A localization framework for singular learning theory

Brennan Dury

Abstract

We develop a framework for singular learning theory with several analytic losses on a possibly non-compact parameter space. Additionally, we support loss functions beyond KL divergence and do not assume relatively finite variance. Using this framework, we show that suitably weighted auxiliary losses localize the Gibbs posterior near a common zero set while preserving the leading learning coefficient of a distinguished loss.

1 Setup and main theorem

1.1 Framework

Let $\Omega \subseteq \mathbb{R}^d$ be connected and open, and let $W \subseteq \Omega$ be a closed non-empty semianalytic parameter set such that $W = \overline{W^\circ}$. On Ω fix a measure defined by

$$\mathrm{d}\mu(w) = \varphi_{\mathrm{an}}(w)\varphi_{\mathrm{sm}}(w)\,\mathrm{d}w, \qquad \varphi_{\mathrm{an}} \geq 0 \text{ analytic and not identically zero}, \quad \varphi_{\mathrm{sm}} > 0 \text{ smooth.} \tag{1}$$

We assume that $\mu(W) = 1$, and we regard $\mu|_W$ as a Bayesian prior on W. Let $(X, \mathcal{A})$ be a measurable observation space with probability measure ν. For each $j \in \{1, \ldots, \ell\}$, let $f_j : X \times \Omega \to \mathbb{R}$ be a jointly measurable loss such that $f_j(\cdot, w) \in L^2(X, \nu)$ and the map

$$\Phi_j : \Omega \longrightarrow L^2(X, \nu), \qquad \Phi_j(w) := f_j(\cdot, w), \tag{2}$$

is real analytic. Define the expected loss by

$$K_j(w) = \mathbb{E}_\nu f_j(X, w). \tag{3}$$

We assume $K_j \geq 0$ on W and that its minimum is zero and attained. The functions K_j are real analytic on Ω, since expectation is bounded on $L^2(X, \nu)$.

Example 1.1. Let $\{P_w : w \in \Omega\}$ be a family of model probability measures. Assume $\nu \ll P_w$ and write

$$r(x, w) = \frac{d\nu}{dP_w}(x).$$

Suppose $\log r(\cdot, w) \in L^2(X, \nu)$ for every $w \in \Omega$ and the map $\Phi(w) := \log r(\cdot, w)$ is real analytic into $L^2(X, \nu)$. Then $\mathrm{KL}(\nu \| P_w)$ is finite. Choose $w^0 \in W$ minimizing this divergence on W, and define the centered loss $f : X \times \Omega \to \mathbb{R}$ by

$$f(x, w) = \log \frac{r(x, w)}{r(x, w^0)}.$$

The map $\Phi_c(w) := f(\cdot, w)$ is real analytic, given by $\Phi(w) - \Phi(w^0)$. Its excess expected loss is

$$K(w) = \mathrm{KL}(\nu \| P_w) - \mathrm{KL}(\nu \| P_{w^0}) \geq 0, \qquad w \in W. \tag{4}$$

Since W is not assumed compact, we impose the following conditions to control concentration at infinity. There exist $\theta \in [0,1]^\ell$, $C \geq 0$, and $u \in (0,\infty]^\ell$ such that

$$\log \mathbb{E}_\nu e^{-\sum_{j=1}^{\ell} s_j f_j(X,w)} \leq -\sum_{j=1}^{\ell} s_j \theta_j K_j(w) + C\|s\|_2^2 \tag{5}$$

for every $w \in W$ and every $s \in [0,\infty)^\ell$ satisfying $s_j \leq u_j$. For the same θ, we assume that

$$\left\{ w \in W : \sum_{j=1}^{\ell} \theta_j K_j(w) \leq a \right\} \quad \text{is compact} \tag{6}$$

for some $a > 0$.

Remark 1.2. Condition (5) is equivalent, up to constants and reducing the bounds u_j, to

$$\log \mathbb{E}_\nu e^{-s_j f_j(X,w)} \leq -\theta_j s_j K_j(w) + C_j s_j^2, \qquad 0 \leq s_j \leq u_j, \quad w \in W, \quad \forall j \in \{1,\dots,\ell\}. \tag{7}$$

Indeed, take $s = s_j e_j$ to obtain the componentwise bounds with $C_j = C$ and the same coordinate bounds u_j. Conversely, Hölder gives

$$\log \mathbb{E} e^{-\sum_j s_j f_j} \leq \frac{1}{\ell} \sum_j \log \mathbb{E} e^{-\ell s_j f_j} \leq -\sum_j s_j \theta_j K_j + \ell \sum_j C_j s_j^2.$$

Thus the vector bound holds on $\prod_j [0, u_j/\ell]$ with $C = \ell \max_j C_j$.

Dropping the index j, condition (7) is satisfied in the following cases.

1. *Deterministic.* If $f(X,w) = K(w)$ almost surely, then

$$\log \mathbb{E}_\nu e^{-sf(X,w)} = -sK(w).$$

Take $\theta = 1$, $C = 0$, and $u = \infty$.

2. *Sub-Gaussian.* Suppose the centered losses are uniformly sub-Gaussian: for some $\sigma^2 < \infty$,

$$\log \mathbb{E}_\nu e^{s(f(X,w)-K(w))} \leq \sigma^2 s^2 \qquad (s \in \mathbb{R},\ w \in W).$$

Then

$$\log \mathbb{E}_\nu e^{-sf(X,w)} = -sK(w) + \log \mathbb{E}_\nu e^{-s(f(X,w)-K(w))} \leq -sK(w) + \sigma^2 s^2.$$

Take $\theta = 1$, $C = \sigma^2$, and $u = \infty$.

3. *Realizable.* In the KL divergence example, if $P_{w^0} = \nu$, then $f = \log r$. Put $A(w) = \{x : r(x,w) > 0\}$ and $m(w) = P_w(A(w)) > 0$. For $0 \leq s \leq 1$, Jensen's inequality for the probability measure $P_w|_{A(w)}/m(w)$ gives

$$\begin{aligned} \mathbb{E}_\nu e^{-sf(X,w)} &= \int_{A(w)} r^{-s}\, d\nu = \int_{A(w)} r^{1-s}\, dP_w \\ &\leq m(w) \left(\frac{1}{m(w)} \int_{A(w)} r\, dP_w \right)^{1-s} = m(w)^s \leq 1. \end{aligned}$$

Take $\theta = 0$, $C = 0$, and $u = 1$.

4. Suppose that, for constants $A, B \geq 0$ and $s^* \in (0, \infty]$,

$$\log \mathbb{E}_\nu e^{-s(f(X,w)-K(w))} \leq s^2(A + BK(w)), \qquad w \in W, \quad 0 \leq s \leq s^*.$$

Then

$$\log \mathbb{E}_\nu e^{-sf(X,w)} \leq -s(1 - Bs)K(w) + As^2.$$

Take any $\theta \in (0, 1)$, $C = A$, and

$$u = \min\left\{s^*, \frac{1-\theta}{B}\right\},$$

where $(1 - \theta)/B = +\infty$ when $B = 0$.

