1 Some inequalities for Gaussian random variables

Lemma 1 *Let $\eta \sim \mathcal{N}(0,1)$. Then, for all $x > 0$,*

$$P(|\eta| > x) \ \le \ \sqrt{\frac{2}{\pi}}\frac{e^{-x^2/2}}{x}, \qquad (1)$$

$$\mathbf{E}[\eta^2 I(|\eta| > x)] \ \le \ \sqrt{\frac{2}{\pi}}\Big(x + \frac{2}{x}\Big)e^{-x^2/2}. \qquad (2)$$

Proof. Inequality (1) follows from the fact that

$$\forall x > 0: \quad \int_x^\infty e^{-u^2/2}du \le \int_x^\infty \frac{u}{x}e^{-u^2/2}du = \frac{1}{x}e^{-x^2/2}.$$

The proof of (2) is similar :

$$\mathbf{E}[\eta^2 I(|\eta| > x)] = \frac{2}{\sqrt{2\pi}}\int_x^\infty u^2 e^{-u^2/2}du \le \sqrt{\frac{2}{\pi}}\int_x^\infty \frac{u^3}{x}e^{-u^2/2}du = \sqrt{\frac{2}{\pi}}\Big(x + \frac{2}{x}\Big)e^{-x^2/2}.$$

Lemma 2 *Let $\eta_1, \dots, \eta_M, M \ge 2$, be $\mathcal{N}(0,1)$ random variables. Then*

$$\forall t > \sqrt{2}: \quad P\Big(\max_{1\le j\le M}|\eta_j| > t\sqrt{\log M}\Big) \le M^{1-t^2/2}.$$

and

$$\forall u > 0: \quad P\Big(\max_{1\le j\le M}|\eta_j| > \sqrt{2\log M} + u\Big) \le \frac{1}{\sqrt{\pi\log M}}e^{-u^2/2}.$$

Proof. Using the union bound we get

$$\begin{aligned} P\Big(\max_{1\le j\le M}|\eta_j| > t\sqrt{\log M}\Big) &\le \sum_{j=1}^M P(|\eta_j| > t\sqrt{\log M}) \\ &= MP(|\eta_1| > t\sqrt{\log M}) \end{aligned}$$

We now apply Lemma 1 for $x = t\sqrt{\log M} \ge \sqrt{2\log M} \ge \sqrt{2\log 2}$. This gives the bound

$$P(|\eta_1| > t\sqrt{\log M}) \le \frac{1}{\sqrt{\pi\log 2}}e^{-\frac{t^2\log M}{2}} < M^{-t^2/2}.$$

Similarly, for any $u > 0$,

$$P\Big(\max_{1\le j\le M}|\eta_j| > \sqrt{2\log M} + u\Big) \le \frac{M}{\sqrt{\pi\log M}}e^{-(\sqrt{2\log M}+u)^2/2} \le \frac{e^{-u^2/2}}{\sqrt{\pi\log M}}.$$

Lemma 3 *Let $\sigma > 0, M \ge 2$ and $Y_1, \dots, Y_M$ be random variables such that*

$$\forall s > 0, \forall i, \quad \mathbf{E}(e^{sY_i}) \le e^{s^2\sigma^2/2}.$$

Then

$$\mathbf{E}\left(\max_{1\le i\le M} Y_i\right) \le \sigma\sqrt{2\log M}.$$

If, in addition, $\mathbf{E}(e^{-sY_i}) \le e^{s^2\sigma^2/2}$, then

$$\mathbf{E}\left(\max_{1\le i\le M} |Y_i|\right) \le \sigma\sqrt{2\log(2M)}.$$

Proof. We first use Jensen's inequality for the convex function $x \mapsto \exp(sx)$:

$$\begin{aligned} \forall s > 0, \quad \exp\left(s\mathbf{E}\left(\max_{1\le i\le M} Y_i\right)\right) &\le \mathbf{E}\left(\exp\left(s\max_{1\le i\le M} Y_i\right)\right) \\ &= \mathbf{E}\left(\max_{1\le i\le M}\exp(sY_i)\right) \\ &\le \sum_{i=1}^M \mathbf{E}(\exp(sY_i)) \\ &\le Me^{s^2\sigma^2/2}. \end{aligned}$$

This implies

$$\forall s > 0, \quad \mathbf{E}\left(\max_{1\le i\le M} Y_i\right) \le \frac{\log M}{s} + \frac{s\sigma^2}{2}.$$

Putting here $s = \sqrt{\frac{2\log M}{\sigma^2}}$ we get the first inequality. The proof of the second one is analogous.

