Spring 2018: STA 6448
Advanced Probability and Inference II
Lecture 5

Yun Yang

- Concentration inequality

Gaussian concentration: Proof

We prove the theorem with a weaker constant in the exponent. In addition, we may assume f to be differentiable (why?).

Lemma

Suppose $f : \mathbb{R}^n \to \mathbb{R}$ is differentiable, then for any convex function $\phi : \mathbb{R} \to \mathbb{R}$, we have

$$\mathbb{E}\big[\phi\big(f(X) - \mathbb{E}[f(X)]\big)\big] \le \mathbb{E}\big[\phi\big(\frac{\pi}{2}\langle \nabla f(X), Y\rangle\big)\big],$$

where $X, Y \sim \mathcal{N}(0, I_n)$ are independent standard n-dim Gaussians.

Gaussian concentration: Proof

Now, we prove the theorem using this lemma.

Let $Y = (Y_1, \dots, Y_n)$ be i.i.d. copies of X_k's. For fixed $\lambda \in \mathbb{R}$, apply the lemma with $\phi(t) = e^{\lambda t}$, we obtain

$$
\begin{aligned}
&\mathbb{E}\big[\exp\big(\lambda\big\{f(X) - \mathbb{E}[f(X)]\big\}\big)\big] \\
\leq &\,\mathbb{E}\big[\exp\big(\frac{\pi\lambda}{2}\sum_{k=1}^{n} Y_k \frac{\partial f}{\partial x_k}(X)\big)\big] \\
= &\,\mathbb{E}_X\big[\exp\big(\frac{\pi^2\lambda^2}{8}\|\nabla f(X)\|^2\big)\big] \leq \exp\big(\frac{\pi^2\lambda^2}{8}L^2\big).
\end{aligned}
$$

Therefore, $f(X) - \mathbb{E}[f(X)]$ is sub-Gaussian with parameter $\pi L/2$, and

$$
\mathbb{P}\Big[\big|f(X) - \mathbb{E}[f(X)]\big| \geq t\Big] \leq 2\,e^{-\frac{2t^2}{\pi^2 L^2}} \quad \text{for all } t > 0.
$$

Lemma: Proof

Will apply the Slepian smart path interpolation:

$$Z_k(\theta) = X_k \sin\theta + Y_k \cos\theta, \quad \text{for } \theta \in [0, \frac{\pi}{2}] \text{ and } k = 1, 2, \ldots, n.$$

Observe that $Z_k(0) = Y_k$, $Z_k(1) = X_k$, and $(Z_k(\theta), Z_k'(\theta))$ are independent standard Gaussian variables.

By the convexity of ϕ, we have

$$\mathbb{E}_X\big[\phi\big(f(X) - \mathbb{E}_Y[f(Y)]\big)\big] \le \mathbb{E}_{X,Y}\big[\phi\big(f(X) - f(Y)\big)\big].$$

Notice that

$$f(X) - f(Y) = f(Z(1)) - f(Z(0)) = \int_0^{\frac{\pi}{2}} \langle \nabla f(Z(\theta)), Z'(\theta)\rangle \, d\theta$$

Lemma: Proof

Therefore,

$$\begin{aligned}
\mathbb{E}_{X,Y}\big[\phi\big(f(X)-f(Y)\big)\big] &= \mathbb{E}_{X,Y}\big[\phi\big(\frac{2}{\pi}\int_0^{\frac{\pi}{2}}\frac{\pi}{2}\langle\nabla f(Z(\theta)),\, Z'(\theta)\rangle\, d\theta\big)\big] \\
&\overset{(i)}{\leq} \frac{2}{\pi}\int_0^{\frac{\pi}{2}}\mathbb{E}_{X,Y}\big[\phi\big(\frac{\pi}{2}\langle\nabla f(Z(\theta)),\, Z'(\theta)\rangle\big)\big]\, d\theta \\
&= \mathbb{E}\big[\phi\big(\frac{\pi}{2}\langle\nabla f(\widetilde{X}),\, \widetilde{Y}\rangle\big)\big],
\end{aligned}$$

for independent standard n-dim Gaussian variables $(\widetilde{X}, \widetilde{Y})$, where step (i) follows by the convexity of ϕ.

