Statistics 291: Lecture 5 (February 6, 2024)

Kac-Rice I: General Formula

Instructor: Mark Sellke

Scribe: Zad Chin

1 Addendum on $\nabla H_{N,p}(x)$

Recall that we are interested in

$$\max_{x\in S_N} H_{N,p}(x) \le \bar{M} \equiv N^{-(p-1)/2} \max_{x^{(1)},\cdots,x^{(p)}\in S_N} \langle G_N^{(p)}, x^{(1)} \otimes \cdots \otimes x^{(p)} \rangle \overset{\text{whp}}{\le} CN,$$

where $C = C(p)$ for some constant $C(p)$.

Proposition 1.1. *The following equality holds*

$$\sup_{x\in S_N} \left\| \nabla H_{N,p}(x) \right\| \le p\bar{M}/\sqrt{N} \overset{\text{whp}}{\le} pC\sqrt{N}.$$

Proof. By a simple exercise in tensor algebra, we obtain

$$\begin{aligned}
\langle \nabla H_{N,p}(x), y\rangle &= N^{-(p-1)/2} \frac{d}{dt} \langle G_N^{(p)}, (x+ty) \otimes \cdots \otimes (x+ty)\rangle \mid_{t=0} \\
&= N^{-(p-1)/2} \langle G_N^{(p)}, y \otimes x^{\otimes p-1} + x \otimes y \otimes x^{\otimes p-2} + \cdots + x^{\otimes p-1} \otimes y \rangle \\
&\le p\bar{M}.
\end{aligned}$$

□

A related exercise is to prove that

$$\max_{x\in S_N} \left\| \nabla^2 H_{N,p}(x) \right\|_{\text{op}} \overset{\text{whp}}{\le} C',$$

where the operator norm above means the maximum absolute eigenvalue and $C' = C'(p)$ is some constant.

2 Critical Points

In previous lectures, we observed that the overlaps are close to 1 for low temperature. Mathematically,

$$\mathbb{E}|R(x,\tilde{x})| \ge 1 - O(1/\beta),$$

where x and $\tilde{x}$ are i.i.d. from μ_β. This suggests that μ_β clusters are around (near-global) local optima. As such, we will discuss how to count critical points, local maxima, and so on. This is relevant for optimization.

2.1 The 1-Dimensional Case

The question we are interested in solving is the following:

Question 2.1. *Let $f\colon I \to \mathbb{R}$, for $I \subseteq \mathbb{R}$ an interval, be a random smooth function. Then, how many zeros does f have on average? In other words, what is $\mathbb{E}|zeros(f)|$?*

One guess might be the approximation

$$\mathbb{E}|\text{zeros}(f)| = \int_I \varphi_{f(u)}(0)\, du,$$

where $\varphi_{f(u)}$ is the probability density for the random variable $u \mapsto f(u)$. However, this fails since it is not even scale-invariant: clearly $\mathbb{E}|\text{zeros}(f)| = \mathbb{E}|\text{zeros}(2f)|$, but our approximation for these two quantities is different (one is half of the other!). A semi-rigorous derivation of the correct formula is as follows:

$$\begin{aligned}
\mathbb{E}|\text{zeros}(f)| &= \mathbb{E} \sum_{z\in \text{zeros}(f)} 1 \\
&\approx \mathbb{E} \sum_{z\in \text{zeros}(f)} \frac{1}{2\delta} \int_{z-\sqrt{\delta}}^{z+\sqrt{\delta}} 1_{|f(u)|\le\delta} \left|f'(u)\right| du \\
&= \mathbb{E}\frac{1}{2\delta} \int_I 1_{|f(u)|\le\delta} \left|f'(u)\right| du \\
&\overset{d\downarrow 0}{\approx} \int_I \varphi_{f(u)}(0)\mathbb{E}[|f'(u)| \mid f(u) = 0]\, du,
\end{aligned}$$

where the last equation follows from noting that

$$\mathbb{E}[1_{|f(u)|\le\delta}|f'(u)|] = \mathbb{P}[|f(u)| \le \delta]\mathbb{E}[|f'(u)| \mid |f(u)| \le \delta].$$

In fact, this is a big general theorem.

Theorem 2.2. *(Kac-Rice Formula) Call a gaussian process $f\colon I \to \mathbb{R}$ is nice if*

- *$f \in C'(I)$ a.s.;*
- *$\text{Var}[f(u)] > 0$ for all $u \in I$;*
- *There are (a.s.) no double roots (i.e. u such that $f(u) = f'(u) = 0$).*

Then, we have the equality

$$\mathbb{E}|zeros(f)| = \int_I \varphi_{f(u)}(0)\mathbb{E}[|f'(u)| \mid f(u) = 0]\, du,$$

and this is known as the Kac-Ric formula.

