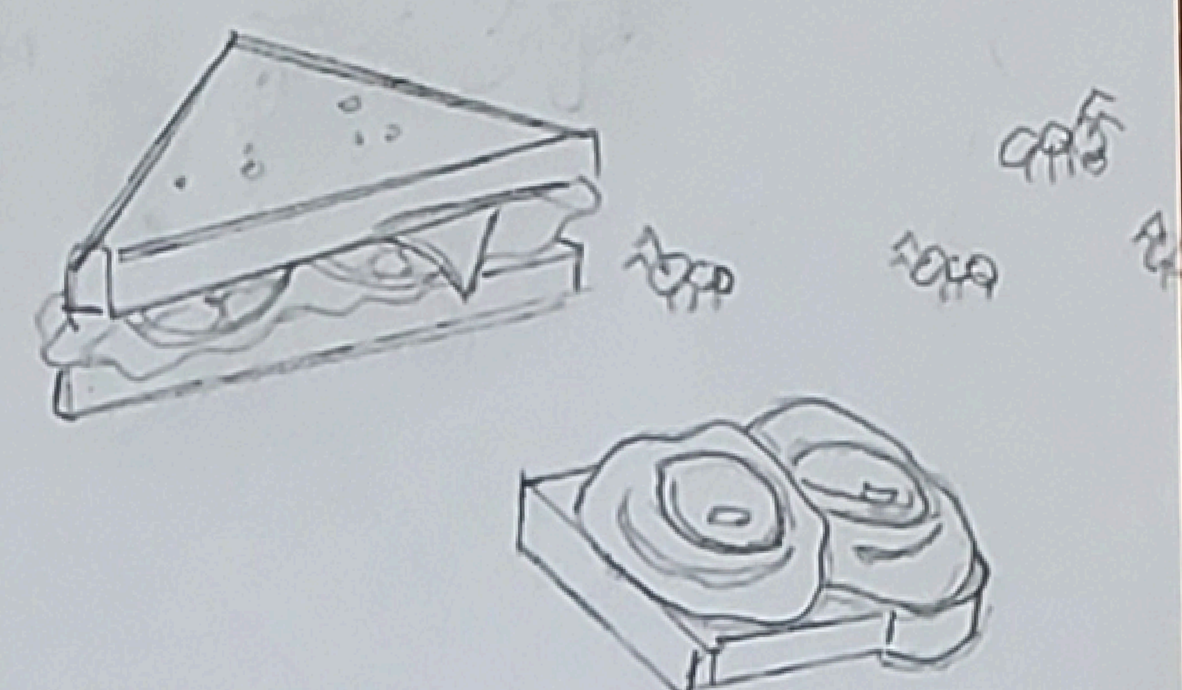
Chapter 14: Localization and Uniform Laws:

(Proposition 14.25) For any 1-bounded and star-shaped class F, the population and empirical radii satisfies the sandwich relation:



$$\frac{1}{4}\delta \leq \delta \leq 3\delta$$

14.1
$$\mathbb{E}\Big[\sup_{\|f\|_2 \leq t} \|f\|_n\Big]^2 = \mathbb{E}\Big[\sup_{\|f\|_2 \leq t} (\|f\|_n^2 - \|f\|_n^2)\Big]$$
$$\leq 2 \cdot \mathbb{E}\Big[\sup_{\|f\|_2 \leq t} \Big(\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(X_i)\Big)\Big]$$
$$\leq 4 \cdot t^2$$

$$\mathbb{E}\Big[\sup_{\|f\|_2 \leq t} \|f\|_n\Big] \leq \sqrt{4t^2}$$
$$\leq \sqrt{5t^2}$$
$$\leq \sqrt{5}\, t$$

Lemma 13.6) $\frac{G(\delta; F)}{\delta} \leq c \cdot \delta$

14.2 a) $\frac{\tilde{R}_n(s)}{s} \simeq \frac{G(\delta; F)}{\delta}$

$0 < s \leq t$ $\quad \frac{\tilde{R}_n(s)}{\delta_n} \simeq R(\delta; F) \leq \delta_n$