For example, let $\nu = N(a, T)$ and $P_w = N(m(w), \Sigma)$, with fixed positive-definite covariance matrices T, Σ and optimal parameter w^0. Write $\|v\|^2_{\Sigma^{-1}} = v^\top \Sigma^{-1} v$, and put

$$m_0 = m(w^0), \qquad d_0^2 = \|m_0 - a\|^2_{\Sigma^{-1}}, \qquad L = \lambda_{\max}(\Sigma^{-1/2} T \Sigma^{-1/2}).$$

For the centered negative log-likelihood,

$$K(w) = \frac{1}{2}\left(\|m(w) - a\|^2_{\Sigma^{-1}} - d_0^2\right),$$
$$f(X, w) - K(w) = -(X - a)^\top \Sigma^{-1}(m(w) - m_0).$$

With $v = m(w) - m_0$, we have

$$v^\top \Sigma^{-1} T \Sigma^{-1} v \leq L\|v\|^2_{\Sigma^{-1}} \leq L\left(2\|m(w) - a\|^2_{\Sigma^{-1}} + 2d_0^2\right) = 4L(K(w) + d_0^2).$$

The Gaussian moment-generating function therefore gives

$$\log \mathbb{E}_\nu e^{-s(f-K)} = \frac{s^2}{2} v^\top \Sigma^{-1} T \Sigma^{-1} v \leq s^2\left(2Ld_0^2 + 2LK(w)\right).$$

Take $A = 2Ld_0^2$, $B = 2L$, and $s^* = \infty$.

1.2 Localization and the main theorem

Suppose the KL divergence loss f comes from a model realizable at w_*, and write $K(w) = \mathbb{E}_\nu f(X, w)$. As above, we may take $\theta = C = 0$ and $u = 1$. With this choice, if there are no additional losses and W is noncompact, (6) cannot hold. A simple fix is to add a deterministic loss $R(w)$ such that $\{w \in W : R(w) \leq a\}$ is compact for some $a > 0$. For instance, for a minimizer $w_* \in W$ of K, set

$$R(w) = \|w - w_*\|^2.$$

For the pair (f, R), $\theta = (0, 1)$, $C = 0$, and $u = (1, \infty)$ satisfy (5). Since W is closed, the sublevel sets of R are compact, so (6) also holds.

Although this additional loss prevents concentration at infinity, it can alter the local geometry of the loss landscape. The following conjecture asks whether weights can be chosen on each function so that the Gibbs estimator has the same asymptotic loss as if the prior were restricted to a sufficiently small relative ball $W \cap B_\delta(w_*)$. It formalizes the conjecture in Appendix D of *The Local Learning Coefficient: A Singularity-Aware Complexity Measure* [1], showing that the estimator proposed by the authors is consistent.

Given independent observations $X_1, X_2, \ldots$ with distribution ν, define the empirical loss

$$\widehat{K}_n(w) = \frac{1}{n}\sum_{i=1}^{n} f(X_i, w).$$

Assume $K \not\equiv 0$ on Ω. For $z \in W$ with $K(z) = 0$, define the local RLCT relative to the prior by

$$\lambda_z(K; W, \mu) = \sup\left\{ s \geq 0 : \int_{W\cap B} K(w)^{-s}\,\mathrm{d}\mu(w) < \infty \text{ for some open } B \subseteq \Omega \text{ containing } z \right\}. \quad (8)$$

Conjecture 1.3. *There exist sequences $\beta_n, \gamma_n > 0$ with*

$$\beta_n \longrightarrow 0, \qquad n\beta_n \longrightarrow \infty, \qquad \gamma_n \longrightarrow \infty,$$

such that, for every KL divergence loss as in Example 1.1 with $K \not\equiv 0$ and every $w_* \in W$ such that $P_{w_*} = \nu$,

$$n\beta_n \frac{\displaystyle\int_W \widehat{K}_n(w) e^{-n\beta_n \widehat{K}_n(w) - \gamma_n R(w)}\,\mathrm{d}\mu(w)}{\displaystyle\int_W e^{-n\beta_n \widehat{K}_n(w) - \gamma_n R(w)}\,\mathrm{d}\mu(w)} = \lambda_{w_*}(K; W, \mu) + o_{\mathbb{P}}(1).$$

We prove a more general result. Index a sequence of estimators by $t \in \mathbb{N}$. For each loss j, let $n_{j,t} \in \mathbb{N}$ be sample sizes with $n_{j,t} \to \infty$, and let $X_{j,1}, X_{j,2}, \ldots$ be independent observations with distribution ν within that sample stream. Put

$$K_{j,t}(w) = \frac{1}{n_{j,t}} \sum_{i=1}^{n_{j,t}} f_j(X_{j,i}, w). \quad (9)$$

Notably, we impose no restriction on dependence between different streams. We attach to each loss an inverse temperature $\beta_{j,t}$ and put

$$A_t = n_{1,t}\beta_{1,t}, \qquad \widehat{Q}_t(w) = \sum_{j=1}^{\ell} n_{j,t}\beta_{j,t} K_{j,t}(w).$$

For each data set, whenever the normalization constant is finite and positive, define the empirical Gibbs posterior and its posterior expectation functional by

$$\mathrm{d}\widehat{\Pi}_t(w) = \frac{e^{-\widehat{Q}_t(w)}\,\mathrm{d}\mu(w)}{\int_W e^{-\widehat{Q}_t}\,\mathrm{d}\mu}, \qquad \langle g \rangle_t := \int_W g(w)\,\mathrm{d}\widehat{\Pi}_t(w), \quad (10)$$

for $g \in L^1(W, \widehat{\Pi}_t)$. Let $\mathcal{H}$ be the class of continuous functions $h : \mathbb{R} \to \mathbb{R}$ such that

$$|h(x)| \leq C_h e^{\eta|x|} \quad (x \in \mathbb{R}) \qquad \text{for some } C_h > 0 \text{ and } 0 < \eta < 1. \quad (11)$$

This class contains all bounded continuous functions and all polynomials. For $\lambda > 0$, write γ_λ for the Gamma probability measure with shape λ and rate 1:

$$\mathrm{d}\gamma_\lambda(x) = \frac{x^{\lambda-1}e^{-x}}{\Gamma(\lambda)} \mathbf{1}_{(0,\infty)}(x)\,\mathrm{d}x.$$

Then $\mathcal{H} \subseteq L^1(\mathbb{R}, \gamma_\lambda)$ for every $\lambda > 0$.

