Regularly Parametrised Models

Thomas Waring

March 8, 2022

This talk introduces the RLCT of an ideal, which we can use, via *regularly parametrised models*, to simply calculations. The main reference is [Lin11], see §§1.4 and 4.1. Throughout, let $W \subset \mathbb{R}^N$ be compact and semianalytic, and $\mathcal{A}_W$ be the ring of (real) analytic functions on W.

Definition 1. Let $f, \varphi \in \mathcal{A}_W$ be non-negative analytic functions on W. The RLCT of f, with the prior φ, is the pair (λ, θ) of real numbers, such that the *partition function*:

$$Z(n) := \int_W e^{-nf(w)} \varphi(w)\, dw$$

obeys the asymptotic expression:

$$-\log Z(n) = \lambda \log n - (\theta - 1) \log \log n + O(1). \tag{1}$$

Notation: $\mathrm{RLCT}_W(f; \varphi) = (\lambda, \theta)$.

Remark 2. See [Lin11, §3] for the proof that this is well-defined. In fact λ is rational and θ a positive integer, and they can be respectively calculated as the smallest pole of

$$\zeta(z) = \int_W f(w)^{-z} \varphi(w)\, dw$$

and its multiplicity. We we write $(\lambda, \theta) < (\lambda', \theta')$, if $\lambda < \lambda'$ or $\lambda = \lambda'$ and $\theta > \theta'$, which is equivalent to

$$\lambda \log n - (\theta - 1) \log \log n < \lambda' \log n - (\theta' - 1) \log \log n$$

for all sufficiently large n.

Lemma 3. Let f, g be real analytic on W. If there is a positive constant c such that $f \leq cg$ on W, then

$$\mathrm{RLCT}_W(f; \varphi) \leq \mathrm{RLCT}_W(g; \varphi),$$

for any prior φ.

Proof. Observe that:

$$Z_f(n) = \int_W dw\, e^{-nf(w)} \geq \int_W dw\, e^{-cng(w)} = Z_g(cn),$$

so we have,

$$\begin{aligned}\lambda_f \log n - (\theta_f - 1)\log\log n \le \lambda_g \log(cn) - (\theta_g - 1)\log\log(cn) + O(1) \\ = \lambda_g \log n - (\theta_g - 1)\log\log n + \lambda_g \log c \\ - (\theta_g - 1)\log\left(1 + \frac{\log c}{\log n}\right) + O(1).\end{aligned}$$

The last two terms are $O(1)$, so we have, for sufficiently large n:

$$\lambda_f \log n - (\theta_f - 1)\log\log n \le \lambda_g \log n - (\theta_g - 1)\log\log n.$$

□

Corollary 4. If there are positive constants c, d such that $cg(w) \le f(w) \le dg(w)$, then $\mathrm{RLCT}_W(f;\varphi) = \mathrm{RLCT}_W(g;\varphi)$. Such functions are called *comparable*.

Corollary 5. Let $I = (f_1, \dots, f_s)$ and $J = (g_1, \dots, g_r)$ be ideals (with a choice of generators) of $\mathcal{A}_W$. If $I \subset J$, then

$$\mathrm{RLCT}_W(f_1^2 + \cdots + f_r^2;\varphi) \le \mathrm{RLCT}_W(g_1^2 + \cdots + g_r^2;\varphi).$$

Proof. Writing each f_i in terms of the g_j, we have:

$$f_i = \sum_{j=1}^{r} h_{ij} g_j,$$

for some $h_{ij} \in \mathcal{A}_W$. By the Cauchy-Schwartz inequality:

$$f_i^2 = \left(\sum_{j=1}^{r} h_{ij} g_j\right)^2 \le \left(\sum_{j=1}^{r} h_{ij}^2\right)\left(\sum_{j=1}^{r} g_j^2\right).$$

and so,

$$\sum_{i=1}^{s} f_i^2 \le \left(\sum_{i=1}^{s}\sum_{j=1}^{r} h_{ij}^2\right)\left(\sum_{j=1}^{r} g_j^2\right).$$

As the h_{ij} are analytic (continuous) on the compact set W, there exists a constant c so that,

$$\sup_{w\in W}\left(\sum_{i=1}^{s}\sum_{j=1}^{r} h_{ij}(w)^2\right) = c$$

and we win. □

The last corollary makes the next definition independent of the choice of (finitely many) generators for the ideal I.