$t \leq s$ $\quad \frac{\tilde{R}_n(s)}{\delta_n} \leq \delta_n$

$\tilde{R}_n(s) \leq \delta_n^2$

$$\tilde{R}_n(s) \leq \max[\delta_n, \delta_n^2]$$

b) $\tilde{R}_n(t) \le \delta_n^2 \le C \cdot t^2$

$$\frac{\delta}{\sqrt{c}} \le t$$

14.3

$$\mathbb{E}_{\varepsilon}\Big[\sup_{\substack{f \in P \\ \|f_\delta\|_2 \le \delta}} \Big|\frac{1}{n}\sum_{i=1}^{n} \varepsilon_i f(x_i)\Big|\Big]$$

$$\le \int_0^{\delta} \sqrt{\frac{\log N(x, F, \|\circ\|_n)}{n}}\, dx$$

$$\le \int_0^{\delta} \sqrt{\frac{r \log(1 + 2\delta/x)}{n}}\, dx$$

$$\le \delta \sqrt{\frac{r}{n}} \int_0^{\delta^2 = 1} \sqrt{\log(1 + 2/u)}\, du$$

$$\cong \delta \sqrt{\frac{r}{n}} \cong c' \sqrt{\frac{1}{n}} \quad \text{where } c' = \delta\sqrt{r}$$

An inequality with a relationship to Students' t-distribution. Rademacher's $[\tilde{R}_n(t)]$ are less than the critical values in a t-table, specifically, Rademachers per constant (e.g $C = \sigma^2$). A real example involves binary images (black & white) above or below average intensity. The null hypothesis, the image has a homogeneous intensity against other images.

The expectation derived from "ball" packing problems. The lowest error; the packing no. or lower bound, and highest error; the covering no. or upper bound.

Packing #	Covering #

$$M(2\delta; \Pi, \rho) \le N(\delta; \Pi, \rho) \le M(\delta; \Pi, \rho)$$

"Lower Bound"	"Upper bound"

$$d \log(1/\delta) \le \log N(\delta; B, \|\circ\|) \le d \log(1 + 2/\delta)$$

"Lower Error"	"Upper Error"

$$r \log(1/\delta) \le \log N(x; F, \|\circ\|) \le r \log(1 + 2\delta/x)$$

14.4 (Equation 14.14a) $\tilde{R}(\delta, F) \leq \sqrt{\frac{2}{n}} \sqrt{\sum_{j=1}^{\infty} \min\{\mu_i, \delta^2\}}$

a) $f(x) = \sum_{j=1}^{r} \theta_j \phi_j$; $\sup_{\substack{\|f\|_H \leq 1 \\ \|f\|_2 \leq 1}} \left| \sum_{j=1}^{r} \varepsilon_j f(x_j) \right| \leq \sup_{\theta \in \mathbb{R}} \left| \sum_{j=1}^{r} \theta_j z_j \right|$ where $z_j = \sum_{j=1}^{r} \varepsilon_i \phi_j(x)$

Mercer's Theorem matches

$$\langle f(X_j), Z_j \rangle = \sum_{j=1}^{r} \frac{\langle f(X_j), \phi_j \rangle \langle Z_j, \phi_j \rangle}{\mu_j}$$

From Corollary 12.26, the relationship matches a space between $\langle f(X), Z \rangle$.

$$\mathbb{H} := \left\{ f(X_j) = \sum_{j=1}^{\infty} \beta_j \phi_j(x) \,\middle|\, (\beta_j)_{j=1}^{\infty} \in \ell^2(\mathbb{N}), \ \sum_{j=1}^{\infty} \frac{\beta_j^2}{\mu_j} \leq \infty \right\}$$

= where $\beta_j = \theta_j$ and $\mathbb{H} = D$.

b) $\mathcal{E} := \left\{ (\theta)_{j=1}^{\infty} \,\middle|\, \sum_{j=1}^{r} \frac{\theta_j^2}{\eta_j} \leq 2 \right\}$

$:= \left\{ (\theta)_{j=1}^{\infty} \,\middle|\, \sum_{j=1}^{r} \frac{\theta_j^2}{\min[\delta^2 + \mu_j]} \leq 2 \right\}$

$:= \left\{ (\theta)_{j=1}^{\infty} \,\middle|\, \sum_{j=1}^{r} \frac{\theta_j^2}{\delta^2} + \frac{\theta_j^2}{\mu_j} \leq 2 \right\}$