Theorem 1.4. *Suppose*

$$\varnothing \neq Z = \{w \in W : K_1(w) = \cdots = K_\ell(w) = 0\},$$

conditions (5) *and* (6) *are satisfied,* $K_1 \not\equiv 0$*, and*

$$\begin{aligned} &n_{1,t}\beta_{1,t} \longrightarrow \infty, \\ &\sqrt{n_{j,t}}\beta_{j,t} \longrightarrow 0, && \forall j \in \{1,\dots,\ell\}, \\ &\frac{n_{j,t}\beta_{j,t}}{\log(n_{1,t}\beta_{1,t})} \longrightarrow \infty, \qquad \frac{\log(n_{j,t}\beta_{j,t})}{\log(n_{1,t}\beta_{1,t})} \longrightarrow 0, && \forall j \in \{2,\dots,\ell\}. \end{aligned} \tag{12}$$

Then $\lambda_Z = \min_{z\in Z} \lambda_z(K_1; W, \mu) > 0$ *is attained. For all sufficiently large* t*,* $\widehat{\Pi}_t$ *is well defined almost surely. For every* $h \in \mathcal{H}$*,* $h(A_tK_{1,t}) \in L^1(W, \widehat{\Pi}_t)$ *almost surely for all sufficiently large* t*, and*

$$\big\langle h(A_tK_{1,t})\big\rangle_t = \int_{\mathbb{R}} h \,\mathrm{d}\gamma_{\lambda_Z} + o_{\mathbb{P}}(1). \tag{13}$$

In particular,

$$\langle A_tK_{1,t}\rangle_t = \lambda_Z + o_{\mathbb{P}}(1). \tag{14}$$

2 Reduction to local population integrals

2.1 Control at infinity

Put

$$S = \sum_j \theta_j K_j, \qquad b_t = \min_j n_{j,t}\beta_{j,t}, \qquad \sigma_t^2 = \sum_j n_{j,t}\beta_{j,t}^2, \qquad Y_{j,t} = n_{j,t}\beta_{j,t}K_{j,t}.$$

The temperature assumptions give $b_t/\log A_t \to \infty$ and $\sigma_t \to 0$. For a measurable set $D \subseteq W$, define the tilted partition functions

$$\begin{aligned} \mathcal{Z}_{D,t}(c) &= \int_D e^{-cA_tK_1 - \sum_{j\geq 2} n_{j,t}\beta_{j,t}K_j} \,\mathrm{d}\mu, \\ \widehat{\mathcal{Z}}_{D,t}(c) &= \int_D e^{-cY_{1,t} - \sum_{j\geq 2} Y_{j,t}} \,\mathrm{d}\mu, \qquad c > 0. \end{aligned} \tag{15}$$

Varying c will give the posterior exponential moments $\langle e^{sA_tK_{1,t}}\rangle_t$. Take

$$V = \overline{W \cap \{S < a\}}.$$

This set is a closed subset of the compact set in (6), and is semianalytic by standard properties of semianalytic sets. Since $W = \overline{W^\circ}$ and $\{S < a\}$ is open, we have

$$V = \overline{W^\circ \cap \{S < a\}} = \overline{V^\circ}.$$

Since $S = 0$ on Z and $a > 0$, we also have $Z \subset \operatorname{int}_W V$, while $S \geq a$ on $W \setminus V$.

Applying (5) on the jth coordinate axis and using independence of the observations within the jth sample stream gives

$$\log \mathbb{E}e^{-sY_{j,t}(w)} = n_{j,t} \log \mathbb{E}_\nu e^{-s\beta_{j,t}f_j(X,w)} \leq -s\theta_j n_{j,t}\beta_{j,t}K_j(w) + Cs^2 n_{j,t}\beta_{j,t}^2,$$

provided $s\beta_{j,t} \le u_j$. Fix $c > 0$ and write $c_* = \min(c, 1)$. Since $\beta_{j,t} \to 0$, for all sufficiently large t this estimate applies with $s = \ell c$ in the first stream and $s = \ell$ in the remaining streams. As in Remark 1.2, Hölder's inequality gives

$$\begin{aligned}\log \mathbb{E}e^{-cY_{1,t}-\sum_{j\ge 2} Y_{j,t}} &\le \frac{1}{\ell}\log \mathbb{E}e^{-\ell c Y_{1,t}} + \frac{1}{\ell}\sum_{j\ge 2}\log \mathbb{E}e^{-\ell Y_{j,t}} \\ &\le -c\theta_1 A_t K_1 - \sum_{j\ge 2}\theta_j n_{j,t}\beta_{j,t}K_j + C_c\sigma_t^2.\end{aligned} \tag{16}$$

Therefore

$$\mathbb{E}e^{-cY_{1,t}(w)-\sum_{j\ge 2} Y_{j,t}(w)} \le e^{-c_* b_t S(w) + C_c\sigma_t^2} \qquad (w \in W). \tag{17}$$

Since $S \ge a$ outside V and $\mu(W) = 1$, Tonelli's theorem gives

$$\begin{aligned}\mathbb{E}\widehat{\mathcal{Z}}_{W\setminus V,t}(c) &\le e^{-ac_* b_t + C_c\sigma_t^2}, \\ \mathcal{Z}_{W\setminus V,t}(c) &\le e^{-ac_* b_t}.\end{aligned} \tag{18}$$

In particular, the empirical partition functions are finite almost surely for all sufficiently large t.

2.2 Removing the sampling noise on V

Write

$$Q_t = \sum_j n_{j,t}\beta_{j,t}K_j, \qquad \Delta_{j,t} = n_{j,t}\beta_{j,t}(K_{j,t} - K_j), \qquad \Delta_t(c) = c\Delta_{1,t} + \sum_{j\ge 2}\Delta_{j,t}.$$

Analyticity into $L^2(\nu)$ gives $\sup_{w\in V,j}\mathbb{E}_\nu f_j(X,w)^2 < \infty$. Choose $s_0 > 0$ with $s_0 \le \min_j u_j$. By (5) on each coordinate axis and $K_j \ge 0$, $\sup_{w\in V,j}\mathbb{E}_\nu e^{-s_0 f_j(X,w)} < \infty$. Since $x^2 e^{-s_0 x/2} \le C_{s_0}(1 + e^{-s_0 x})$ for $x \in \mathbb{R}$,

$$\sup_{w\in V,j}\mathbb{E}_\nu\bigl[f_j(X,w)^2(1 + e^{-s_0 f_j(X,w)/2})\bigr] < \infty.$$

For $0 \le s \le s_0/2$, $0 \le r \le 1$, and $x \in \mathbb{R}$, we have $e^{-rsx} \le 1 + e^{-s_0 x/2}$. Taylor's formula with integral remainder gives

$$\begin{aligned}e^{-sx} &= 1 - sx + s^2x^2\int_0^1 (1-r)e^{-rsx}\,\mathrm{d}r \\ &\le 1 - sx + \frac{s^2x^2}{2}(1 + e^{-s_0 x/2}).\end{aligned}$$

Taking expectations with $x = f_j(X,w)$ and using the preceding uniform bound gives the first inequality below. The second follows from $\log z \le z - 1$ and $\mathbb{E}_\nu f_j = K_j$:

$$\mathbb{E}_\nu e^{-sf_j} \le 1 - sK_j + C_V s^2, \qquad \log \mathbb{E}_\nu e^{-s(f_j - K_j)} \le C_V s^2 \quad (0 \le s \le s_0/2). \tag{19}$$