In particular, the results of Lemma 3 hold for $Y_i \sim \mathcal{N}(0,\sigma^2)$. Note that in Lemmas 2 and 3 we do not assume independence of $\eta_1, \dots, \eta_M$ or of $Y_1, \dots, Y_M$.

2 Oracle properties of hard thresholding

Theorem 4 *Consider the linear regression model under Assumption (ORT). Then the following holds.*

(i) **(Sparsity oracle inequality)** *If $\tau = \sigma\sqrt{\frac{2\log M}{n}}$, then*

$$\mathbf{E}|\hat\theta^H - \theta^*|_2^2 \le 2\sigma^2\left(\frac{|\theta^*|_0}{n}\log M\right)\left(1 + \frac{4}{\sqrt{\log M}}\right).$$

(ii) **(Selection of variables)** *If $\tau = B\sigma\sqrt{\frac{\log M}{n}}$ with $B > \sqrt{2}$ and*

$$\min_{j:\theta_j^*\ne 0}|\theta_j^*| > 2\tau,$$

then, with probability at least $1 - M^{1-B^2/2}$ we have:

$$\hat{J} = J(\theta^*),$$

where $J(\theta^*) = \{j : \theta_j^* \neq 0\}$ and $\hat{J} = \{j : \hat{\theta}_j^H \neq 0\}$.

Proof.
(i). If $\theta_j^* = 0$, then

$$|\hat{\theta}_j^H - \theta_j^*| = |y_j| I(|y_j| > \tau) = \varepsilon|\eta_j| I(|\eta_j| > \sqrt{2\log M}),$$

while for $\theta_j^* \neq 0$ we have the bound

$$|\hat{\theta}_j^H - \theta_j^*| = |y_j I(|y_j| > \tau) - \theta_j^*| \le |y_j - \theta_j^*| + |y_j| I(|y_j| \le \tau) \le \varepsilon|\eta_j| + \tau.$$

Therefore,

$$\begin{aligned} \mathbf{E}|\hat{\theta}^H - \theta|_2^2 &= \sum_{j=1}^M \mathbf{E}|\hat{\theta}_j^H - \theta_j^*|^2 \qquad (3) \\ &\le M\varepsilon^2 \mathbf{E}[\eta_1^2 I(|\eta_1| > \sqrt{2\log M})] + |\theta^*|_0 \mathbf{E}[(\varepsilon|\eta_j| + \tau)^2]. \end{aligned}$$

Since $\mathbf{E}(\eta_1^2) = 1$ and $\mathbf{E}|\eta_1| = \sqrt{2/\pi}$,

$$\begin{aligned} \mathbf{E}[(\varepsilon|\eta_j| + \tau)^2] &= \varepsilon^2 + \frac{4\varepsilon\tau}{\sqrt{2\pi}} + \tau^2 \qquad (4) \\ &= \varepsilon^2 \left(1 + 4\sqrt{\frac{\log M}{\pi}} + 2\log M\right). \end{aligned}$$

By Lemma 1

$$\mathbf{E}[\eta_1^2 I(|\eta_1| > \sqrt{2\log M})] \le \sqrt{\frac{2}{\pi}} \left(\frac{1}{\sqrt{2\log M}} + 2\sqrt{2\log M}\right) M^{-1}. \qquad (5)$$

Plugging (4) and (5) in (3) and using that $1 + \frac{1}{\sqrt{\pi \log M}} \le 4\sqrt{\frac{\log M}{\pi}}$ for all $M \ge 2$ and the inequality $6/\sqrt{\pi} \le 4$, we obtain the result.
(ii). Consider the random event

$$\mathcal{A} = \{|y_j - \theta_j^*| \ge \tau, j = 1, \dots, M\}.$$

By Lemma 2, the probability of the complementary event $\mathcal{A}^c$ satisfies

$$\begin{aligned} P(\mathcal{A}^c) &= P\left\{\max_{1\le j\le M} |\zeta_j| > \tau\right\} \\ &= P\left\{\varepsilon \max_{1\le j\le M} |\eta_j| > B\varepsilon\sqrt{\log M}\right\} \\ &\le M^{1-B^2/2}. \end{aligned}$$

