Lecture 9 Concentration of functions of random variables.

Next, we will look at concentration of $f(x_1, \dots, x_n)$ where f is a "well-behaved" function and $x_1, \dots, x_n$ are independent r.v.'s.

Bounded differences inequality (McDiarmid)

So far, we have focused on concetration of the average $\frac{1}{n}\sum_{i=1}^{n} x_i$:

Useful insight:

As long as $f(x_1, \dots, x_n)$ depends weakly on individual x_i, concentration holds!

<u>Thm</u>: (McDiarmid) Suppose that $f: \mathcal{X}^n \to \mathbb{R}$ has the bounded difference property: $\exists L_1, \dots, L_n$ such that

$$|f(x_1, \dots, x_k, \dots, x_n) - f(x_1, \dots, x_k', \dots, x_n)| \le L_k \quad \forall x, x' \in \mathcal{X}^n$$

that differ only in k'th entry.

Then for independent rv's $X = (X_1, \dots, X_n)$ have

$$\mathbb{P}[|f(X) - \mathbb{E}f(X)| > t] \le 2e^{-\frac{2t}{\|L\|_2^2}}$$

<u>pf</u>: We will use the Martingale method.

Define

$$Y_0 = \mathbb{E}f(X) \quad \text{and} \quad Y_i = \mathbb{E}[f(X) | X_1, \dots, X_i] \ \forall i$$

Observe

$$Y_i = Y_0 + \sum_{j=0}^{i-1} \underbrace{(Y_{j+1} - Y_j)}_{=: D_{j+1}} = Y_0 + \sum_{j=1}^{i} D_j$$

and

$$\mathbb{E}[Y_i | X_1, \dots, X_{i-1}] = \mathbb{E}[f(X) | X_1, \dots X_{i-1}] = Y_{i-1}$$

$$\Rightarrow \mathbb{E}[D_{j+1} | X_1, \dots, X_j] = 0$$

$$\Rightarrow \mathbb{E}[e^{\lambda(f(X) - \mathbb{E}f(X))}] = \mathbb{E}[e^{\lambda(Y_n - Y_0)}]$$

$$\mathbb{E}\left[e^{\lambda(f(X)-\mathbb{E}f(X))}\right] = \mathbb{E}\left[e^{\lambda(Y_n - Y_0)}\right]$$

$$= \mathbb{E}\left[e^{\lambda \sum_{j=1}^{n} D_j}\right]$$

$$= \mathbb{E}\left[e^{\lambda(Y_{n-1}-Y_0)} e^{\lambda D_n}\right]$$

$$= \mathbb{E}\left[e^{\lambda(Y_{n-1}-Y_0)} \underbrace{\mathbb{E}\left[e^{\lambda D_n} \mid X_1, \ldots, X_{n-1}\right]}\right]$$

Let X' differ from X in $X_i' \overset{\text{i.i.d.}}{\sim} X_i$. Then

$$\mathbb{E}\left[e^{\lambda D_i} \mid X_1, \ldots, X_{i-1}\right] = \mathbb{E}\left[e^{\lambda(Y_i - Y_{i-1})} \mid X_1, \ldots, X_{i-1}\right]$$

$$= \mathbb{E}\left[e^{\lambda \mathbb{E}\left[f(X)-f(X') \mid X_1, \ldots, X_i\right]} \;\middle|\; X_1, \ldots, X_{i-1}\right]$$

Jensen

$$\leq \mathbb{E}\left[e^{\lambda (f(X)-f(X'))} \mid X_1, \ldots, X_{i-1}\right]$$

bounded by L_i

$$\Rightarrow \mathbb{E}\left[e^{\lambda(f(X)-f(X'))} \mid X_1, \ldots, X_{i-1}\right] \leq e^{\frac{\lambda^2 L_i^2}{8}}$$

So $\mathbb{E}\left[e^{\lambda(f(X)-\mathbb{E}f(X))}\right] \leq e^{\frac{\lambda^2}{8}\|L\|^2}$. Apply Chernoff!

$\square$

Lipschitz transformation of Gaussians

Thm: Let $X_1, \dots, X_n \overset{i.i.d.}{\sim} N(0,1)$ and let $F: \mathbb{R}^n \to \mathbb{R}$ be L-Lipschitz:

$$|F(x) - F(y)| \le L \|x-y\|_2 \quad \forall x, y \in \mathbb{R}^n$$

Then

$F(X) - \mathbb{E}F(X)$ is $\frac{\pi L}{\sqrt{2}}$-subGaussian.

We'll need the following:

Exercise:
Suppose that (X, Y) are jointly normal.
Then
X and Y are independent iff
$\mathbb{E}[XY] = (\mathbb{E}X)(\mathbb{E}Y)$

pf of thm:

We can assume w/log:

- $L=1$
- $\mathbb{E}F(X_1,\dots,X_n)=0$
- F is C^1-smooth (otherwise approximate)

Symmetrization:

Let Y be an independent realization of X. Then

$$\mathbb{E}\exp(\lambda F(X)) = \mathbb{E}\exp(\lambda F(X) - \mathbb{E}_Y F(Y))$$

$$\text{Jensen} \rightarrow \quad \leq \mathbb{E}\exp(\lambda(F(X)-F(Y)))$$

Let's write:

$$F(X)-F(Y) = \int_0^{\pi/2} (F\circ\gamma)'(\theta)\, d\theta$$

where

$$\gamma(\theta) := y\cos(\theta) + X\sin(\theta)$$

Note

$$\dot{\gamma}(\theta) = -y\sin(\theta) + X\cos(\theta)$$

So

$(\gamma(\theta), \dot{\gamma}(\theta))$ jointly normal with $\mathrm{Cor}(\gamma(\theta), \dot{\gamma}(\theta)) = 0$

independent

So

$$\exp(\lambda(F(X) - F(y))) = \exp\left(\frac{1}{\pi/2}\int_0^{\pi/2} \frac{\pi}{2}\lambda (F\circ\gamma)'(\theta)\, d\theta\right)$$

$$\overset{\text{Jensen}}{\leq} \frac{2}{\pi}\int_0^{\pi/2} \exp\left(\frac{\pi}{2}\lambda (F\circ\gamma)'(\theta)\right) d\theta$$

So

$$\mathbb{E}\exp(\lambda(F(X)-F(Y)))$$

$$\le \frac{2}{\pi}\int_0^{\pi/2} \mathbb{E}\exp\Big(\frac{\pi}{2}\lambda\langle \nabla F(\underbrace{\gamma(\theta)}_{V}), \underbrace{\dot\gamma(\theta)}_{W}\rangle d\theta\Big)\, d\theta$$

independent

So (V,W) is jointly normal and independent

And

$$\mathbb{E}[\langle \nabla F(V), W\rangle^2 | V] = \operatorname{Var}_W\Big(\sum_i \frac{\partial F}{\partial x_i}(V) W_i\Big)$$

$$= \|\nabla F(V)\|^2 \le 1$$

So $\mathbb{E}\langle \nabla F(V), W\rangle^2 \le 1$

Therefore we deduce

$$\mathbb{E}\Big[\exp\Big(\frac{\pi}{2}\lambda\langle \nabla F(V), W\rangle\Big) \Big| V\Big] \le \exp\Big(\frac{\pi^2\lambda^2 L^3}{24}\Big)$$

Remove conditioning on V with tower rule. $\boxtimes$