Remarks.

(a) The gaussian hypothesis is not really needed. However, it makes using the formula much easier because we know how to deal with conditioning by gaussians much better than even other simple continuous random variables, e.g. uniforms.

(b) The original paper Kac (1943) did a similar gaussian example. However, proving that this formula holds more universally is still considered modern research.

(c) An example of a gaussian process that is "not nice" is the following: let $f(u) = gu^2$, where $g \sim \mathcal{N}(0,1)$. Then, the Kac-Rice formula gets "confused" and yields $0\cdot\infty$ at $u = 0$.

Example. We will consider the example

$$f(u) = \sum_{i=0}^{\infty} g_i u^i, \quad I = [-a, a], \quad 0 < a < 1,$$

with the g_i i.i.d. $\mathcal{N}(0,1)$. Note that $(f(u), f'(u))$ is a centered gaussian with:

- $\mathbb{E}[f(u)^2] = 1 + u^2 + u^4 + \cdots = \frac{1}{1-u^2}$;
- $\mathbb{E}[f'(u)^2] = 1 + 4u^2 + 9u^4 + \cdots = \frac{1+u^2}{(1-u^2)^3}$;
- $\mathbb{E}[f(u)f'(u)] = u + 2u^3 + 3u^5 + \cdots = \frac{u}{(1-u^2)^2}$.

Moreover, $\varphi_{f(u)}(0) = \sqrt{\frac{1-u^2}{\pi}}$, and $\mathrm{Var}[f'(u) \mid f(u)] = \frac{1}{(1-u^2)^3}$ and so

$$\mathbb{E}[|f'(u)| \mid f(u) = 0] = \sqrt{\frac{2}{\pi}}(1-u^2)^{-3/2}.$$

Then, by the Kac-Rice formula,

$$\begin{aligned}
\mathbb{E}|\mathrm{zeros}(f)| &= \int_{-a}^{a} \varphi_{f(u)}(0)\mathbb{E}[|f'(u)| \mid f(u) = 0]\, du \\
&= \int_{-a}^{a} \sqrt{\frac{1-u^2}{\pi}}\sqrt{\frac{2}{\pi}}(1-u^2)^{-3/2}\, du \\
&= \frac{1}{\pi}\int_{-a}^{a} \frac{du}{1-u^2} \\
&= \frac{2\,\mathrm{arctanh}(a)}{\pi} \\
&\approx \frac{\log\left(\frac{1}{1-a}\right)}{\pi}.
\end{aligned}$$

2.2 The N-Dimensional Case

Remarks. In this case, "nice" implies that there are no x such that $f(x) = 0$ and $\det JF(x) = 0$.

Theorem 2.3. *In the N-dimensional case, we have the following two results.*

(a) *Let $f : \mathbb{R}^N \to \mathbb{R}^N$ be a nice gaussian process, and $U \subseteq \mathbb{R}^N$ an open set. Then,*

$$\mathbb{E}|\mathrm{zeros}_U(F)| = \int_U \varphi_{F(x)}(\vec{0})\mathbb{E}\left[|\det JF(x)| \mid F(x) = \vec{0}\right]\, dx.$$

(b) *Let $H \colon \mathbb{R}^N \to \mathbb{R}$ have a nice gradient. Then, the average number of critical points of H on U is*

$$\mathbb{E}\,|\mathrm{Crt}_U(H)| = \int_U \varphi_{\nabla H(x)}(\vec{0})\mathbb{E}\left[|\det \nabla^2 H(x)| \mid \nabla H(x) = \vec{0}\right]\, dx.$$

3 The Situation On S_N

Definition 3.1. We say that $x \in S_N$ is a *critical point* on S_N if $\nabla H_{N,p}(x)$ is parallel to x.

Definition 3.2. The spherical *gradient* is defined as

$$\nabla_{\mathrm{sph}} H_{N,p}(x) = P_x^{\perp}\nabla H_{N,p}(x),$$

where $P_x^{\perp} = \mathbb{I}_N - \frac{xx^T}{N}$ is a rank $N-1$ projection.

