{R}$*).*

4 Discussion

We have proved posterior consistency of nonparametric Poisson regression models, under a modest assumption on the prior: that it put positive mass on all uniform balls in the parameter space. This condition, as demonstrated in the example, is easy to verify for most priors, especially those based on series expansions or splines. We introduced a truncation approach to construct tests that are applicable here and to the more general setting of positive stochastically-ordered random variables. We expect that our approach may be extended to nonparametric regression models in other exponential families.

It is interesting to note that we require the parameter space Θ_γ to be bounded away from zero. This is not a mere artifact of our proof: if (nearly) all observations in a sample from a Poisson distribution are zero, then it impossible to estimate the (log) mean without some condition such as a bound away from zero. We expect that our approach may be extended to obtain bounds on convergence rates for nonparametric Poisson regression models.

Appendix

Proof of Lemma 2.2. Notice that

$$\mathbb{E}_{\theta_0}(\Phi_n) = \mathbb{P}_{\theta_0}\left\{\sum_{i=1}^{n}\big[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\big] > n^2\epsilon\right\}.$$

Since $\mathbb{E}_{\theta_0}\big[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\big] = 0$ and $-n \leq \big[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\big] \leq n$, by Hoeffding's inequality (Hoeffding, 1963, Theorem 1)

$$\mathbb{P}_{\theta_0}\left\{\sum_{i=1}^{n}\big[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\big] > n^2\epsilon\right\} \leq \exp\big(-2n^4\epsilon^2/n(2n)^2\big) = \exp(-n\epsilon^2/2).$$

To obtain the type II error rate of equation (5b), first note

$$\begin{aligned}
\mathbb{E}_{\theta_1}(1-\Phi_n) &= \mathbb{P}_{\theta_1}\left\{\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right] \le n^2\epsilon\right\} \\
&= \mathbb{P}_{\theta_1}\left\{\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] + \sum_{i=1}^n \left[\theta_1^{(n)}(X_i) - \theta_0^{(n)}(X_i)\right] \le n^2\epsilon\right\} \\
&= \mathbb{P}_{\theta_1}\left\{-\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right]\right. \\
&\qquad\qquad \left. \ge -n^2\epsilon + \sum_{i=1}^n \left[\theta_1^{(n)}(X_i) - \theta_0^{(n)}(X_i)\right]\right\}. \qquad (11)
\end{aligned}$$

Since $\theta_0(X_i) + \epsilon < \theta_1(X_i)$ for all $i \in \mathbb{N}$, under the (stated) hypothesis that θ_i is the location parameter for the distribution of Y_i, or that the family of p.d.f.'s have a Monotone Likelihood Ratio in θ_i, we have $\mathbb{P}_{\theta_0(X_i)}(Y_i \le c) \ge \mathbb{P}_{\theta_1(X_i)}(Y_i \le c)$ for each $c > 0$ and, since the function $f(y) \equiv (y \wedge n)$ is nondecreasing,

$$\theta_0^{(n)}(X_i) \equiv \mathbb{E}_{\theta_0(X_i)}\left[(Y_i \wedge n)\right] \le \mathbb{E}_{\theta_1(X_i)}\left[(Y_i \wedge n)\right] \equiv \theta_1^{(n)}(X_i) \qquad (12)$$

for all $n \in \mathbb{N}$ and all $i \in \mathbb{N}$. Thus each $\left[\theta_1^{(n)}(X_i) - \theta_0^{(n)}(X_i)\right] \ge 0$ is positive, so the inequality (11) can be bounded by

$$\begin{aligned}
\mathbb{E}_{\theta_1}(1-\Phi_n) &\le \mathbb{P}_{\theta_1}\left\{-\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] \ge -n^2\epsilon\right\} \\
&\le \exp(-n\epsilon^2/2) \qquad (13)
\end{aligned}$$

by another application of Hoeffding's inequality.

Proof of Remark 2.3. In this case the type II error rate follows from the following simple application of Hoeffding's inequality:

$$\begin{aligned}
\mathbb{E}_{\theta_1}(1-\Phi_n) &= \mathbb{P}_{\theta_1}\left\{\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] > n^2\epsilon\right\} \\
&\le \exp(-2n^4\epsilon^2/n(2n)^2) = \exp(-n\epsilon^2/2).
\end{aligned}$$

For the type I error,

$$\begin{aligned}
\mathbb{E}_{\theta_0}(\Phi_n) &= \mathbb{P}_{\theta_0}\left\{\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_1^{(n)}(X_i)\right] \le n^2\epsilon\right\} \\
&= \mathbb{P}_{\theta_0}\left\{-\sum_{i=1}^n \left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right]\right. \\
&\qquad\qquad \left. \ge -n^2\epsilon + \sum_{i=1}^n \left[\theta_0^{(n)}(X_i) - \theta_1^{(n)}(X_i)\right]\right\}. \qquad (14)
\end{aligned}$$

Once again the differences $\left[\theta_0^{(n)}(X_i) - \theta_1^{(n)}(X_i)\right]$ will be positive, so equation (14) can be bounded above by

$$\mathbb{E}_{\theta_0}(\Phi_n) \leq \mathbb{P}_{\theta_0}\left\{-\sum_{i=1}^{n}\left[(Y_i \wedge n) - \theta_0^{(n)}(X_i)\right] \geq -n^2\epsilon\right\} \leq \exp(-n\epsilon^2/2)$$

by another application of Hoeffding's inequality. ■

Acknowledgments

The authors would like to thank Jayanta K. Ghosh and Subhashis Ghoshal for their encouragement and helpful conversations. This work was supported in part by the National Science Foundation under Grant Numbers DMS–0112069, DMS–0342172, DMS–0422400 and DMS–0406115. Any opinions, findings, and conclusions or recommendations expressed in this material are those of the authors and do not necessarily reflect the views of the National Science Foundation.

References

Abramowitz, M. & Stegun, I. A. (Eds.) (1974). *Handbook of Mathematical Functions With Formulas, Graphs, and Mathematical Tables.* New York, NY: Dover.

Amewou-Atisso, M., Ghosal, S., Ghosh, J. K. & Ramamoorthi, R. V. (2003). Posterior consistency for semi-parametric regression problems. *Bernoulli* **9**, 291–312.

Choi, T. & Schervish, M. J. (2007). On posterior consistency in nonparametric regression problems. *J. Multivariate Anal.* **0**, in press.

Ghosal, S. & Roy, A. (2006). Posterior consistency of Gaussian process prior for nonparametric binary regression. *Ann. Stat.* **34**, 2413–2429.

Ghosal, S. & van der Vaart, A. W. (2007a). Convergence rates of posterior distributions for noniid observations. *Ann. Stat.* **35**, 192–223.

Ghosal, S. & van der Vaart, A. W. (2007b). Posterior convergence rates of Dirichlet mixtures at smooth densities. *Ann. Stat.* **35**, in press.

Hoeffding, W. (1963). Probability inequalities for sums of bounded random variables. *J. Am. Stat. Assoc.* **58**, 13–30.

van der Vaart, A. W. & Wellner, J. A. (1996). *Weak Convergence and Empirical Processes.* Springer Series in Statistics. New York, NY: Springer-Verlag.