Fix $c > 0$. Independence within the jth sample stream gives

$$\mathbb{E}e^{-q\Delta_{j,t}(w)} = \Bigl(\mathbb{E}_\nu e^{-q\beta_{j,t}(f_j(X,w)-K_j(w))}\Bigr)^{n_{j,t}} \le e^{C_V q^2 n_{j,t}\beta_{j,t}^2} \qquad (0 \le q\beta_{j,t} \le s_0/2).$$

Since $\beta_{j,t} \to 0$, for all sufficiently large t this applies with $q = \ell r c$ in the first stream and $q = \ell r$ in the others, uniformly for $0 \le r \le 2$. Hölder's inequality then gives

$$\begin{aligned} 1 \le \mathbb{E}e^{-r\Delta_t(c)} &\le \left(\mathbb{E}e^{-\ell r c \Delta_{1,t}}\right)^{1/\ell} \prod_{j\ge 2} \left(\mathbb{E}e^{-\ell r \Delta_{j,t}}\right)^{1/\ell} \\ &\le \exp\left(C_V \ell r^2 \left[c^2 n_{1,t}\beta_{1,t}^2 + \sum_{j\ge 2} n_{j,t}\beta_{j,t}^2 \right] \right) \\ &\le e^{C_{V,c} r^2 \sigma_t^2} \qquad (0 \le r \le 2,\ w \in V). \end{aligned} \tag{20}$$

Here the lower bound is Jensen's inequality, since $\mathbb{E}\Delta_t(c) = 0$, and the last bound uses $C_{V,c} = C_V \ell \max(c^2, 1)$. Consequently,

$$\begin{aligned} \mathbb{E}|e^{-\Delta_t(c)} - 1|^2 &= \mathbb{E}e^{-2\Delta_t(c)} - 2\mathbb{E}e^{-\Delta_t(c)} + 1 \\ &\le e^{4C_{V,c}\sigma_t^2} - 1 = O(\sigma_t^2), \end{aligned} \tag{21}$$

uniformly on V. By Cauchy–Schwarz, $\mathbb{E}|e^{-\Delta_t(c)} - 1| \le C_{V,c}\sigma_t$. Multiplying by $e^{-cA_tK_1 - \sum_{j\ge 2} n_{j,t}\beta_{j,t}K_j}$, integrating over D, and applying Tonelli's theorem gives

$$\mathbb{E}|\widehat{\mathcal{Z}}_{D,t}(c) - \mathcal{Z}_{D,t}(c)| \le C_{V,c}\sigma_t \mathcal{Z}_{D,t}(c) \qquad (D \subseteq V \text{ measurable}). \tag{22}$$

2.3 Localization near Z

Fix a measurable set $D \subseteq W$ with $Z \subseteq \operatorname{int}_W D$. If $D = W$, the tails are zero. Otherwise, compactness of V and the inequality $\sum_j K_j \ge S \ge a$ outside V give

$$\delta_D := \inf_{w \in W\setminus D} \sum_j K_j(w) > 0. \tag{23}$$

For $c > 0$ and $c_* = \min(c, 1)$, the population exponent satisfies

$$cA_tK_1 + \sum_{j\ge 2} n_{j,t}\beta_{j,t}K_j \ge c_* b_t \sum_j K_j.$$

Combining (18) and (20) at $r = 1$, we obtain

$$\begin{aligned} \mathbb{E}\widehat{\mathcal{Z}}_{W\setminus D,t}(c) &\le \mathbb{E}\widehat{\mathcal{Z}}_{W\setminus V,t}(c) + \mathbb{E}\widehat{\mathcal{Z}}_{V\setminus D,t}(c) \\ &\le e^{-ac_*b_t + C_c\sigma_t^2} + e^{C_{V,c}\sigma_t^2}\mathcal{Z}_{V\setminus D,t}(c) \\ &\le e^{-ac_*b_t + C_c\sigma_t^2} + e^{-c_*b_t\delta_D + C_{V,c}\sigma_t^2} \\ &= o(A_t^{-M}) \qquad \text{for every } M > 0. \end{aligned} \tag{24}$$

The last step uses $b_t/\log A_t \to \infty$.

Proposition 2.1. *Under the hypotheses of Theorem 1.4, let $D \subseteq W$ be measurable with $Z \subseteq \operatorname{int}_W D$. For every fixed $c > 0$,*

$$\frac{\mathbb{E}|\widehat{\mathcal{Z}}_{W,t}(c) - \mathcal{Z}_{D,t}(c)|}{\mathcal{Z}_{D,t}(c)} \longrightarrow 0. \tag{25}$$

Proof. Put $H = \sum_j K_j$. For all sufficiently large t, $n_{j,t}\beta_{j,t} \le A_t$ for every j, so

$$cA_tK_1 + \sum_{j\ge 2} n_{j,t}\beta_{j,t}K_j \le \max(c,1)A_tH.$$

The nonzero analytic function H vanishes on Z. Since Z is compact, choose a compact semianalytic set $D' \subseteq D$ with $Z \subseteq \operatorname{int}_W D'$ by taking a finite union of closed balls intersected with W. The standard scalar analytic Laplace asymptotics on D' [5, 3] give some $C_0 > 0$ such that

$$\mathcal{Z}_{D,t}(c) \ge \int_{D'} e^{-\max(c,1)A_tH}\,\mathrm{d}\mu \ge A_t^{-C_0} \qquad \text{for all sufficiently large } t. \tag{26}$$

The leading term is positive because D' contains a relative neighborhood in W of a point $z \in Z \subseteq \operatorname{supp}(\mu|_W)$. By (18),

$$\mathcal{Z}_{D\setminus V,t}(c) \le \mathcal{Z}_{W\setminus V,t}(c) = o(A_t^{-M}) \qquad \text{for every } M > 0. \tag{27}$$

Decomposing both integrals over $D \cap V$ gives

$$\begin{aligned}\widehat{\mathcal{Z}}_{W,t}(c) - \mathcal{Z}_{D,t}(c) = \widehat{\mathcal{Z}}_{D\cap V,t}(c) - \mathcal{Z}_{D\cap V,t}(c)\\ + \widehat{\mathcal{Z}}_{W\setminus(D\cap V),t}(c) - \mathcal{Z}_{D\setminus V,t}(c).\end{aligned}$$

For the first difference, (22) gives

$$\mathbb{E}|\widehat{\mathcal{Z}}_{D\cap V,t}(c) - \mathcal{Z}_{D\cap V,t}(c)| \le C_{V,c}\sigma_t\mathcal{Z}_{D\cap V,t}(c) \le C_{V,c}\sigma_t\mathcal{Z}_{D,t}(c).$$

Since $Z \subseteq \operatorname{int}_W(D \cap V)$, (24) also gives

$$\mathbb{E}\widehat{\mathcal{Z}}_{W\setminus(D\cap V),t}(c) = o(A_t^{-M}) \qquad \text{for every } M > 0.$$

Together with (27), the triangle inequality therefore yields

$$\mathbb{E}|\widehat{\mathcal{Z}}_{W,t}(c) - \mathcal{Z}_{D,t}(c)| \le C_{V,c}\sigma_t\mathcal{Z}_{D,t}(c) + o(A_t^{-M}).$$

