Lecture 12

Lecture date: September 24, 2007 *Scribe: Richard Liang*

1 Stein's method for concentration inequalities

The purpose of this lecture will be to prove the following theorem.

Theorem 1 *Suppose you have an exchangeable pair (X, X') of random objects. Suppose f and F are two functions such that*

(1) $F(X, X') = -F(X', X)$ a.s.; and

(2) $\mathbf{E}[f(X, X') \mid X] = f(X)$ a.s.

Let $v(X) = \frac{1}{2}\mathbf{E}\big[|(f(X) - f(X'))F(X, X')| \,\big|\, X\big]$. Then

(a) $\mathbf{E}[f(X)] = 0$ and

$$\begin{aligned} \mathrm{Var} f(X) &= \frac{1}{2}\mathbf{E}[(f(X) - f(X'))F(X, X')] \\ &\leq \mathbf{E}[v(X)]. \end{aligned} \tag{1}$$

(b) Suppose $\mathbf{E}(e^{\theta f(X)}|F(X, X')|)$ is finite for all θ. If B and C are constants such that $v(X) \leq Bf(X) + C$ a.s., then

$$\mathbf{P}\{|f(X)| > t\} \leq 2\exp\left(-\frac{t^2}{2Bt + 2C}\right). \tag{2}$$

(c) $\mathbf{E}\big[f(X)^{2k}\big] \leq (2k-1)^k \mathbf{E}\big[v(X)^k\big]$ for all $k \in \mathbb{N}$.

Exercise 2 *(You may be able to do this after the proof.)*

Extend (c) to all real $k > 1/2$. (We think that $(2k-1)^k$ remains unchanged for $k \geq 1$ but are not sure for $1/2 < k < 1$.)

Proof: First, note that $\mathbf{E}[f(X)] = \mathbf{E}[F(X,X')] = 0$ since (X,X') is an exchangeable pair and F is antisymmetric.

Further, we will assume

$$\begin{aligned} \mathbf{E}[v(X)] &< \infty \text{ for part (a); and} \quad (3) \\ \mathbf{E}\Big[e^{\theta f(X)}|F(X,X')|\Big] &< \infty \text{ for all } \theta \text{ in part (b).} \quad (4) \end{aligned}$$

We start by showing (a).

$$\begin{aligned} \text{Var} f(X) &= \mathbf{E}\big[f(X)^2\big] \\ &= \mathbf{E}\big[f(X)F(X,X')\big] \\ &= \mathbf{E}\big[f(X')F(X',X)\big] \text{ since } (X,X') \text{ exchangeable} \\ &= -\mathbf{E}\big[f(X')F(X,X')\big] \text{ by antisymmetry of } F \\ &= \frac{1}{2}\mathbf{E}\big[\big(f(X)-f(X')\big)\,F(X,X')\big] \\ &\leq \mathbf{E}[v(X)], \end{aligned}$$

which proves (a).

(b): Let $m(\theta) = \mathbf{E}\big[e^{\theta f(X)}\big]$; then $m'(\theta) = \mathbf{E}\big[f(X)e^{\theta f(X)}\big]$. Write this as

$$\begin{aligned} m'(\theta) &= \mathbf{E}\Big[F(X,X')e^{\theta f(X)}\Big] \\ &= \frac{1}{2}\mathbf{E}\Big[F(X,X')\left(e^{\theta f(X)} - e^{\theta f(X')}\right)\Big] \end{aligned}$$

via the antisymmetry of F and the exchangeability of (X,X').

We'll use the inequality

$$|e^x - e^y| \leq \frac{1}{2}|x-y|(e^x+e^y). \quad (5)$$

To see this, suppose $y < x$;

$$\begin{aligned} e^x - e^y &= \int_0^1 \frac{d}{dt}\left(e^{tx+(1-t)y}\right) dt \\ &= (x-y)\int_0^1 e^{tx+(1-t)y}\, dt \\ &\leq (x-y)\int_0^1 (te^x + (1-t)e^y)\, dt \text{ by Jensen's inequality} \\ &= (x-y)\frac{1}{2}\left(e^x+e^y\right), \end{aligned}$$

and similarly for $y \geq x$.

