Statistics 291: Lecture 3 (January 30, 2024)

Concentration-Enhanced Second Moment Method

Instructor: Mark Sellke

Scribe: Jarell Cheong Tze Wen

1 A Short Tensor Warm-up

Recall from last class that we defined

$$H_{N,p}(x) := N^{-(p-1)/2} \sum_{i_1,\ldots,i_p=1}^{N} g_{i_1\cdots i_p} x_{i_1} \cdots x_{i_p} = N^{-(p-1)/2} \langle G_N^{(p)}, x^{\otimes p} \rangle.$$

This second expression might seem a little mysterious if one has not seen tensors, so to begin, we will first say a few things about tensors and how they will be used in this class. To begin, despite all the definitions for tensors involving category theory and the universal property, tensors, for this class, simply mean vectors indexed by tuples of numbers $(i_1,\ldots,i_p) \in \{1,\ldots,N\}^p$. Then, inner product of tensors is simply given by

$$\langle S, T \rangle = \sum_{i_1,\ldots,i_p=1}^{N} s_{i_1\cdots i_p} t_{i_1\cdots i_p}$$

if S and T are two p-tensors. For more about tensors, refer to Hillar-Lim (2013), "Most Tensor Problems are NP-Hard." There is one fact about tensors that we will use many times, which is given below.

Proposition 1.1. *Define $x^{\otimes p}$ by $(x^{\otimes p})_{i_1\cdots i_p} = x_{i_1} \cdots x_{i_p}$. Then, $\langle x^{\otimes p}, y^{\otimes p} \rangle = \langle x, y \rangle^p$. In particular, we have*

$$\|x^{\otimes p}\| = \|x\|^p.$$

2 More on Second Moment Calculations

Recall, from last lecture, that our goal was to show that for small β,

$$\lim_{N\to\infty} \mathbb{E} F_N(\beta) = \frac{\beta^2}{2}.$$

We have previously established the upper bound

$$\begin{aligned}
\mathbb{E} F_N(\beta) &= \frac{1}{N} \mathbb{E} \log Z_N(\beta) \\
&\leq \frac{1}{N} \log \mathbb{E} Z_N(\beta)
\end{aligned}$$

$$= \frac{\beta^2}{2},$$

where the inequality is Jensen's inequality, and we defined the last term above to be $F_N^{\text{ann}}(\beta)$. First, we will explain the final equality above in slightly more detail than last lecture. This equality follows from two results, which we present below.

Lemma 2.1. *If $g \sim \mathcal{N}(0,1)$, then $\mathbb{E}e^{\lambda g} = e^{\lambda^2/2}$.*

Proof. This is easily obtained by completing the square as follows.

$$\mathbb{E}e^{\lambda g} = \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-u^2/2+\lambda u}\, du = e^{\lambda^2/2} \frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{(u-\lambda)^2/2}\, du = e^{\lambda^2/2}(1) = e^{\lambda^2/2}. \qquad \square$$

Lemma 2.2. $\mathbb{E}Z_N(\beta) = e^{\beta^2 N/2}$.

Proof. Interchanging the expectation and the integral, observe that

$$\mathbb{E}Z_N(\beta) = \int_{S_N} \mathbb{E}e^{\beta H_{N,p}(x)}\, dx.$$

For fixed $x \in S_N$, $H_{N,p}(x) \sim \mathcal{N}(0,N)$, so plugging $\lambda = \beta\sqrt{N}$ into Lemma 2.1, we get

$$\mathbb{E}e^{\beta H_{N,p}(x)} = e^{\beta^2 N/2}$$

for all $x \in S_N$. Thus, averaging over the integral, we get $\mathbb{E}Z_N(\beta) = e^{\beta^2 N/2}$. $\square$

With this, recall that we managed to compute the second moment

$$\begin{aligned}
\mathbb{E}[Z_N(\beta)^2] &= \mathbb{E}\int_{S_N}\int_{S_N} \exp(\beta H_{N,p}(x) + \beta H_{N,p}(\tilde{x}))\, dx d\tilde{x} \\
&= \exp\left(N \max_{-1 \le R \le 1} \left\{\beta^2(1+R^p) + \frac{1}{2}\log(1-R^2)\right\} + o(N)\right),
\end{aligned}$$

where $R(x,\tilde{x}) = \langle x, \tilde{x}\rangle / N$ is the overlap. Last lecture, for the sake of time, we used a quick geometric argument to show why the function

$$f_\beta(R) := \beta^2(1+R^p) + \frac{1}{2}\log(1-R^2)$$

was maximized at $\beta = 0$. We will prove this more formally now.