$:= \mathbb{H} = \left\{ (\theta)_{j=1}^{\infty} \,\middle|\, \sum_{j=1}^{n} \frac{\theta_j^2}{\delta^2} \right\} + D = \left\{ (\theta)_{j=1}^{\infty} \,\middle|\, \sum_{j=1}^{n} \frac{\theta_j^2}{\mu_j} \right\}$

$:= \mathbb{H} \cap D$

c) $\mathbb{E}_{\varepsilon, x} \left[\sup_{\substack{\|f\|_H \leq 1 \\ \|f\|_2 \leq \delta}} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i f(x_i) \right| \right] = \mathbb{E}_{\varepsilon, x} \left[\sup_{\substack{\|f\|_H \leq 1 \\ \|f\|_2 \leq \delta}} \left| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_i \theta_j \phi_j \right| \right]$

$$\leq \mathbb{E}_{\varepsilon, x} \left[\sup_{\substack{\|f\|_H \leq 1 \\ \|f\|_2 \leq 1}} \left| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_i \sqrt{\eta}\, \phi_j \right| \right]$$

$$\leq \mathbb{E}_{\varepsilon, x} \left[\sup_{\substack{\|f\|_H \leq 1 \\ \|f\|_2 \leq 1}} \left| \frac{1}{n} \sum_{j=1}^{n} \varepsilon_i \sqrt{2 \cdot \min\{\delta_j^2, \mu_j\}}\, \phi_j \right| \right]$$

$$\leq \sqrt{\frac{2}{n}} \cdot \sqrt{\sum_{j=1}^{\infty} \min\{\delta^2, \mu_j\}} \cdot \mathbb{E}\left[\sup_{\substack{\|f\|_H \leq 1 \\ \|f\|_2 \leq \delta}} \left| \sum_{j=1}^{\infty} \varepsilon_i \phi_j \right| \right]$$

$$\leq \sqrt{\frac{2}{n}} \cdot \sqrt{\sum_{j=1}^{\infty} \min\{\delta^2, \mu_j\}}$$

14.5 (Theorem 12.20) $T_K(\phi_j) = \mu_j \phi_j$; $K(x,z) = \sum_{j=1}^{\infty} \mu_j \phi_j(x) \phi_j(z)$

$$R_x(\cdot) = K(\cdot, x)$$

(Equation 12.11a) $T_K(f)(x) = \int_{\mathcal{X}} K(x,z) f(z) \cdot d\mathbb{P}(z)$

a) $\mathbb{E}[\hat{T}_K] = \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} [R_{x_i} \otimes R_{x_i}](f)\right]$

$$= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i) R_{x_i}\right]$$

$$= \mathbb{E}\left[\frac{1}{n}\sum_{i=1}^{n} f(x_i) \cdot K(\cdot, x)\right] \text{ "...from equation 12.11a"}$$

$$= \int_K \frac{f(x) \cdot K(\cdot, x)}{n} dx$$

$$= T_K$$

b) $|||\hat{T}_K - T_K|||_n := \sup_{\|f\|_H \leq 1} \|(\hat{T}_K - T_K)(f)\|_H$

$$:= \sup_{\|f\|_H \leq 1} |\langle f (\hat{T}_K - T_K) f \rangle|_H$$

$$:= |\langle f, \hat{T}_K f \rangle - \langle f, T_K f \rangle|$$

$$:= \left| \|f\|_n^2 - \|f\|_2^2 \right|$$

c) $\| \hat{T}_K(\phi_j) - \mu_j \phi_j \|_H$

$$\mu \| \hat{T}_K(\phi_j) - \mu_j \phi_j \| \leq \| \hat{T}_K - T_K \|$$

$$\| \hat{T}_K(\phi_j) - \mu_j \phi_j \| \leq \frac{\| \hat{T}_K - T_K \|_H}{\mu_j}$$

14.6 (Equation 14.22b) $\mathbb{E}[f^4(x)] \leq c^2 (\mathbb{E}[f^2(x)])^2$

(Equation 2.55a,b,c) $\mathbb{E}[X^{2k}] = \frac{(2k!)}{2^k(k!)} \circ (c\tau)^{2k}$

$$\mathbb{E}[e^{\lambda k}] \leq \sum_{k=0}^{\infty} \frac{2^k \lambda^{2k} \mathbb{E}[X^{2k}]}{2k!}$$

