Corrections to "Generalization Bounds via Information Density and Conditional Information Density"

Fredrik Hellström, *Graduate Student Member, IEEE*, and Giuseppe Durisi, *Senior Member, IEEE*

***Abstract*—An error in the proof of the data-dependent tail bounds on the generalization error presented in Hellström and Durisi (2020) is identified, and a correction is proposed. Furthermore, we note that the absolute continuity requirements in Hellström and Durisi (2020) need to be strengthened to avoid measurability issues.**

I. Data-Dependent Bounds in [1, eqs. (26), (34), (95), and (98)]

In the proof of [1, eq. (26)], we incorrectly claimed that [1, eq. (32)] implies [1, eq. (26)]. The issue is that [1, eq. (32)] holds for a *fixed* λ, whereas, for [1, eq. (26)] to hold, [1, eq. (32)] needs to hold uniformly over all $\lambda \in \mathbb{R}$.

This issue can be fixed as follows. Since $\mathrm{gen}(w, \boldsymbol{Z})$ is $\sigma/\sqrt{n}$-sub-Gaussian with zero mean under $P_{\boldsymbol{Z}}$ for all w, we can apply [2, Th. 2.6.(IV)] (with $\lambda = 1 - 1/n$ therein) to conclude that

$$\mathbb{E}_{P_{\boldsymbol{Z}}}\left[\exp\left(\frac{n-1}{2\sigma^2}(\mathrm{gen}(w,\boldsymbol{Z}))^2\right)\right] \le \sqrt{n}. \tag{1}$$

Taking the expectation with respect to P_W, changing measure to $P_{W\boldsymbol{Z}}$, and rearranging terms, we obtain

$$\mathbb{E}_{P_{W\boldsymbol{Z}}}\left[\exp\left(\frac{n-1}{2\sigma^2}(\mathrm{gen}(W,\boldsymbol{Z}))^2 - \log\sqrt{n} - \imath(W,\boldsymbol{Z})\right)\right] \le 1. \tag{2}$$

Proceeding as in [1, Corollary 2], with an additional use of Jensen's inequality, we find that with probability at least $1-\delta$ under $P_{\boldsymbol{Z}}$,

$$\begin{aligned}&\left|\mathbb{E}_{P_{W|\boldsymbol{Z}}}[\mathrm{gen}(W,\boldsymbol{Z})]\right| \\ &\le \sqrt{\frac{2\sigma^2}{n-1}\left(D(P_{W|\boldsymbol{Z}}\,||\,P_W) + \log\frac{\sqrt{n}}{\delta}\right)}.\end{aligned} \tag{3}$$

Similarly, proceeding as in the proof of [1, eq. (34)], we find that with probability at least $1-\delta$ under $P_{W\widetilde{\boldsymbol{Z}}\boldsymbol{S}}$,

$$|\mathrm{gen}(W,\boldsymbol{Z})| \le \sqrt{\frac{2\sigma^2}{n-1}\left(\imath(W,\boldsymbol{Z}) + \log\frac{\sqrt{n}}{\delta}\right)}. \tag{4}$$

The issue reported in this note also affects the data-dependent tail bounds for the random-subset setting reported in [1, eqs. (95) and (98)]. To fix it, we use that for any fixed $(w, \widetilde{z})$, the random variable $\widehat{\mathrm{gen}}(w, \widetilde{z}, \boldsymbol{S})$ is $1/\sqrt{n}$-sub-Gaussian with zero mean under $P_{\boldsymbol{S}}$. Applying [2, Th. 2.6.(IV)] with $\lambda = 1 - 1/n$ we obtain

$$\mathbb{E}_{P_{\boldsymbol{S}}}\left[\exp\left(\frac{n-1}{2}(\widehat{\mathrm{gen}}(w,\widetilde{z},\boldsymbol{S}))^2\right)\right] \le \sqrt{n}. \tag{5}$$

Taking the expectation with respect to $P_{W\widetilde{\boldsymbol{Z}}}$, changing measure to $P_{W\widetilde{\boldsymbol{Z}}\boldsymbol{S}}$, and rearranging terms, we conclude that

$$\begin{aligned}&\mathbb{E}_{P_{W\widetilde{\boldsymbol{Z}}\boldsymbol{S}}}\left[\exp\left(\frac{n-1}{2}(\widehat{\mathrm{gen}}(W,\widetilde{\boldsymbol{Z}},\boldsymbol{S}))^2 - \log\sqrt{n} - \imath(W,\boldsymbol{S}|\widetilde{\boldsymbol{Z}})\right)\right] \\ &\le 1.\end{aligned} \tag{6}$$