Proposition 2.3. *If $\beta \le \beta_0(p)$ is small and $p \ge 2$, then*

$$\max_{-1 \le R \le 1} f_\beta(R) = f_\beta(0) = \beta^2 = 2F_N^{\text{ann}}(\beta).$$

Proof. By easy computations, $f_\beta(0) = \beta^2$ and $f'_\beta(0) = 0$. Therefore, it suffices to show that $f''_\beta(0) \le 0$ for all $-1 \le R \le 1$. Indeed, first note that

$$f''_\beta(0) = \beta^2 p(p-1)R^{p-2} - \frac{1}{2}\left(\frac{1}{(1+R)^2} + \frac{1}{(1-R)^2}\right).$$

The first term above is bounded above by $\beta^2 p(p-1)$ since R is at most 1, while the second term is bounded below by $\frac{1}{2}$ since the larger of $1/(1+R)^2$ and $1/(1-R)^2$ is at least 1. So if $\beta \le 1/p$, it follows that

$$f''_\beta(0) \le \beta^2 p(p-1) - \frac{1}{2} \le 0. \qquad \square$$

Remark. We performed all these calculations for the spherical p-spin model, but these results also hold for the Ising p-spin model after some modifications. Recall that for the Ising p-spin model, the underlying measure is uniform on the Boolean cube, i.e.

$$\mu \sim \mathrm{Unif}(\{\pm 1\}^N).$$

Since the Boolean cube is contained in the sphere, the exact same calculation shows that

$$\mathbb{E}Z_N(\beta) = e^{\beta^2 N/2}.$$

Meanwhile, if we let h be the entropy function

$$h(q) = q\log(1/q) + (1-q)\log\left(\frac{1}{1-q}\right),$$

running through the second moment calculations instead shows that

$$\mathbb{E}[Z_N(\beta)^2] = \exp\left(N \max_{-1\le R\le 1}\left\{\beta^2(1+R^p) - h\left(\frac{1+R}{2}\right)\right\} + o(N)\right),$$

which is again just $\exp(\beta^2 N + o(N))$ for small β.

So far, our work with the second moment method is enough for us to conclude that for $\beta \le \beta_0$,

$$r_N(\beta) \coloneqq \frac{\mathbb{E}[Z_N(\beta)^2]}{\mathbb{E}[Z_N(\beta)]^2} \le e^{N\delta_N}$$

where δ_N is a deterministic sequence with $\lim_{N\to\infty}\delta_N = 0$. Now, to make further progress on the limiting expected free energy, we need two further insights: a form of "weak" Chebyshev inequality and the general technique of concentration of measure. We address the former first, leaving the latter to the next section.

Theorem 2.4 (Paley-Zygmund inequality)**.** *Let Z be a nonnegative random variable such that $\mathbb{E}Z > 0$ and $\mathbb{E}Z^2 < \infty$. Then, $\mathbb{P}[Z \ge \mathbb{E}Z/2] \ge \mathbb{E}[Z]^2/4\mathbb{E}Z^2$. Approximate equality holds for*

$$Z = \begin{cases} \mathbb{E}Z^2 & \textit{with probability } 1/\mathbb{E}Z^2, \\ 0 & \textit{otherwise,} \end{cases}$$

if we assume, without loss of generality, that $\mathbb{E}Z = 1$.

Proof. By rescaling if necessary, assume, without loss of generality, that $\mathbb{E}Z = 1$. Therefore,

$$\begin{aligned} 1 &= \mathbb{E}Z \\ &= \mathbb{E}[Z\cdot 1_{Z\le 1/2}] + \mathbb{E}[Z\cdot 1_{Z>1/2}] \\ &\le 1/2 + \sqrt{\mathbb{E}[Z^2]\mathbb{P}[Z>1/2]}, \end{aligned}$$

where in the last line we used the fact that $Z\cdot 1_{Z\le 1/2} \le 1/2$ almost surely and the Cauchy-Schwarz inequality. The desired inequality follows after some rearrangement. □

In particular, for $\beta \le \beta_0$, the Paley-Zygmund inequality yields

$$\mathbb{P}[Z_N(\beta) \ge e^{\beta^2 N/2}/2] \ge \frac{1}{4r_N(\beta)} \ge \frac{e^{-N\delta_N}}{4},$$

and moreover, the leftmost probability is equal to $\mathbb{P}[F_N(\beta) \ge \beta^2/2 - \log 2/N]$. Thus, we have a lower bound for the tail probability, and it remains to find an upper bound for the tail probability, which we will do next.