Since $f_\theta(x) = \langle \theta, x \rangle$, which is also Gaussian:

$$\mathbb{E}[\langle x, \theta \rangle]^4 = (c^2 \mathbb{E}[\langle x, \theta \rangle^2])^2 = \left(c^2 [1 + \theta^2 \mathbb{E}[X^2] + \theta^4 \mathbb{E}[X^2]^2 + \cdots]\right)^2$$

$$\mathbb{E}[\langle x, \theta \rangle]^4 = ([B+6]^2 \cdot \mathbb{E}[\langle x, \theta \rangle^2])^2 = \left([B+6]^2 \circ [1 + \theta^2 \mathbb{E}[X^2] + \theta^4 \mathbb{E}[X^2]^2 + \cdots]\right)^2$$

Term #1: $c^2 = (B+6)^2$

Term #2: $c^2 \circ 2\theta^2 \mathbb{E}[x^2] = (B+6)^2 \circ 2 \circ \theta^2 \mathbb{E}[x^2]$

Term #3: $c^2 \circ 3\theta^4 \mathbb{E}[x^2]^2 = (B+6)^2 \circ 3 \circ \theta^4 \mathbb{E}[x^2]^2$

The term derives from Equation 2.55a no matter the constant. Page 464 shows another derivation.

14.7. (Theorem 14.12)

$$\|f\|_n^2 \geq \frac{1}{2} \|f\|_2^2$$

If: $f_\theta = \langle X, \theta \rangle$, then:

$$\|f_\theta\|_n^2 = \|\langle X, \theta \rangle\|_n^2$$

$$= \frac{\|X\theta\|_2^2}{n}$$

$$\geq \frac{1}{2}\|f_\theta\|_2^2$$

$$\geq \|\sqrt{\Sigma}\,\theta\|_2^2$$

$$\geq \gamma_{min}(\Sigma)\|\theta\|_2^2$$

$$\frac{\|X\theta\|_2^2}{n} \geq \gamma_{min}(\Sigma)\|\theta\|_2^2$$

$$\frac{\|X\theta\|_2^2}{\gamma_{min}(\Sigma)\|\theta\|_2^2} \geq n$$

Reversal

$$n \geq -\frac{\|X\theta\|_2^2}{\gamma_{min}(\Sigma)\|\theta\|_2^2}$$

$$\geq -\frac{\mathbb{E}[\sup \langle X, \theta \rangle]_2^2}{\gamma_{min}(\Sigma)\|\theta\|_2^2}$$

$$\geq -\frac{\mathrm{trace}(\langle X, \theta \rangle)\,\mathbb{E}[\langle X, \theta \rangle]_2^2}{\gamma_{min}(\Sigma)\|\theta\|_2^2}$$

$$\geq \frac{-\rho^2(\Sigma)\, s\log(ed/s)}{\gamma_{min}(\Sigma) \cdot n}$$

$$\geq \frac{c_0 \cdot \rho^2(\Sigma)}{\gamma_{min}(\Sigma)}\, s\log(ed/s) \quad \text{where} \quad c_0 = -1/n$$

From Exercise 4.11 and Exercise 5.7:

$$\mathbb{E}[\langle X, \theta \rangle]_2^2 \leq \frac{s\log(ed/s)}{n}$$

14.8 $\tilde{R}(\delta; G) \leq \delta^2$ $\quad \tilde{R}(\sqrt{d}\,\varepsilon; G) \leq \tilde{R}(\delta; G)$

$\tilde{R}(\sqrt{d}\,\varepsilon; G) \leq d\varepsilon^2$ $\quad \sum_{i=1}^{d} \frac{\|g_i\|^2}{d\varepsilon^2} \leq \sum_{i=1}^{d} \frac{\|g_i\|^2}{\delta^2}$

$$\delta^2 \leq d\varepsilon^2$$

14.9. (Equation 14.56) $\hat{f} \in \underset{f \in F}{\operatorname{argmin}}\, \mathbb{P}(-\log f(x)) = \underset{f \in F}{\operatorname{argmin}} \left\{ -\frac{1}{n} \sum_{i=1}^{n} \log f(x_i) \right\}$