Finally, divide by $\mathcal{Z}_{D,t}(c)$ and use (26). For $M > C_0$,

$$\begin{aligned}\frac{\mathbb{E}|\widehat{\mathcal{Z}}_{W,t}(c) - \mathcal{Z}_{D,t}(c)|}{\mathcal{Z}_{D,t}(c)} &\le C_{V,c}\sigma_t + \frac{o(A_t^{-M})}{\mathcal{Z}_{D,t}(c)}\\ &\le C_{V,c}\sigma_t + o(A_t^{C_0-M}) \longrightarrow 0.\end{aligned} \tag{28}$$

□

3 Local geometry

3.1 Resolution of singularities

We study the population integrals

$$\mathcal{Z}_{D,t}(c) = \int_D \exp\left(-cA_tK_1(w) - \sum_{j=2}^{\ell} n_{j,t}\beta_{j,t}K_j(w)\right) \mathrm{d}\mu(w), \qquad c > 0.$$

Proposition 2.1 allows us to take D to be a neighborhood of Z. We now describe these integrals in local coordinates. Since the integral is unaffected by K if it is identically 0, in this section we omit identically zero losses and relabel the remaining ones as $K_1, \dots, K_\ell$. We assume K_1 is not identically 0. Put

$$F = (K_1, \dots, K_\ell) : W \longrightarrow [0,\infty)^\ell.$$

For a measurable neighborhood D of Z relative to W, define

$$\eta_D = F_*(\mu|_D). \tag{29}$$

Theorem 3.1. *There is an open relatively compact semianalytic neighborhood D_* of Z such that, for every measurable $D \subseteq W$ with*

$$Z \subseteq \operatorname{int}_W D, \qquad D \subseteq W \cap D_*, \tag{30}$$

we have a finite decomposition

$$\eta_D = \sum_{\alpha,\sigma} (F_{\alpha\sigma})_* m_{\alpha\sigma,D}. \tag{31}$$

Each term is defined on $(0,1)^d$ by

$$F_{\alpha\sigma,i}(x) = K_i(\pi_{\alpha\sigma}(x)) = u_{\alpha\sigma,i}(x)x^{A_{\alpha i}}, \qquad A_{\alpha i} \in \mathbb{N}_0^d \setminus \{0\}, \tag{32}$$

$$\mathrm{d}m_{\alpha\sigma,D}(x) = b_{\alpha\sigma}(x)\rho_{\alpha\sigma,D}(x)x^{\beta_\alpha - \mathbf{1}}\,\mathrm{d}x, \qquad \beta_{\alpha j} \geq 1. \tag{33}$$

The maps $\pi_{\alpha\sigma}$ extend to the origin, with $\pi_{\alpha\sigma}(0) = z_\alpha \in Z$. The units $u_{\alpha\sigma,i}$ and $b_{\alpha\sigma}$ are bounded above and below by positive constants. The cutoff satisfies $0 \leq \rho_{\alpha\sigma,D} \leq 1$ and is bounded below by a positive constant on some $(0,\varepsilon_D)^d$.

Proof. Use the local standard form and orthant integration formula of [4, Theorems 7 and 8], resolving the losses, the analytic prior factor, and the local defining functions of W. $K_i(z_\alpha) = 0$ implies $A_{\alpha i} \neq 0$. □

Proposition 3.2. *Use the charts of Theorem 3.1 for $(K_1, \dots, K_\ell)$, and put*

$$\lambda_\alpha = \min_{j : A_{\alpha 1 j} > 0} \frac{\beta_{\alpha j}}{A_{\alpha 1 j}}, \qquad \lambda = \min_\alpha \lambda_\alpha. \tag{34}$$

For every D satisfying (30)*,*

$$\sup\{s : \int_D K_1^{-s}\,\mathrm{d}\mu < \infty\} = \lambda = \min_{z \in Z} \lambda_z(K_1; W, \mu) = \lambda_Z.$$

For semianalytic D, $-\lambda_Z$ is the rightmost real pole of $\int_D K_1^z\,\mathrm{d}\mu$.

Proof. Apply the chart formula for the RLCT [5, p. 874, equation (15)] to the charts of Theorem 3.1 over Z. The pole characterization follows from [4, Theorem 9]. □

The ratios in (34) are ratios of positive integers, so λ_Z is a positive rational number. We call a semianalytic D satisfying (30) *admissible* and hold it fixed as the weights vary.

3.2 Distribution asymptotics

The posterior concentrates near Z, where all losses vanish. We use dilation operators on distributions to zoom in around Z. For $r, \mathbf{t} \in (0,\infty)^\ell$, put

$$B(r) = \prod_{i=1}^{\ell} [0, r_i), \qquad b_D(r) = \eta_D(B(r)) = \mu(D \cap \{K_i < r_i \text{ for all } i\}).$$

Define

$$\delta_{\mathbf{t}}(x) = (t_1 x_1, \dots, t_\ell x_\ell), \qquad \eta_{D,\mathbf{t}} = (\delta_{\mathbf{t}})_* \eta_D, \qquad \mathcal{V}_{D,r} = \frac{\eta_{D,r^{-1}}}{b_D(r)}. \tag{35}$$

Vector reciprocals and products are coordinatewise. The chart bounds imply $0 < b_D(r) < \infty$, and this normalization gives $\mathcal{V}_{D,r}(B(\mathbf{1})) = 1$. For comparison with [2, Section 4.1], write

$$D_\omega = \delta_{e^{-\omega}}, \qquad T_\omega \eta = (D_\omega)_* \eta, \qquad \eta_{D,\mathbf{t}} = T_{-\log \mathbf{t}} \eta_D.$$

The following lemma gives a nonzero subsequential limit of $\mathcal{V}_{D,r}$ along any sequence of scales.

Lemma 3.3. *Put $B_s = [0,s)^\ell$. If η is a finite nonempty sum of the chart measures of the form (32)–(33), there are constants $C, D_0 > 0$ such that*

$$0 < \eta(B(r)) < \infty, \qquad \eta(B(sr)) \le C s^{D_0} \eta(B(r)) \quad (s \ge 1,\ r \in (0,\infty)^\ell). \tag{36}$$

Consequently the measures

$$\mathcal{V}_r := \frac{(x \mapsto x/r)_* \eta}{b_r}, \qquad b_r := \eta(B(r)), \qquad r \in (0,\infty)^\ell,$$

where division is coordinatewise, are sequentially pre-compact in the strong topology of $\mathcal{S}'(\mathbb{R}^\ell)$, and every subsequential limit is positive and nonzero.