Let us show that $\hat{J} \subseteq J(\theta^*)$ on the event $\mathcal{A}$. Let $\hat{\theta}_j^H \neq 0$. In this case,

$$\hat{\theta}_j^H = y_j \Leftrightarrow |y_j| > \tau \Leftrightarrow |\theta_j^* + \varepsilon\eta_j| > \tau,$$

which implies $|\theta_j^*| > \tau - |\varepsilon\eta_j| \ge \tau - \tau = 0$ on the event $\mathcal{A}$. Therefore, $\theta_j^* \neq 0$.
Let us show that $J(\theta^*) \subseteq \hat{J}$ on the event $\mathcal{A}$. Let $\theta_j^* \neq 0$. Then $|\theta_j^*| > 2\tau$, which yields

$$\begin{aligned} |y_j| &= |\theta_j^* + \varepsilon\eta_j| > 2\tau - |\varepsilon\eta_j| \\ &\ge 2\tau - \tau = \tau \end{aligned}$$

on the event $\mathcal{A}$. On the other hand, by definition of $\hat{\theta}^H$,

$$|y_j| > \tau \Rightarrow \hat{\theta}_j^H = y_j$$

Thus, $\hat{\theta}_j^H \neq 0$ with probability 1.

Definition 1 *The trigonometric basis in $L_2[0,1]$ is defined by*

$$\begin{aligned} \varphi_1(x) &\equiv 1, \\ \varphi_{2k}(x) &= \sqrt{2}\cos(2\pi kx), \\ \varphi_{2k+1}(x) &= \sqrt{2}\sin(2\pi kx), \quad k = 1,2,\ldots, \end{aligned}$$

for $x \in [0,1]$.

Suppose that function f is sufficiently regular, namely it belongs to the following class of functions.

Definition 2 *Let $\beta \in \{1,2,\ldots\}$, $L > 0$. The functional class of Sobolev $W(\beta, L)$ is defined by*

$$\begin{aligned} W(\beta,L) &= \Big\{f \in [0,1] \to \mathbf{R} : f^{(\beta-1)} \text{ is absolutely continuous and} \\ &\int_0^1 (f^{(\beta)}(x))^2 dx \le L^2,\ f^{(j)}(0) = f^{(j)}(1),\ j = 0,1,\ldots,\beta-1\Big\}. \end{aligned}$$

Any function $f \in W(\beta, L)$ is periodic (for $f \in [0,1] \to \mathbf{R}$ it means $f(0) = f(1)$) and admits the representation

$$f(x) = \theta_1\varphi_1(x) + \sum_{k=1}^{\infty}(\theta_{2k}\varphi_{2k}(x) + \theta_{2k+1}\varphi_{2k+1}(x)),\ \forall\, x \in [0,1],$$

in terms of its Fourier coefficients θ_j, where the sequence $\theta = \{\theta_j\}_{j=1}^{\infty}$ belongs to the space

$$\ell^2(\mathbf{N}) = \Big\{\theta : \sum_{j=1}^{\infty}\theta_j^2 < \infty\Big\}$$

and $\{\varphi_j\}_{j=1}^{\infty}$ is the trigonometric basis. Define

$$a_j = \begin{cases} j^{\beta}, & \text{for } j \text{ even}, \\ (j-1)^{\beta}, & \text{for } j \text{ odd}. \end{cases} \tag{1}$$

The next proposition proved in the Appendix of the book Tsybakov (2009) shows under which conditions on θ the function f belongs to the class $W(\beta, L)$.

Proposition 3 *Let $\beta \in \{1,2,\ldots\}$, $L > 0$, and $\{\varphi_j\}_{j=1}^{\infty}$ the trigonometric basis. The the function $f = \sum_{j=1}^{\infty}\theta_j\varphi_j$ belongs to the class $W(\beta, L)$ if and only if the vector θ of its coefficients of Fourier belongs to the l'ellipsoid in $\ell^2(\mathbf{N})$ defined by*

$$\Theta(\beta, Q) = \Big\{\theta \in \ell^2(\mathbf{N}) : \sum_{j=1}^{\infty} a_j^2\theta_j^2 \le Q\Big\}, \tag{2}$$

where $Q = L^2/\pi^{2\beta}$ and a_j is given in (1).