Example: χ^2 variable revisit

Consider chi-squared random variable

$$Y = \sum_{k=1}^{n} Z_k^2, \qquad Z_k \overset{iid}{\sim} \mathcal{N}(0,1).$$

In this example, we consider an alternative approach to obtain the χ^2 concentration inequality.

Define $V = \sqrt{Y}/\sqrt{n} = \|Z\|_2/\sqrt{n}$. Since Euclidean norm is 1-Lipschitz, Gaussian concentration implies

$$\mathbb{P}\big[V - \mathbb{E}[V] \geq \delta\big] \leq e^{-n\,\delta^2/2} \quad \text{for all } \delta > 0.$$

Moreover, we have $\mathbb{E}[V] \leq \sqrt{\mathbb{E}[V^2]} = 1$. Therefore,

$$\mathbb{P}\Big[\frac{Y}{n} \geq (1+\delta)^2\Big] \leq e^{-n\,\delta^2/2}, \quad \text{or} \quad \mathbb{P}\big[Y \geq n\,(1+3t)\big] \leq e^{-n\min(t,\,t^2)/2}.$$

Example: Order statistics

Given a random vector $(X_1, X_2, \ldots, X_n)$, its order statistics are obtained by re-ordering its components in a non-increasing manner,

$$X_{(1)} \geq X_{(2)} \geq \cdots \geq X_{(n)}.$$

In particular, $X_{(1)} = \max_k X_k$ and $X_{(n)} = \min_k X_k$.

It can be shown (leave as an exercise) that $|X_{(k)} - Y_{(k)}| \leq \|X - Y\|_2$ for any $k = 1, \ldots, n$. Therefore, each order statistics is a 1-Lipschitz function. When X is a Gaussian random vector, then

$$\mathbb{P}\big[|X_{(k)} - \mathbb{E}[X_{(k)}]| \geq t\big] \leq 2\, e^{-t^2/2} \quad \text{for all } t > 0.$$

Example: Gaussian complexity

This example is related to the previous example of Rademacher complexity. For any set $A \in \mathbb{R}^n$, define

$$Z = \sup_{a \in A} \left(\sum_{k=1}^{n} w_k \, a_k \right) = \sup_{a \in A} \langle w, \, a \rangle,$$

where $w = (w_1, \ldots, w_n)$ is a sequence of i.i.d. $\mathcal{N}(0, 1)$. Its expectation $\mathcal{G}(A) = \mathbb{E}[Z]$ is known as the Gaussian complexity of set A.

Viewing Z as a function $f(w_1, \ldots, w_n)$, it is easy to verify that f is Lipschitz with parameter $\sup_{a \in A} \|a\|_2$.

Property

Z is sub-Gaussian with parameter $\sup_{a \in A} \sum_{k=1}^{n} a_k^2$.

Example: Singular values of Gaussian random matrices

For integers $n > d$, consider the random matrix $X \in \mathbb{R}^{n \times d}$ with i.i.d. $\mathcal{N}(0,1)$ entries, and let

$$\gamma_1(X) \geq \gamma_2(X) \geq \cdots \geq \gamma_d(X) \geq 0$$

be its ordered singular values. By Weyl's inequality,

$$\max_{k=1,2,\ldots,n} |\gamma_k(X) - \gamma_k(Y)| \leq \|X - Y\|_{\mathsf{op}} \leq \|X - Y\|_{\mathsf{F}}.$$

Therefore, each singular value $\gamma_k(X)$ is a 1-Lipschitz function of the random matrix (viewed as a nd-dim vector).

Property

$$\mathbb{P}\left(\left|\gamma_k(X) - \mathbb{E}[\gamma_k(X)]\right| \geq t\right) \leq 2\, e^{-t^2/2} \quad \text{for all } t > 0.$$