# Errata for

Biau, G. (2012). Analysis of a random forests model,
*Journal of Machine Learning Research*, **Vol. 13**, pp. 1063–1095

June 2, 2018

The following corrections have been pointed out by Qiannan Gao, Markus Pauly, and Jason Klusowski.

- In Proposition 2, Theorem 5, and Corollary 6, the constant $C$ should be
$$C = \frac{576}{\pi}\left(\frac{\pi\log 2}{16}\right)^{S/2d}.$$
- Modify the end of the proof of Proposition 2 from line 9 of page 1085 as follows: Clearly,
$$\begin{aligned}\lambda\left(A_{nj}(\mathbf{X},\Theta)\cap A_{nj}(\mathbf{X},\Theta')\right) &\leq 2^{-\max(K_{nj},K'_{nj})}\\ &= 2^{-(K_{nj}+K'_{nj})/2}2^{-|K_{nj}-K'_{nj}|/2}\end{aligned}$$
and, consequently,
$$\prod_{j=1}^{d}\lambda\left(A_{nj}(\mathbf{X},\Theta)\cap A_{nj}(\mathbf{X},\Theta')\right) \leq 2^{-\lceil\log_2 k_n\rceil}\prod_{j=1}^{d}2^{-|K_{nj}-K'_{nj}|/2}$$
(since, by Fact 1, $\sum_{j=1}^{d}K_{nj} = \sum_{j=1}^{d}K'_{nj} = \lceil\log_2 k_n\rceil$). Plugging this inequality into (3) and applying Hölder's inequality, we obtain
$$\begin{aligned}\mathbb{E}\left[\bar{r}_n(\mathbf{X})-\tilde{r}_n(\mathbf{X})\right]^2 &\leq \frac{12\sigma^2 k_n}{n}\mathbb{E}\left[\prod_{j=1}^{d}2^{-|K_{nj}-K'_{nj}|/2}\right]\\ &= \frac{12\sigma^2 k_n}{n}\mathbb{E}\left[\mathbb{E}\left[\prod_{j=1}^{d}2^{-|K_{nj}-K'_{nj}|/2}\mid\mathbf{X}\right]\right]\\ &\leq \frac{12\sigma^2 k_n}{n}\mathbb{E}\left[\prod_{j=1}^{d}\mathbb{E}^{1/d}\left[2^{-d|K_{nj}-K'_{nj}|/2}\mid\mathbf{X}\right]\right]\\ &\leq \frac{12\sigma^2 k_n}{n}\mathbb{E}\left[\prod_{j=1}^{d}\mathbb{E}^{1/d}\left[2^{-|K_{nj}-K'_{nj}|}\mid\mathbf{X}\right]\right]\\ &\quad\text{(since } d\geq 2\text{)}.\end{aligned}$$

Therefore,

$$\mathbb{E}\left[\bar{r}_n(\mathbf{X}) - \tilde{r}_n(\mathbf{X})\right]^2 \le \frac{24\sigma^2 k_n}{n}\mathbb{E}\left[\prod_{j=1}^{d}\mathbb{E}^{1/d}\left[2^{-(K_{nj}-K'_{nj})}\mathbf{1}_{[K_{nj}\ge K'_{nj}]}\,|\,\mathbf{X}\right]\right].$$

Each term in the product may be bounded by technical Proposition 13, and this leads to

$$\begin{aligned}\mathbb{E}\left[\bar{r}_n(\mathbf{X}) - \tilde{r}_n(\mathbf{X})\right]^2 &\le \frac{576\sigma^2 k_n}{\pi n}\prod_{j=1}^{d}\min\left(1, \left[\frac{\pi}{16\lceil\log_2 k_n\rceil p_{nj}(1-p_{nj})}\right]^{1/2d}\right)\\ &\le \frac{576\sigma^2 k_n}{\pi n}\prod_{j=1}^{d}\min\left(1, \left[\frac{\pi\log 2}{16(\log k_n)p_{nj}(1-p_{nj})}\right]^{1/2d}\right).\end{aligned}$$

Using the assumption on the form of the $p_{nj}$, we finally conclude that

$$\mathbb{E}\left[\bar{r}_n(\mathbf{X}) - \tilde{r}_n(\mathbf{X})\right]^2 \le C\sigma^2\left(\frac{S^2}{S-1}\right)^{S/2d}(1+\xi_n)\frac{k_n}{n(\log k_n)^{S/2d}},$$

where

$$C = \frac{576}{\pi}\left(\frac{\pi\log 2}{16}\right)^{S/2d}$$

and

$$1+\xi_n = \prod_{j\in\mathcal{S}}\left[(1+\xi_{nj})^{-1}\left(1-\frac{\xi_{nj}}{S-1}\right)^{-1}\right]^{1/2d}.$$

Clearly, the sequence $(\xi_n)$, which depends on the $\{(\xi_{nj}) : j \in \mathcal{S}\}$ only, tends to 0 as $n$ tends to infinity.

- Statement $(iii)$ of Lemma 12 page 1091 is incorrect and should be: For all $d \ge 1$,

$$\mathbb{E}\left[2^{-d(Z_1-Z_2)}\mathbf{1}_{[Z_1\ge Z_2]}\right] \le \frac{24}{\pi}\int_0^1 \exp\left(-4Np(1-p)t^2\right)\mathrm{d}t.$$

  The first two lines of the proof should be modified accordingly, and, everywhere, $\mathbb{E}\left[2^{-d(Z_1-Z_2)_+}\right]$ should be replaced by $\mathbb{E}\left[2^{-d(Z_1-Z_2)}\mathbf{1}_{[Z_1\ge Z_2]}\right]$.

- In Proposition 13, replace $\mathbb{E}\left[2^{-d(Z_1-Z_2)_+}\right]$ by $\mathbb{E}\left[2^{-d(Z_1-Z_2)}\mathbf{1}_{[Z_1\ge Z_2]}\right]$.