Proceeding as in [1, Corollary 6], we conclude that with probability at least $1-\delta$ under $P_{\widetilde{\boldsymbol{Z}}\boldsymbol{S}}$,

$$\begin{aligned}&\mathbb{E}_{P_{W|\widetilde{\boldsymbol{Z}}\boldsymbol{S}}}[\widehat{\mathrm{gen}}(W,\widetilde{\boldsymbol{Z}},\boldsymbol{S})] \\ &\le \sqrt{\frac{2}{n-1}\left(D(P_{W|\widetilde{\boldsymbol{Z}}\boldsymbol{S}}\,||\,P_{W|\widetilde{\boldsymbol{Z}}}) + \log\frac{\sqrt{n}}{\delta}\right)}.\end{aligned} \tag{7}$$

Furthermore, with probability at least $1-\delta$ under $P_{W\widetilde{\boldsymbol{Z}}\boldsymbol{S}}$,

$$\left|\widehat{\mathrm{gen}}(W,\widetilde{\boldsymbol{Z}},\boldsymbol{S})\right| \le \sqrt{\frac{2}{n-1}\left(\imath(W,\boldsymbol{S}|\widetilde{\boldsymbol{Z}}) + \log\frac{\sqrt{n}}{\delta}\right)}. \tag{8}$$

To summarize, the data-dependent tail bounds reported in [1, eqs. (26), (34), (95), and (98)] should be replaced with (3), (4), (7), and (8) respectively.

Note that the data-independent tail bounds that we provide in [1, eqs. (27), (35), (41), (42), (96), (99), (101), and (102)] still hold verbatim, although their proofs need to be modified. Specifically, for a fixed λ, one needs to first replace the information measure appearing in the bounds with its data-independent relaxation. The desired bounds then follow by setting λ equal to a suitably chosen, data-independent constant. Consider for example the data-independent bound in [1, eq. (27)]. To obtain it, we first use [1, eq. (33)] in [1, eq. (32)], which results in

$$\begin{aligned}P_{\boldsymbol{Z}}\Bigg[&\frac{\lambda^2\sigma^2}{2n} - \lambda\,\mathbb{E}_{P_{W|\boldsymbol{Z}}}[\mathrm{gen}(W,\boldsymbol{Z})] + \frac{\mathbb{E}_{P_{\boldsymbol{Z}}}^{1/t}[D(P_{W|\boldsymbol{Z}}\,||\,P_W)^t]}{\delta^{1/t}} \\ &+ \log\frac{1}{\delta} \ge 0\Bigg] \ge 1 - 2\delta.\end{aligned} \tag{9}$$

Manuscript received June 8, 2021; accepted June 8, 2021. Date of current version September 20, 2021. *(Corresponding author: Fredrik Hellström.)*

The authors are with the Department of Electrical Engineering, Chalmers University of Technology, 412 96 Gothenburg, Sweden (e-mail: frehells@chalmers.se; durisi@chalmers.se).

Digital Object Identifier 10.1109/JSAIT.2021.3088240

The desired result follows by setting $\lambda = \pm\sqrt{\frac{a}{b}}$, where $a = \mathbb{E}_{P_{\mathbf{Z}}}^{1/t}[D(P_{W\,|\,\mathbf{Z}}\,||\,P_W)^t]/\delta^{1/t} + \log\frac{1}{\delta}$ and $b = \sigma^2/(2n)$, and then replacing δ with $\delta/2$.

II. Absolute Continuity Assumption

In the statement of [1, Th. 1], we assumed that $P_{W\mathbf{Z}} \ll P_W P_{\mathbf{Z}}$. To avoid measurability issues, we should also assume that $P_W P_{\mathbf{Z}} \ll P_{W\mathbf{Z}}$. Similarly, in [1, Th. 4], we should also assume that $P_{W|\widetilde{\mathbf{Z}}} P_{\widetilde{\mathbf{Z}}} P_{\mathbf{S}} \ll P_{W\widetilde{\mathbf{Z}}\mathbf{S}}$.

References

[1] F. Hellström and G. Durisi, "Generalization bounds via information density and conditional information density," *IEEE J. Sel. Areas Inf. Theory*, vol. 1, no. 3, pp. 824–839, Nov. 2020.

[2] M. J. Wainwright, *High-Dimensional Statistics: A Non-Asymptotic Viewpoint*. Cambridge, U.K.: Cambridge Univ. Press, 2019.