3 Concentration of Free Energy

Definition 3.1. A function $F\colon \mathbb{R}^d \to \mathbb{R}$ is L-Lipshitz if for all $G, \tilde{G} \in \mathbb{R}^d$,

$$|F(G) - F(\tilde{G})| \le L\|G - \tilde{G}\|.$$

In this section, we reduce the proof of our main result for this lecture to the following two lemmas.

Lemma 3.2. *$F_N(\beta) = F_N(\beta; G_N^{(\beta)})$ is $\beta/\sqrt{N}$-Lipshitz in $G_N^{(p)}$.*

Lemma 3.3. *If $F\colon \mathbb{R}^d \to \mathbb{R}$ is L-Lipshitz and $G \sim \mathcal{N}(0, \mathbb{I}_d)$, then*

$$\mathbb{P}[|F(G) - \mathbb{E}F(G)| \ge \lambda] \le 2e^{-\lambda^2/8L^2}.$$

In particular, the above probability is small once $\lambda \gg L$.

Corollary 3.4. *By Lemma 3.2 and Lemma 3.3,* $\mathbb{P}[|F_N(\beta) - \mathbb{E}F_N(\beta)| \ge \lambda] \le 2e^{-N\lambda^2/8\beta^2}$.

The statement of Lemma 3.3 should be surprising if this is the first time one sees it. After all, it is a statement about probabilities of random variables of arbitrary dimension that does not depend on the dimension! A naive proof attempt, just to get some intuition about why this result is so surprising, goes as follows: let

$$G, \tilde{G} \sim \mathcal{N}(0, \mathbb{I}_d)$$

be i.i.d. standard gaussians. Then, $G - \tilde{G} \sim \mathcal{N}(0, 2\mathbb{I}_d)$, so with high probability,

$$\|G - \tilde{G}\| \le O(\sqrt{d}).$$

Since F is L-Lipshitz, with high probability, we can reasonably expect that $|F(G) - F(\tilde{G})| \le O(L\sqrt{d})$. However Lemma 3.3 instead states that F fluctuates at the much smaller scale $O(L)$.

Remark. The "least concentrated" Lipshitz functions are linear.

Remark. In this class, we shall apply Lemma 3.3 to all kinds of Lipshitz functions F. Besides $F_N(\beta)$, we can take F to be the eigenvalues of matrices, or even the performance of optimization algorithms, say for optimizing p-spin models.

If we assume both Lemma 3.2 and Lemma 3.3 for now, we can finish the main proof about the limiting value of $F_N(\beta)$. Recall that up to this point, we know that $\mathbb{E}F_N(\beta) \le \beta^2/2$ for all N.

Proposition 3.5. *The inequality* $\limsup_{N\to\infty} \mathbb{E}F_N(\beta) \ge \beta^2/2$ *holds, and thus* $\lim_{N\to\infty} \mathbb{E}F_N(\beta) = \beta^2/2$.

Proof. Fix $\epsilon > 0$, and suppose, for contradiction, that for N arbitrarily large,

$$\mathbb{E}F_N(\beta) \le \frac{\beta^2}{2} - \epsilon.$$

Then, we obtain the following string of inequalities:

$$\begin{aligned} 2e^{-\epsilon^2 N/32\beta^2} &\ge \mathbb{P}[|F_N(\beta) - \mathbb{E}F_N(\beta)| \ge \epsilon/2] \\ &\ge \mathbb{P}[F_N(\beta) \ge \beta^2/2 - \log 2/N] \\ &\ge e^{-N\delta_N}/4. \end{aligned}$$

The first inequality follows from setting $\lambda = \epsilon/2$ in Corollary 3.4, the second inequality follows from observing, say by drawing a number line with all the relevant values, that the event in the former probability contains (as subsets of $\mathbb{R}$) the event in the latter probability, and the third inequality follows from our work near the end of the previous section. However, for sufficiently large N and $\delta_N \to 0$, we see that

$$2e^{-\epsilon^2 N/32\beta^2} \le e^{-N\delta_N}/4,$$

which contradicts the string of inequalities above. $\square$

Remark. To gain a feeling for the types of arguments used in the proofs above, try out the problem in the first homework about the replica method, which says that $\mathbb{E}[Z_N(\beta)^{\epsilon/N}]$ "recovers the free energy."