Now, we will attempt to find an appropriate Hessian on the sphere. A plausible guess for what this might be is the tangential Hessian

$$\nabla^2_{\tan} H_{N,p}(x) \equiv P_x^{\perp} \nabla^2 H_{N,p}(x) P_x^{\perp}.$$

However, this is a bad idea! Intuitively, due to the curvature of sphere, the Hessian will depend more on functions outside of the sphere than just on functions on the sphere. For an example, consider

$$\hat{H}_{N,p}(x) = H_{N,p}(x) + (\|x\|^2 - N)$$

Note that these two hamiltonians are exactly the same on S_N, but

$$\nabla^2_{\tan} \hat{H}_{N,p}(x) \neq \nabla^2_{\tan} H_{N,p}(x).$$

To derive the "morally as well as formally correct" Hessian, first consider the points

$$x = (\sqrt{N}, 0 \cdots, 0), \quad x^{(\epsilon)} = \left(\sqrt{N} - \frac{\epsilon^2}{2}\sqrt{N} + O(\epsilon^4), \epsilon, 0, \cdots, 0\right),$$

with $\epsilon \ll 1/N$. Let $\partial_{x_i} := \frac{\delta}{\delta x_i}$. Then, to second order,

$$H_{N,p}(x^{(\epsilon)}) = H_{N,p}(x) + \epsilon \partial_{x_2} H_{N,p}(x) + \frac{\epsilon}{2}\left(\partial^2_{x_2} H_{N,p}(x) - N^{-1/2} \partial_{x_1} H_{N,p}(x)\right).$$

Definition 3.3. The correct analogue of the Hessian on S_N is the *spherical Hessian*

$$\nabla^2_{\mathrm{sph}} H_{N,p}(x) = \nabla^2_{\tan} H_{N,p}(x) - \nabla_{\mathrm{rad}} H_{N,p}(x) P_x^{\perp},$$

where $P_x^{\perp}$ can be identified as $\mathbb{I}_{N-1}$, and

$$\nabla_{\mathrm{rad}} H_{N,p}(x) = \frac{1}{N} \langle \nabla H_{N,p}(x), x \rangle.$$

Moreover, for pure p-spin models,

$$\nabla_{\mathrm{rad}} H_{N,p}(x) = \frac{p H_{N,p}(x)}{N},$$

which can be verified directly from a gradient computation or observed from the fact that

$$H_{N,p}(x) = A(x) t^p$$

follows from homogeneity. Combining what we have learned, the big lesson is

$$\begin{aligned}
\mathbb{E}\left|\mathrm{Crt}_{S_N}(H_{N,p})\right| &= \int_{S_N} \varphi_{\nabla_{\mathrm{sph}} H_{N,p}(x)}(\vec{0}) \mathbb{E}\left[|\det \nabla^2_{\mathrm{sph}} H_{N,p}(x)| \mid \nabla_{\mathrm{sph}} H_{N,p}(x) = \vec{0} \right] d\mu^{\mathrm{Vol}}_{N-1}(x) \\
&= \mathrm{Vol}_{N-1}(S_N) \varphi_{\nabla_{\mathrm{sph}} H_{N,p}(x)}(\vec{0}) \mathbb{E}\left[|\det \nabla^2_{\mathrm{sph}} H_{N,p}(x)| \mid \nabla_{\mathrm{sph}} H_{N,p}(x) = \vec{0} \right].
\end{aligned}$$

Here, $d\mu^{\mathrm{Vol}}_{N-1}(x)$ is the non-normalized $(N-1)$ volume and

$$\mathrm{Vol}_{N-1}(S_N) = \frac{2\pi^{N/2} N^{(N-1)/2}}{\Gamma(N/2)} = (2\pi e)^{n/2} e^{o(N)}.$$

In our next class, we discuss how to compute the expectation in the formula with the following joint law.

Proposition 3.4. *The terms* $\left(H_{N,p}(x), \nabla_{sph} H_{N,p}(x), \nabla^2_{\tan} H_{N,p}(x)\right)$ *are independent, and moreover:*

- $H_{N,p}(x) \sim \mathcal{N}(0, N)$;
- $\nabla_{\mathrm{sph}} H_{N,p}(x) \sim \sqrt{p}\mathcal{N}(0, P_x^{\perp} \approx \mathbb{I}_{N-1})$;

- $\nabla^2_{\tan} H_{N,p}(x) \sim \sqrt{p(p-1)\frac{N-1}{N}}\mathrm{GOE}(N-1)$

where $\mathrm{GOE}(M)$ *is a* $M \times M$ *symmetric gaussian matrix* A *such that:*

- $A_{ij} \sim \mathcal{N}(0, 1/M)$, $i \neq j$;
- $A_{ii} \sim \mathcal{N}(0, 2/M)$;
- *The* A_{ij} *are independent for* $1 \leq i \leq j \leq M$.