Example Sheet 1

Topics in Statistical Theory
Part III, Lent 2016 - Quentin Berthet*

Exercise 1

a) Let B be a random variable taking values in $[a, b]$. Show that $B - \mathbf{E}[B] \in \mathsf{sG}\big(\frac{(b-a)^2}{4}\big)$.

b) Let $X_1, \ldots, X_n$ be independent random variables such that $X_i \in [a_i, b_i]$, and $\bar{X} = \frac{1}{n}\sum X_i$. Provide an upper bound on $\mathbf{P}(\bar{X} - \mathbf{E}[\bar{X}] > t)$.

Exercise 2

Let X_1 and X_2 be two random variables such that $X_1 \in \mathsf{sG}(\sigma_1^2)$ and $X_2 \in \mathsf{sG}(\sigma_2^2)$ (note that no independence is assumed). Show that $X_1 + X_2 \in \mathsf{sG}\big(2(\sigma_1^2 + \sigma_2^2)\big)$.

Exercise 3

We recall that for $X \in \mathsf{sG}(\sigma^2)$, there is an explicit bound on $\mathbf{E}[|X|^k]$. Show that conversely, for any random variable Z with $\mathbf{E}[Z] = 0$, if $\mathbf{E}[|Z|^k] \leq 2\big(\sqrt{2}\sigma\big)^k\,\Gamma\big(\frac{k}{2}+1\big)$ for some $\sigma > 0$ and all $k \geq 2$, then $Z \in \mathsf{sG}(c\sigma^2)$ for some constant $c \geq 1$.

Exercise 4

a) Show that there exists a random variable $X \in \mathsf{sG}(1)$ such that $X \notin \mathsf{sG}(\tau^2)$ for all $\tau \in (0, 1)$.

b) Show that for all $\varepsilon \in (0, 1)$, there exists X with the above property such that $\mathbf{E}[X^2] = \varepsilon$.

c) Show that it is not possible to obtain a bound valid for all $X \in \mathsf{sG}(1)$ and all $t > 0$, of the form

$$\mathbf{P}(X > t) \leq e^{-\frac{t^2}{2\mathbf{E}[X^2]}}\,.$$

Exercise 5

Let $X \in \mathsf{sG}(\sigma^2)$, and $\Delta_4 = \mathbf{E}[X^4]^{1/2}$.

a) Show that $\mathbf{E}[X^2] \leq \Delta_4 \leq 4\sigma^2$. Give an example where Δ_4 is much smaller than σ^2.

b) Show that for all $t > 0$, it holds that

$$\mathbf{P}(X > t) \leq \exp\Big(-\frac{t^2}{2\Delta_4 + 2\sqrt{2}\sigma t}\Big)\,.$$

*q.berthet@statslab.cam.ac.uk

Exercise 6

Let $X \in \mathsf{sG}_d(\sigma^2)$. Show that for all $\delta \in (0,1)$, it holds with probability $1-\delta$ that

$$\|X\|_1 \le C\sigma d + \sigma\sqrt{2d\log(1/\delta)}\,,$$

for some constant $C > 0$.

Exercise 7

Let A be a random matrix of $\mathbf{R}^{m\times n}$ with independent coefficients such that $A_{ij} \in \mathsf{sG}(\sigma^2)$.

a) Show that $A \in \mathsf{sG}_{m\times n}(\sigma^2)$, i.e. that $\mathbf{Tr}(AU) \in \mathsf{sG}(\sigma^2)$ for all $\|U\|_F^2 = \sum_{i,j} U_{ij}^2 = 1$.

b) Show that for some constant $C > 0$

$$\mathbf{E}[\|A\|] \le C\sigma(\sqrt{m} + \sqrt{n})\,,$$

where $\|A\|$ is the operator norm of A defined by $\max_{x\in\mathbf{R}^m} \|Ax\|_2/\|x\|_2$.

Exercise 8

Consider the following testing problem, for $Z \in \mathsf{sG}_d(1)$ and some $\mathcal{S} \subset \mathbf{R}^d$

$$\begin{aligned} H_0 &: X = Z \\ H_1 &: X = \theta + Z,\ \theta \in \mathcal{S}\,. \end{aligned}$$

a) For $m \ge 2$, let $d = \binom{m}{2}$ and $X \in \mathbf{R}^d$ be the weights observed along the edges of a complete graph with m vertices. For some $\mu > 0$, let $\mathcal{S}$ be defined as

$$\mathcal{S} = \{\mu \mathbf{1}_{E_s} \mid \text{for all } 1 \le s \le m\}\,,$$

where E_s is the set of edges of a *star graph*, i.e. a graph whose only edges are those between vertex s and the $m-1$ other vertices. For all $\delta \in (0,1)$, describe a test that has probability of error less than δ when

$$\mu \ge 2\sqrt{\frac{2\log(m)}{m}} + 2\sqrt{\frac{2\log(1/\delta)}{m}}\,.$$

Is the test easily implementable?

b) With the same setting as above, consider for $k \le m$

$$\mathcal{S} = \{\mu \mathbf{1}_{K_S} \mid \text{for all } S,\ |S| = k\}\,,$$

where K_S is the set of edges of a *clique*, i.e. a graph whose only edges are those between vertices in S. For all $\delta \in (0,1)$, describe a test that has probability of error less than δ when

$$\mu \ge C\sqrt{\frac{2\log(me/k)}{k-1}} + C'\sqrt{\frac{2\log(1/\delta)}{k(k-1)}}\,,$$

for some constants $C, C' > 0$. Is the test easily implementable?