Proof. For one chart write $\eta^{\mathrm{ch}} = F_* m$ and

$$E(r) = F^{-1}(B(r)), \qquad \mathrm{d}m = b\rho x^{\beta - \mathbf{1}}\, \mathrm{d}x.$$

Choose constants satisfying

$$0 < u_- \le u_i \le u_+, \qquad 0 < b_- \le b \le b_+, \qquad 0 \le \rho \le \rho_+ \quad \text{on } (0,1)^d,$$
$$\rho \ge \rho_- > 0 \text{ on } (0,\varepsilon)^d.$$

Since $|A_i|_1 = \sum_j A_{ij} \ge 1$, choose $c > 0$ with

$$0 < c < \min\{1, \varepsilon\}, \qquad \frac{u_+}{u_-} c^{|A_i|_1} \le 1 \quad (1 \le i \le \ell).$$

For $s \ge 1$, put $\kappa_s(x) = cx/s$. Then

$$\begin{aligned} &\kappa_s((0,1)^d) \subset (0,\varepsilon)^d, \\ &\frac{F_i(\kappa_s(x))}{F_i(x)} = \frac{u_i(\kappa_s(x))}{u_i(x)} c^{|A_i|_1} s^{-|A_i|_1} \le s^{-1}, \\ &\kappa_s(E(sr)) \subset E(r). \end{aligned} \tag{37}$$

Put $d_0 = \sum_j \beta_j$. The change of variables $x \mapsto \kappa_s(x)$ gives

$$(\kappa_s(x))^{\beta-\mathbf{1}} |\det D\kappa_s| = (c/s)^{d_0} x^{\beta-\mathbf{1}}.$$

Since $\rho \geq \rho_-$ on the image of κ_s, we obtain

$$\begin{aligned} \eta^{\mathrm{ch}}(B(r)) = m(E(r)) &\geq m(\kappa_s(E(sr))) \\ &\geq b_-\rho_- c^{d_0} s^{-d_0} \int_{E(sr)} x^{\beta-\mathbf{1}} \,\mathrm{d}x \\ &\geq c_* s^{-d_0} m(E(sr)) = c_* s^{-d_0} \eta^{\mathrm{ch}}(B(sr)), \end{aligned} \tag{38}$$

where $c_* = (b_-\rho_-)/(b_+\rho_+)\, c^{d_0} > 0$. Taking maxima over the finitely many charts in (38) gives (36).

Consequently $\mathcal{V}_r(B_1) = 1$ and $\mathcal{V}_r(B_s) \leq C s^{D_0}$ for $s \geq 1$. For $M > D_0$ and $p_M(h) = \sup_x (1 + \|x\|)^M |h(x)|$, a dyadic decomposition gives

$$|\langle \mathcal{V}_r, h\rangle| \leq \left(1 + C 2^{D_0} \sum_{j=0}^{\infty} 2^{-j(M-D_0)}\right) p_M(h).$$

Thus $\{\mathcal{V}_r\} \subset \mathcal{S}'$ is equicontinuous. On equicontinuous subsets of the dual of a Montel space, the weak-* and strong dual topologies coincide. Since $\mathcal{S}$ is Montel, Banach–Alaoglu and separability of $\mathcal{S}$ imply that every equicontinuous subset of $\mathcal{S}'$ is relatively weak-* sequentially compact, hence relatively strongly sequentially compact. Choose $\chi \in C_c^\infty(\mathbb{R}^\ell)$ with $\chi \geq 0$ and $\chi = 1$ on $[0,1]^\ell$. For every strong subsequential limit $\mathcal{V}$,

$$\mathcal{V} \geq 0, \qquad \langle \mathcal{V}, \chi\rangle = \lim \langle \mathcal{V}_r, \chi\rangle \geq 1,$$

so $\mathcal{V} \neq 0$. □

For an admissible D, the Mellin transform of η_D is the zeta function of [4, Definition 4]:

$$M_D(z) = \int_{D \cap \{K_1 \cdots K_\ell > 0\}} \prod_{i=1}^{\ell} K_i(w)^{z_i} \,\mathrm{d}\mu(w). \tag{39}$$

Choose $\varepsilon > 0$ with $\beta_{\alpha j} - \varepsilon \sum_i A_{\alpha i j} > 0$ for every chart and every j. The decomposition in Theorem 3.1 gives

$$\int_D \prod_{i=1}^{\ell} K_i(w)^{-\varepsilon} \,\mathrm{d}\mu(w) < \infty. \tag{40}$$

More generally, the same decomposition shows that M_D is holomorphic on

$$\left\{ z \in \mathbb{C}^\ell : \beta_{\alpha j} + \sum_{i=1}^{\ell} A_{\alpha i j} \operatorname{Re} z_i > 0 \text{ for every } \alpha, j \right\}.$$

The meromorphic continuation and polynomial growth results [4, Theorem 9 and Corollary 3] apply on finitely many local semianalytic pieces of $\overline{D}$. They give continuation of M_D to $\mathbb{C}^\ell$, with possible poles on a locally finite real affine hyperplane arrangement with finitely many normal directions. On each compact set of real parts, away from this arrangement,

$$|M_D(z)| \leq C(1 + \|\operatorname{Im} z\|)^N.$$

We verify the hypotheses of [2, Section 4.1] as follows:

(M1) The measure η_D is finite and compactly supported, hence tempered. It has no coordinate-boundary mass, so the cutoff extension of [2, Proposition 4.1] satisfies $E\eta_D = \eta_D$.

(M2) For $c \in (-\varepsilon, 0)^\ell$, the measure
$$\eta_c = (\log)_*(x^c \eta_D|_{(0,\infty)^\ell})$$
is finite by (40), and $\widehat{\eta_c}(-\xi) = M_D(c + i\xi)$.

(M3) The continuation and vertical growth just stated give the required polar arrangement and growth bounds.

The homogeneous leading term theorem [2, Theorem 8.4] then gives the following consequence.

Proposition 3.4. *Fix an admissible D. Suppose $\mathbf{t}_n \in (0,\infty)^\ell$ and*
$$\omega_n = -\log \mathbf{t}_n, \qquad \|\omega_n\| \longrightarrow \infty, \qquad \frac{\omega_n}{\|\omega_n\|} \longrightarrow \theta.$$
If
$$\frac{\eta_{D,\mathbf{t}_n}}{b_D(\mathbf{t}_n^{-1})} \longrightarrow \mathcal{V} \neq 0 \quad \textit{strongly in } \mathcal{S}',$$
then there is a unique $\lambda \in \mathbb{R}$ such that
$$T_{h\theta}\mathcal{V} = e^{-h\lambda}\mathcal{V} \quad (h \in \mathbb{R}), \qquad \frac{\log b_D(\mathbf{t}_n^{-1})}{\|\omega_n\|} \longrightarrow -\lambda. \tag{41}$$

Proof. The preceding verification gives the Mellin hypotheses, and Lemma 3.3 gives boundedness of the normalized family. Apply [2, Theorem 8.4] with $a_n = b_D(\mathbf{t}_n^{-1})$. □

Under our weight assumptions, $\theta = -e_1$. Hence every subsequential limit is homogeneous under dilation of the first loss coordinate. It remains to identify the exponent and show that it is the same for all subsequences.