Proposition 3 deals only with integer β. It can be extended to all $\beta > 0$ by generalising the definition de $W(\beta, L)$:

Definition 4 *For $\beta > 0$, $L > 0$, the class of Sobolev $W(\beta, L)$ is defined by*

$$W(\beta, L) = \{f \in L_2[0,1] : \theta = \{\theta_j\} \in \Theta(\beta, Q)\},$$

where $\theta_j = \int_0^1 f\varphi_j$ and $\{\varphi_j\}_{j=1}^{\infty}$ is the trigonometric basis. Here $\Theta(\beta, Q)$ is the Sobolev ellipsoid (2) with $Q = L^2/\pi^{2\beta}$ and with coefficients a_j defined in (1).

We will use the following property of the trigonometric basis.

Lemma 5

$$\frac{1}{n}\sum_{s=1}^{n}\varphi_j(s/n)\varphi_k(s/n) = \delta_{jk}, \quad 1 \le j,k \le n-1, \tag{3}$$

where δ_{jk} is the Kronecker symbol.

Proof. Consider the case: $\varphi_j(x) = \sqrt{2}\cos(2\pi mx)$, $\varphi_k(x) = \sqrt{2}\sin(2\pi lx)$, where $j = 2m$, $k = 2l+1$, $j \le n-1$, $k \le n-1$, $n \ge 2$, m, l are integers. Other cases are treated similarly. Set

$$a = \exp\{i2\pi m/n\}, \quad b = \exp\{i2\pi l/n\},$$

where $i = \sqrt{-1}$. Then

$$\begin{aligned} T &= \frac{1}{n}\sum_{s=1}^{n}\varphi_j(s/n)\varphi_k(s/n) = \frac{2}{n}\sum_{s=1}^{n}\frac{(a^s + a^{-s})(b^s - b^{-s})}{4i} \\ &= \frac{1}{2in}\sum_{s=1}^{n}\Big[(ab)^s - (a/b)^s + (b/a)^s - (ab)^{-s}\Big]. \end{aligned}$$

Since $(ab)^n = 1$,

$$\sum_{s=1}^{n}(ab)^s = ab\frac{(ab)^n - 1}{ab - 1} = 0.$$

For the same reason, $\sum_{s=1}^{n}(ab)^{-s} = 0$ and (for $m \ne l$) $\sum_{s=1}^{n}(a/b)^s = \sum_{s=1}^{n}(b/a)^s = 0$. If $m = l$ we have $\sum_{s=1}^{n}(a/b)^s = \sum_{s=1}^{n}(b/a)^s = n$. This implies that $T = 0$.

Proposition 6 *Let $f \in W(\beta, L), \beta \ge 1, L > 0$. Let $\{\varphi_j\}$ be the trigonometric basis and $X_i = i/n$, $i = 1,\ldots,n$. Then, for any $n \ge 1$, $M \ge 1$,*

$$\inf_{\theta \in \mathbf{R}^M}\|f_\theta - f\|^2 \le C(\beta, L)\Big(\frac{1}{M^{2\beta}} + \frac{1}{n}\Big)$$

where $C(\beta, L) > 0$ is a constant depending only on β and L.

Proof. Let $\theta_j = \int_0^1 f\varphi_j, j = 1, 2, \ldots$. For any X_i, we can write

$$f(X_i) = \sum_{j=1}^{\infty} \theta_j \varphi_j(X_i),$$

where the sum converges since $\sum_{j=1}^{\infty} |\theta_j| < \infty$ for $f \in W(\beta, L)$ et $\beta > 1/2$. Indeed,

$$\sum_{j=2}^{\infty} |\theta_j| \le \sqrt{\sum_{j=2}^{\infty} \theta_j^2 a_j^2 \sum_{j=2}^{\infty} \frac{1}{a_j^2}} \le \frac{L}{\pi^\beta} \sqrt{\sum_{j=2}^{\infty} \frac{1}{a_j^2}}. \qquad (4)$$