Now

$$\begin{aligned}
|m'(\theta)| &= \Big|\mathbf{E}\Big[F(X,X')e^{\theta f(X)}\Big]\Big| \\
&\le \frac{1}{2}\Big|\mathbf{E}\Big[F(X,X')\left(e^{\theta f(X)} - e^{\theta f(X')}\right)\Big]\Big| \\
&\le \frac{|\theta|}{4}\mathbf{E}\Big[\big|F(X,X')\left(f(X)-f(X')\right)\big|\left(e^{\theta f(X)} + e^{\theta f(X')}\right)\Big] \\
&\le \frac{|\theta|}{2}\mathbf{E}\Big[\big|F(X,X')\left(f(X)-f(X')\right)\big|\, e^{\theta f(X)}\Big] \text{ by exchangeability of } (X,X') \\
&= |\theta|\mathbf{E}\Big[v(X)e^{\theta f(X)}\Big].
\end{aligned}$$

If $v(X) \le Bf(X) + C$, then the above gives

$$\begin{aligned}
|m'(\theta)| &\le |\theta|\left(B\mathbf{E}\Big[f(X)e^{\theta f(X)}\Big] + C\mathbf{E}\Big[e^{\theta f(X)}\Big]\right) \\
&= B|\theta|m'(\theta) + C|\theta|m(\theta).
\end{aligned}$$

Now, m is a convex function, with $\mathbf{E}[f(X)] = m'(0) = 0$ and taking the value $m(0) = 1$ at its minimum. Suppose $0 < \theta < 1/B$; then

$$m'(\theta)(1 - B\theta) \le C\theta m(\theta).$$

Thus

$$\frac{d}{d\theta}\log m(\theta) = \frac{m'(\theta)}{m(\theta)} \le \frac{C\theta}{1 - B\theta}$$

for $0 < \theta < 1/B$. Therefore,

$$\begin{aligned}
\log m(\theta) &= \int_0^\theta \frac{d}{dt}\log m(t)\,dt \\
&\le \int_0^\theta \frac{Ct}{1 - Bt}\,dt \\
&\le \frac{1}{1 - B\theta}\int_0^\theta Ct\,dt \\
&= \frac{C\theta^2}{2(1 - B\theta)}.
\end{aligned}$$

So, for $\theta > 0$,

$$\begin{aligned}
\mathbf{P}\{f(X) \ge t\} &= \mathbf{P}\Big\{e^{\theta f(X)} \ge e^{\theta t}\Big\} \\
&\le e^{-\theta t}m(\theta) \\
&\le \exp\left(-\theta t + \frac{C\theta^2}{2(1 - B\theta)}\right) \text{ if } 0 < \theta < 1/B.
\end{aligned}$$

Taking

$$\theta = \frac{t}{C + Bt} \in \left(0, \frac{1}{B}\right),$$

we get the desired upper bound, and $\mathbf{P}(f(X) \leq -t)$ can be bounded similarly.

(c): Using a similar manipulation to that in the proof of (a),

$$\begin{aligned}
\mathbf{E}\Big[f(X)^{2k}\Big] &= \mathbf{E}\Big[f(X)^{2k-1}F(X,X')\Big] \\
&= \frac{1}{2}\mathbf{E}\Big[\Big(f(X)^{2k-1} - f(X')^{2k-1}\Big)\,F(X,X')\Big]. \qquad (6)
\end{aligned}$$

Also, similarly to (5), we can show

$$\Big|x^{2k-1} - y^{2k-1}\Big| \leq \frac{2k-1}{2}|x-y|\,\Big|x^{2k-2} + y^{2k-2}\Big|.$$

Plugging this into (6), we get

$$\mathbf{E}\Big[f(X)^{2k}\Big] \leq (2k-1)\mathbf{E}\Big[v(X)f(X)^{2k-2}\Big].$$

Applying Hölder's inequality with $1/p = 1 - 1/k$ and $1/q = 1/k$ gives

$$\mathbf{E}\Big[f(X)^{2k}\Big] \leq (2k-1)\left(\mathbf{E}\Big[f(X)^{2k}\Big]\right)^{(k-1)/k}\left(\mathbf{E}\Big[v(X)^k\Big]\right)^{1/k}$$

and so

$$\left(\mathbf{E}\Big[f(X)^{2k}\Big]\right)^{1/k} \leq (2k-1)\left(\mathbf{E}\Big[v(X)^k\Big]\right)^{1/k}.$$

This completes the proof of (c). □

Example 3 *Suppose we have the Curie-Weiss model on n spins: for $\sigma = (\sigma_1, \ldots, \sigma_n)$,*

$$\mathbf{P}\{\sigma\} = \frac{1}{Z_\beta}\exp\left(\frac{\beta}{n}\sum_{1\leq i<j\leq n}\sigma_i\sigma_j\right).$$

Let

$$m(\sigma) = \frac{1}{n}\sum_i \sigma_i.$$

Construct X' by taking one step in the Gibbs sampler (also known as the Glauber dynamics). Set $F(X,X') = \sigma_I - \sigma'_I$ where I is the updated index. Then

$$\begin{aligned}
f(X) &= \mathbf{E}\big[F(X,X')\,|\,X\big] \\
&\approx m(\sigma) - \tanh(\beta m(\sigma)).
\end{aligned}$$

We'll do this in the next lecture.