Definition 6. Let $I = (f_1, \dots, f_r) \subset \mathcal{A}_W$ be an ideal. Then $\mathrm{RLCT}_W(I;\varphi)$ is defined to be the RLCT associated to the function $f_1^2 + \cdots + f_r^2$, with the prior φ.

Caution: this differs (for convenience) by a factor of 2 from the definition in [Lin11]. As a result, there is a discrepancy between the RLCT of a (non-negative) function, and of the ideal it generates:

$$\begin{aligned} \mathrm{RLCT}_W(f;\varphi) &= (\lambda,\theta), \\ \mathrm{RLCT}_W\big((f);\varphi\big) &= (\lambda/2,\theta)\,. \end{aligned}$$

(Proof: examine the zeta functions.)

Our definition is useful by the following theorem (which is [Lin11, Proposition 4.4]. In particular, the hypotheses are satisfied when $f(w) = D_{\mathrm{KL}}(q \mid\mid p(-|w))$, where $p : W \to \Delta Z$ parametrises probability distributions over some finite set Z. In this case, the fibre ideal is generated by the difference between the component probabilities of p and the true distribution q.

Theorem 7. Let $f : W \to \mathbb{R}$ be real analytic, and suppose that f factors through $u : W \to U$, where $U \subset \mathbb{R}^M$ is compact and semi-analytic.

$$\begin{array}{ccc} W & \xrightarrow{\;u\;} & U \\ & \searrow^{f} & \downarrow g \\ & & \mathbb{R} \end{array}$$

Let $\hat{w} \in W$ be a point with $f(\hat{w}) = 0$, and set $\hat{u} = u(\hat{w})$. If the Hessian of g is positive definite at $\hat{u}$, then there is a semi-analytic neighbourhood $W' \subset W$ of $\hat{w}$ so that $\mathrm{RLCT}_{W'}(f;\varphi) = \mathrm{RLCT}_{W'}(I;\varphi)$, where I is the *fibre ideal*, generated by the components of u:

$$I = (u_i - \hat{u}_i \,:\, i = 1,\dots,m).$$

Proof. Assume that $\hat{u} = 0 \in \mathbb{R}^M$. By the Morse lemma, there is a linear change of coordinates $T : \mathbb{R}^M \to \mathbb{R}^M$ so that $h = g \circ T^{-1} : V \to \mathbb{R}$ has the form:

$$h(v) = (v_1^2 + \cdots + v_m^2)(1 + \tilde{h}(v)),$$

where $V = T(U)$ and $\tilde{h}(0) = 0$. Shrinking to $V' \subset V$, assume that $\tilde{h}(V') \subset [-1/2, 1/2]$. Letting λ, μ be, respectively, the smallest and largest eigenvalues of T^tT, we therefore have:

$$\frac{\lambda}{2}(x_1^2 + \cdots + x_m^2) \le g(u) \le \frac{3\mu}{2}(x_1^2 + \cdots + x_m^2),$$

where the x_i are coordinates on $U' = T^{-1}(V')$. As such, on $W' = u^{-1}(U')$, f is comparable (in the sense of Corollary 4) to the function

$$u_1^2 + \cdots + u_m^2$$

which calculates $\mathrm{RLCT}_{W'}(I;\varphi)$. □

This is useful for two reasons. First, the functions u_i may well be simpler than the original f. For example, in the case of program synthesis on a Turing machine, they are polynomials (see thesis). Second, the ideal definition is more flexible. As well as the freedom to choose generators, it satisfies several other properties which simplify calculations: see here.

References

[Lin11] S. Lin. *Algebraic Methods for Evaluating Integrals in Bayesian Statistics*. PhD thesis, University of California, Berkeley, 2011.