4 Posterior limits

Lemma 4.1. *Fix an admissible D, and put*
$$\boldsymbol{\tau}_t = (A_t, \dots, n_{\ell,t}\beta_{\ell,t}), \qquad \mathfrak{b}_t = b_D(\boldsymbol{\tau}_t^{-1}).$$
Suppose $A_t \to \infty$ and
$$n_{j,t}\beta_{j,t} \longrightarrow \infty, \qquad \frac{\log(n_{j,t}\beta_{j,t})}{\log A_t} \longrightarrow 0 \qquad (2 \le j \le \ell). \tag{42}$$
Then
$$\mathfrak{b}_t = A_t^{-\lambda_Z + o(1)}, \qquad \frac{\log \mathfrak{b}_t}{\log A_t} \longrightarrow -\lambda_Z. \tag{43}$$

Proof. Write $L_t = \log A_t$ and compare the original weights with
$$\varepsilon_t = e^{-\sqrt{L_t}}, \qquad \widetilde{\boldsymbol{\tau}}_t = (A_t, \varepsilon_t, \dots, \varepsilon_t), \qquad \widetilde{\mathfrak{b}}_t = b_D(\widetilde{\boldsymbol{\tau}}_t^{-1}).$$
Their logarithmic scales satisfy
$$\omega_t = -\log \boldsymbol{\tau}_t, \qquad \widetilde{\omega}_t = -\log \widetilde{\boldsymbol{\tau}}_t, \qquad \frac{\omega_t}{L_t}, \frac{\widetilde{\omega}_t}{L_t} \longrightarrow -e_1, \qquad \|\omega_t - \widetilde{\omega}_t\| = o(L_t).$$

From any subsequence, Lemma 3.3 gives a further subsequence on which both normalized measures converge strongly:

$$\frac{T_{\omega_t}\eta_D}{\mathfrak{b}_t} \longrightarrow \mathcal{V} \neq 0, \qquad \frac{T_{\widetilde{\omega}_t}\eta_D}{\widetilde{\mathfrak{b}}_t} \longrightarrow \widetilde{\mathcal{V}} \neq 0.$$

Both limits are positive. Proposition 3.4 gives exponents $\lambda, \widetilde{\lambda}$ with

$$\frac{\log \mathfrak{b}_t}{L_t} \longrightarrow -\lambda, \qquad \frac{\log \widetilde{\mathfrak{b}}_t}{L_t} \longrightarrow -\widetilde{\lambda}.$$

First identify $\widetilde{\lambda}$. Put

$$\widetilde{\mathcal{Z}}_t = \int_D e^{-A_t K_1 - \varepsilon_t \sum_{j\geq 2} K_j}\, \mathrm{d}\mu, \qquad \mathcal{Z}_D^{(0)}(A) = \int_D e^{-AK_1}\, \mathrm{d}\mu.$$

Since the losses are bounded on D,

$$e^{-\varepsilon_t \sum_{j\geq 2} \|K_j\|_{\infty,D}} \leq \frac{\widetilde{\mathcal{Z}}_t}{\mathcal{Z}_D^{(0)}(A_t)} \leq 1.$$

The ratio tends to 1. The scalar Laplace asymptotics [5, 3] and Proposition 3.2 give $\log \mathcal{Z}_D^{(0)}(A_t)/L_t \to -\lambda_Z$. The function $h(x) = e^{-\sum_i x_i}$ on the quadrant has a Schwartz extension, so

$$\frac{\widetilde{\mathcal{Z}}_t}{\widetilde{\mathfrak{b}}_t} \longrightarrow \langle \widetilde{\mathcal{V}}, h\rangle \in (0, \infty).$$

Consequently $\widetilde{\lambda} = \lambda_Z$.

To compare the two normalizations, use the dilation estimate [2, equation (33)]. With

$$p_m(h) = \max_{|\alpha|\leq m} \sup_x (1 + \|x\|)^m |\partial^\alpha h(x)|, \qquad \|V\|_{-m} = \sup_{p_m(h)\leq 1} |\langle V, h\rangle|,$$

it gives

$$\|T_\delta V\|_{-m} \leq C_m e^{c_m \|\delta\|} \|V\|_{-m}.$$

Choose m large enough that both normalized families are bounded in this norm. Their nonzero limits also give positive lower bounds, hence

$$\|T_{\omega_t}\eta_D\|_{-m} \asymp \mathfrak{b}_t, \qquad \|T_{\widetilde{\omega}_t}\eta_D\|_{-m} \asymp \widetilde{\mathfrak{b}}_t.$$

The two measures are related by

$$T_{\widetilde{\omega}_t}\eta_D = T_{\widetilde{\omega}_t - \omega_t}(T_{\omega_t}\eta_D), \qquad T_{\omega_t}\eta_D = T_{\omega_t - \widetilde{\omega}_t}(T_{\widetilde{\omega}_t}\eta_D).$$

Applying the dilation estimate in both directions gives

$$C_m^{-1} e^{-c_m \|\omega_t - \widetilde{\omega}_t\|} \leq \frac{\|T_{\widetilde{\omega}_t}\eta_D\|_{-m}}{\|T_{\omega_t}\eta_D\|_{-m}} \leq C_m e^{c_m \|\omega_t - \widetilde{\omega}_t\|}.$$

Using the norm comparisons above and taking logarithms, we obtain

$$|\log \mathfrak{b}_t - \log \widetilde{\mathfrak{b}}_t| \leq C + c_m \|\omega_t - \widetilde{\omega}_t\| = o(L_t).$$

Thus $\lambda = \widetilde{\lambda} = \lambda_Z$. Since the initial subsequence was arbitrary, $\log \mathfrak{b}_t / L_t \to -\lambda_Z$ along the full sequence. □

Theorem 4.2. *Fix an admissible neighborhood D. Suppose $A_t \to \infty$ and (42) holds. Then*

$$\frac{\mathcal{Z}_{D,t}(c)}{\mathcal{Z}_{D,t}(1)} \longrightarrow c^{-\lambda_Z} \qquad (c>0), \tag{44}$$

locally uniformly in $c>0$.

Proof. Put $\mathcal{V}_t = \eta_{D,\boldsymbol{\tau}_t}/\mathfrak{b}_t = \mathcal{V}_{D,\boldsymbol{\tau}_t^{-1}}$. By Lemma 3.3, from any sequence $t\to\infty$ we can extract a strongly convergent subsequence $\mathcal{V}_t \to \mathcal{V} \neq 0$. The logarithmic direction is

$$\omega_t = (-\log A_t, -\log(n_{2,t}\beta_{2,t}), \dots, -\log(n_{\ell,t}\beta_{\ell,t})), \qquad \frac{\omega_t}{\|\omega_t\|} \longrightarrow -e_1. \tag{45}$$

By (43), $\log \mathfrak{b}_t/\|\omega_t\| \to -\lambda_Z$. Proposition 3.4 therefore identifies the homogeneity exponent as λ_Z, so

$$(x_1, x' \mapsto cx_1, x')_*\mathcal{V} = c^{-\lambda_Z}\mathcal{V} \qquad (c>0).$$

For $c>0$ let

$$h_c(x) = \exp\left(-cx_1 - \sum_{j=2}^{\ell} x_j\right).$$

On the quadrant, h_c has an ambient Schwartz extension: multiply by smooth cutoffs equal to one for nonnegative coordinates and zero below -1. Positivity gives $0<\langle \mathcal{V}, h_1\rangle<\infty$, and homogeneity gives

$$\langle \mathcal{V}, h_c\rangle = c^{-\lambda_Z}\langle \mathcal{V}, h_1\rangle. \tag{46}$$

Hence, along the selected subsequence,

$$\frac{\mathcal{Z}_{D,t}(c)}{\mathcal{Z}_{D,t}(1)} = \frac{\langle \mathcal{V}_t, h_c\rangle}{\langle \mathcal{V}_t, h_1\rangle} \longrightarrow c^{-\lambda_Z}.$$

Every sequence has such a further subsequence with the same limit. This proves (44), locally uniformly in $c>0$ since the extended h_c form bounded Schwartz families.