This sum is finite for $\beta > 1/2$ since $a_j \ge (j-1)^\beta$. Define $\bar\theta = (\theta_1, \ldots, \theta_M)$. Then

$$\min_{\theta \in \mathbf{R}^M} \|f_\theta - f\|^2 \le \|f_{\bar\theta} - f\|^2 = \left\| \sum_{j=M+1}^{\infty} \theta_j \varphi_j \right\|^2$$

Denoting $n_* = \max(M+1, n)$ and adopting the convention $\sum_{j=M+1}^{n-1} = 0$ for $M + 1 \ge n$, we get

$$\begin{aligned}
\left\| \sum_{j=M+1}^{\infty} \theta_j \varphi_j \right\|^2 &= \frac{1}{n} \sum_{i=1}^{n} \sum_{j,k=M+1}^{\infty} \theta_j \theta_k \varphi_j(X_i) \varphi_k(X_i) \\
&\le \frac{2}{n} \sum_{i=1}^{n} \left(\sum_{j,k=M+1}^{n-1} \theta_j \theta_k \varphi_j(X_i) \varphi_k(X_i) \right) + \frac{2}{n} \sum_{i=1}^{n} \left(\sum_{j,k=n_*}^{\infty} \theta_j \theta_k \varphi_j(X_i) \varphi_k(X_i) \right) \\
&\le 2 \sum_{j=M+1}^{n-1} \theta_j^2 + 4 \left(\sum_{j=n_*}^{\infty} |\theta_j| \right)^2,
\end{aligned}$$

where we have used Lemma 5. In the last expression,

$$\sum_{j=M+1}^{n-1} \theta_j^2 \le \sum_{j=M+1}^{\infty} \theta_j^2 a_j^2 \frac{1}{a_j^2} \le \frac{1}{a_{M+1}^2} \sum_{j=1}^{\infty} a_j^2 \theta_j^2 \le \frac{L^2}{\pi^{2\beta} M^{2\beta}}$$

and, for $n_* \ge 3$,

$$\begin{aligned}
\left(\sum_{j=n_*}^{\infty} |\theta_j| \right)^2 &\le \frac{L^2}{\pi^{2\beta}} \sum_{j=n_*}^{\infty} \frac{1}{a_j^2} \le \frac{L^2}{\pi^{2\beta}} \sum_{j=n_*-1}^{\infty} \frac{1}{j^{2\beta}} \\
&\le \frac{L^2}{\pi^{2\beta}} \frac{1}{(2\beta - 1)(n_* - 2)^{2\beta - 1}} \\
&\le \frac{\tilde{C}(\beta, L)}{n}, \quad \text{since } \beta \ge 1, n_* \ge n,
\end{aligned}$$

where $\tilde{C}(\beta, L)$ is a constant depending only on β, L. For $n_* = 2$ analogous bound is immediate from (4).

Theorem 7 *Let $\beta \ge 1$, $L > 0$, $n \ge 1$ and $M = \lceil \alpha n^{\frac{1}{2\beta+1}} \rceil$ for $\alpha > 0$. Let ξ_i be random variables satisfying $\mathbf{E}(\xi_i) = 0$, $\mathbf{E}(\xi_i^2) \le \sigma^2$, $\mathbf{E}(\xi_i \xi_j) = 0$ for $i \ne j$, $i, j = 1, \ldots, n$. Let $\{\varphi_i\}$ be the trigonometric basis and $X_i = i/n$, $i = 1, \ldots, n$. Then,*

$$\sup_{f \in W(\beta, L)} \mathbf{E}\|\hat f^{MC} - f\|^2 \le Cn^{-\frac{2\beta}{2\beta+1}}$$

o $C > 0$ is a constant depending only on $\sigma^2, \beta, L, \alpha$.

Proof. Using Proposition 6,

$$\begin{aligned}
\mathbf{E}\|\hat f^{MC} - f\|^2 &\le \min_{\theta \in \mathbf{R}^M} \|f_\theta - f\|^2 + \frac{\sigma^2}{n} \min(M, n) \\
&\le C(\beta, L)\Big(M^{-2\beta} + \frac{1}{n}\Big) + \frac{\sigma^2 M}{n} \\
&= O\Big(n^{-\frac{2\beta}{2\beta+1}}\Big).
\end{aligned}$$