Proof of Lemma 3.2. Fix $(i_1,\ldots,i_p) \in \{1,\ldots,N\}^p$. Then, a direct computation shows that

$$\begin{aligned}
\frac{\partial}{\partial g_{i_1\cdots i_p}} F_N(\beta) &= \frac{\partial}{\partial g_{i_1\cdots i_p}} \frac{1}{N} \log \int_{S_N} e^{\beta H_{N,p}(x)}\, dx \\
&= \left(N \int_{S_N} e^{\beta H_{N,p}(x)}\, dx \right)^{-1} \frac{\partial}{\partial g_{i_1\cdots i_p}} \int_{S_N} e^{\beta H_{N,p}(x)}\, dx \\
&= \left(N^{(p-1)/2} N \int_{S_N} e^{\beta H_{N,p}(x)}\, dx \right)^{-1} \beta \int_{S_N} x_{i_1} \cdots x_{i_p} e^{\beta H_{N,p}(x)}\, dx \\
&= \frac{\beta}{N^{(p+1)/2}} \mathbb{E}^{x\sim\mu_\beta}[x_{i_1}\cdots x_{i_p}].
\end{aligned}$$

Therefore, the gradient of $F_N(\beta)$ is given by

$$\nabla_{G_N^{(p)}} F_N(\beta) = \frac{\beta}{N^{(p+1)/2}} \mathbb{E}^{x\sim\mu_\beta}[x^{\otimes p}].$$

Now, for each $x \in S_N$, note that we have $\|x^{\otimes p}\| = \|x\|^p = N^{p/2}$, so by Jensen's inequality (norms are convex),

$$\|\mathbb{E}^{x\sim\mu_\beta} x^{\otimes p}\| \le N^{p/2}.$$

This means that the gradient of $F_N(\beta)$ is bounded above by $\beta/\sqrt{N}$, as desired:

$$\left\| \nabla_{G_N^{(p)}} F_N(\beta) \right\| = \left\| \frac{\beta}{N^{(p+1)/2}} \mathbb{E}^{x\sim\mu_\beta}[x^{\otimes p}] \right\| \le \frac{\beta}{N^{(p+1)/2}} N^{p/2} \le \frac{\beta}{\sqrt{N}}. \qquad \square$$

Proof of Lemma 3.3. By smoothing, e.g. using convolution, we may assume that $F \in C^1(\mathbb{R}^d)$. Now, we will use the interpolation method: let

$$G_0, G_{\pi/2} \sim \mathcal{N}(0, \mathbb{I}_d)$$

be i.i.d. standard gaussians. Consider the "path" from G_0 to $G_{\pi/2}$ given by

$$G_\theta = \cos(\theta) G_0 + \sin(\theta) G_{\pi/2}.$$

Let $\tilde{G}_\theta := \frac{d}{d\theta} G_\theta = -\sin(\theta) G_0 + \cos(\theta) G_{\pi/2}$. By the fundamental theorem of calculus and the chain rule,

$$F(G_{\pi/2}) - F(G_0) = \int_0^{\pi/2} \frac{d}{d\theta} F(G_\theta)\, d\theta = \int_0^{\pi/2} \langle \nabla F(G_\theta), \tilde{G}_\theta \rangle\, d\theta. \tag{1}$$

For each θ, G_θ and $\tilde{G}_\theta$ are independent standard gaussians. It is clear that they are both standard gaussians since they are linear combinations of such with weights adding to one. Independence holds because

$$\mathbb{E}[\langle G_\theta, \tilde{G}_\theta \rangle] = \cos\theta \sin\theta (1-1) = 0.$$

By this independence, we get the bound

$$\mathbb{E}[\exp(t\langle \nabla F(G_\theta), \tilde{G}_\theta \rangle)] \le \sup_{G_\theta \in \mathbb{R}^d} \mathbb{E}[\exp(t\langle \nabla F(G_\theta), \tilde{G}_\theta \rangle) \mid G_\theta] = \sup_{G_\theta \in \mathbb{R}^d} \mathbb{E}[\exp(t^2 \|\nabla F(G_\theta)\|^2/2)] \le \exp(t^2 L^2/2).$$

Next lecture, we will finish by applying Jensen's inequality on (1) to bound $\mathbb{E}[\exp(t(F(G_{\pi/2}) - F(G_0)))]$. $\square$