□

We combine Proposition 2.1 with the population partition ratios in Theorem 4.2. The empirical exponent includes all the original losses, including any identically zero population losses omitted from the Mellin analysis.

Proof of Theorem 1.4. Proposition 3.2 proves attainment and positivity of λ_Z. Choose a semianalytic set $D \subseteq V \cap D_*$ with $Z \subseteq \operatorname{int}_W D$. For every fixed $c>0$, (25) gives

$$\mathbb{E}\left|\frac{\widehat{\mathcal{Z}}_{W,t}(c)}{\mathcal{Z}_{D,t}(c)} - 1\right| \longrightarrow 0.$$

Thus, for every $\varepsilon>0$, Markov's inequality gives

$$\mathbb{P}\left(\left|\frac{\widehat{\mathcal{Z}}_{W,t}(c)}{\mathcal{Z}_{D,t}(c)} - 1\right| > \varepsilon\right) \le \frac{1}{\varepsilon}\mathbb{E}\left|\frac{\widehat{\mathcal{Z}}_{W,t}(c)}{\mathcal{Z}_{D,t}(c)} - 1\right| \longrightarrow 0.$$

The same conclusion at $c=1$, followed by continuity of the reciprocal at 1, yields

$$\frac{\widehat{\mathcal{Z}}_{W,t}(c)}{\mathcal{Z}_{D,t}(c)} \xrightarrow{\mathbb{P}} 1, \qquad \frac{\mathcal{Z}_{D,t}(1)}{\widehat{\mathcal{Z}}_{W,t}(1)} \xrightarrow{\mathbb{P}} 1.$$

Therefore, by (44) on the middle factor,

$$\begin{aligned}\frac{\widehat{\mathcal{Z}}_{W,t}(c)}{\widehat{\mathcal{Z}}_{W,t}(1)} &= \frac{\widehat{\mathcal{Z}}_{W,t}(c)}{\mathcal{Z}_{D,t}(c)}\frac{\mathcal{Z}_{D,t}(c)}{\mathcal{Z}_{D,t}(1)}\frac{\mathcal{Z}_{D,t}(1)}{\widehat{\mathcal{Z}}_{W,t}(1)} \\ &\xrightarrow{\mathbb{P}} c^{-\lambda_Z}.\end{aligned} \tag{47}$$

Thus the posterior exponential moments of the scaled empirical loss satisfy

$$M_t(s) := \langle e^{sA_tK_{1,t}}\rangle_t = \frac{\widehat{\mathcal{Z}}_{W,t}(1-s)}{\widehat{\mathcal{Z}}_{W,t}(1)} \xrightarrow{\mathbb{P}} (1-s)^{-\lambda_Z}, \qquad s<1. \tag{48}$$

Write $\rho_t = (A_tK_{1,t})_*\widehat{\Pi}_t$ for the Gibbs posterior of the scaled empirical loss, so that $\int_{\mathbb{R}} h\,\mathrm{d}\rho_t = \langle h(A_tK_{1,t})\rangle_t$. From any subsequence, extract a further subsequence along which (48) holds almost surely at every rational $s \in [-1/2, 1/2]$. Convexity and the continuity theorem for moment-generating functions then give almost sure weak convergence to γ_{λ_Z} along this further subsequence. Since the weak topology on $\mathcal{P}(\mathbb{R})$ is metrizable, the subsequence criterion gives weak convergence in probability.

Let $h \in \mathcal{H}$, choose C_h and η as in (11), and fix $\eta < \eta' < 1$. Let $0 \le \chi_R \le 1$ be a continuous cutoff equal to one on $[-R, R]$ and zero outside $[-R-1, R+1]$. Then

$$\begin{aligned}&\big\langle |h(A_tK_{1,t})|(1-\chi_R(A_tK_{1,t}))\big\rangle_t \\ &\qquad \le C_h e^{-(\eta'-\eta)R}\big(M_t(\eta') + M_t(-\eta')\big).\end{aligned} \tag{49}$$

The term in parentheses converges in probability to a finite constant by (48). The same tail estimate holds under γ_{λ_Z}. By weak convergence in probability, the posterior expectation of the bounded continuous function $h\chi_R$ converges to its integral under γ_{λ_Z}. Letting $R \to \infty$ gives (13). Taking $h(x) = x$ gives (14). □

Proof of Conjecture 1.3. Apply Theorem 1.4 with $K_1 = K$, $K_2 = R = \|w - w_*\|^2$, $t = n$, $n_{1,t} = n_{2,t} = n$, $\beta_{1,t} = \beta_n = n^{-3/4}$, and $\beta_{2,t} = (\log n)^2/n$, setting $\gamma_n = n\beta_{2,t} = (\log n)^2$. Realizability gives $\mathbb{E}e^{-sf(X,w)} \le 1$ for $0 \le s \le 1$, so (5) holds with $\theta = (0,1)$ and $C = 0$. Condition (6) holds since R has compact sublevel sets in W. The common zero set is $\{w_*\}$. The weights satisfy (12). □

References

[1] Edmund Lau, Zach Furman, George Wang, Daniel Murfet, and Susan Wei. *The Local Learning Coefficient: A Singularity-Aware Complexity Measure*. In *Proceedings of the 28th International Conference on Artificial Intelligence and Statistics*, Proceedings of Machine Learning Research **258** (2025), 244–252. `https://proceedings.mlr.press/v258/lau25a.html`.

[2] Brennan Dury. *Mellin Asymptotics under Anisotropic Dilations*. Repository manuscript, `MellinAsymptotics/MellinAsymptotics.tex`.

[3] Brennan Dury. *Meromorphic continuation and possible poles of multivariate absolute-value zeta functions on real analytic spaces*. Repository manuscript, `Zeta/Zeta.tex`.

[4] Brennan Dury. *Multivariate zeta functions, Laplace integrals, and singular learning theory*. Thesis manuscript, `archive/Thesis/Thesis.tex`.

[5] Sumio Watanabe. *A Widely Applicable Bayesian Information Criterion*. *Journal of Machine Learning Research* **14**(27) (2013), 867–897. `https://www.jmlr.org/papers/v14/watanabe13a.html`.