Proof by Deduction #1:

If $f(x) = \{ \frac{1}{n}, \frac{2}{n}, \frac{3}{n}, \ldots, \frac{x}{n} \}$ where $n > x$

then $\lim_{n \to \infty} \hat{f} \in \lim_{n \to \infty} \underset{f \in \mathbb{F}}{\operatorname{argmin}}\, \mathbb{P}(-\log f(x))$

$$= \lim_{n \to \infty} \operatorname{argmin} \mathbb{P}\left(-\log\left(\frac{1}{n}\right) - \log\left(\frac{2}{n}\right) - \log\left(\frac{3}{n}\right) - \cdots - \log \frac{x}{n}\right)$$

$$= \lim_{n \to \infty} \operatorname{argmin} \mathbb{P}\left(-\log\left(\frac{x}{n^x}\right)\right)$$

$$= \infty$$

Proof by Deduction #2:

If $f(x) = e^{-x^2/2\sigma^2}$

then $\lim_{\sigma \to 0} \hat{f} \in \lim_{\sigma \to 0} \underset{f \in \mathbb{F}}{\operatorname{argmin}}\, \mathbb{P}(-\log f(x))$

$$= \lim_{\sigma \to 0} \operatorname{argmin} \mathbb{P}\left(-\log\left(e^{-x^2/2\sigma^2}\right)\right)$$

$$= \lim_{\sigma \to 0} \operatorname{argmin} \mathbb{P}\left(x^2/2\sigma^2\right)$$

$$= \infty$$

14.10 Hellinger Distance : $H(f\|g) = \frac{1}{\sqrt{2}} \int (\sqrt{f(x)} - \sqrt{g(x)}) \, dx$

Kullbeck-Leibler Divergence : $D(f\|g) = \int \log \frac{f(x)}{g(x)} f(x) \, dx$

"A function's distance from another function, applications include: parametric functions, normal distances, region bounds, estimation and estimation tests."

$$D(f\|g) = \int \log \frac{f(x)}{g(x)} f(x) \, dx$$

$$= 2 \cdot \int \log \frac{\sqrt{f(x)}}{\sqrt{g(x)}} f(x) \, dx$$

$$= -2 \int \log \frac{\sqrt{g(x)}}{\sqrt{f(x)}} f(x) \, dx$$

$$\geq 2 \int \left\{ 1 - \frac{\sqrt{g(x)}}{\sqrt{f(x)}} \right\} f(x) \, dx$$

$$\geq \int \left\{ 1 + 1 - 2\sqrt{f(x)}\sqrt{g(x)} \right\} dx$$

$$\geq \int \left\{ \sqrt{f(x)} - \sqrt{g(x)} \right\}^2 dx$$

$$\geq 2H^2(f\|g)$$

$$\boxed{\begin{array}{l} \underline{\text{Taylor Expansion}} \\ \log x = (x-1) - \frac{(x-1)^2}{2} + \frac{(x-1)^3}{3} - \cdots \end{array}}$$

14.11 (Equation 14.62) $\phi_m(x) = \begin{cases} 1 & \text{if } x \in (m-1, m)/T \\ 0 & \text{otherwise} \end{cases}$

$$\hat{f}_T = \sum_{i=1}^{T} \hat{\beta}_m \phi_m \quad \text{Where} \quad \hat{\beta}_m = \frac{1}{n} \sum_{i=1}^{n} \phi_m(x_i)$$

$$\hat{f}_T = \|\hat{f} - f^*\|_2^2 = \sum_{m=1}^{T} \left(\frac{1}{n} \sum_{i=1}^{n} \phi_m(x_i) \right) \phi_m$$

$$= \frac{1}{T^2} [0, 1]^2$$

$$= \frac{\sigma^2}{T^2}$$

$$\leq c \cdot n^{-2/3} \quad \text{when} \quad c = \sigma, \quad n = T^3$$

The $\frac{1}{n^{2/3}}$ bound represents a "sharp" lower bound for the Lipschitz family (pg 495) In "Estimating a Density..." by Lucien Birgé, University Paris X, (1984), the factor $1/n^{2/3}$ restricts Rademacher's from sinusoidal noise.