Chapter 4: Uniform laws of large numbers:

(Glivenko - Cantelli Theorem)

The empirical cumulative distribution function $\hat{F}$ estimates a populations cumulative distribution function in a uniform norm:

$$\|\hat{F}_n - F\|_\infty \xrightarrow{a.s.} 0$$

4.1.

a) $\hat{F}_n$ = C.D.F. By the "Fundamental Theorem of Calculus" also known as the Glivenko-Cantelli theorem: $\|\hat{F}_n - F\|_\infty \xrightarrow{a.s.} 0$

$$\gamma(\|\hat{F}_n - F\|)_\infty \xrightarrow{a.s.} 0$$

$$\gamma(\hat{F}_n) \longrightarrow \gamma(F)$$

b) i) Mean Functional: $F \longmapsto \int x\,dF(x)$

$$\hat{F}_n = \frac{1}{n}\sum_{i=1}^{n} X_i = \frac{(n+1)}{2} \quad ; \quad F = \frac{1}{n}\int_1^n x\,dx = \frac{1}{n}\left[\frac{n^2-1}{2}\right]$$

A continuous function $\hat{F}_n - F \xrightarrow{a.s} 0$, so the mean functional implies discontinuity. (Plot A)

ii) Cramer-von Mises Functional: $F \longmapsto \int [F(x) - F_0(x)]^2 dF_0(x)$

$$\hat{F}_n = \frac{1}{n}\sum_{i=1}^{n}(F(X_i) - F_0(X_i))^2 = \frac{1}{n}\sum_{i=1}^{n} F(X_i)^2 - \frac{1}{n}\left[\sum_{i=1}^{n} F_0(X_i)\right]^2$$

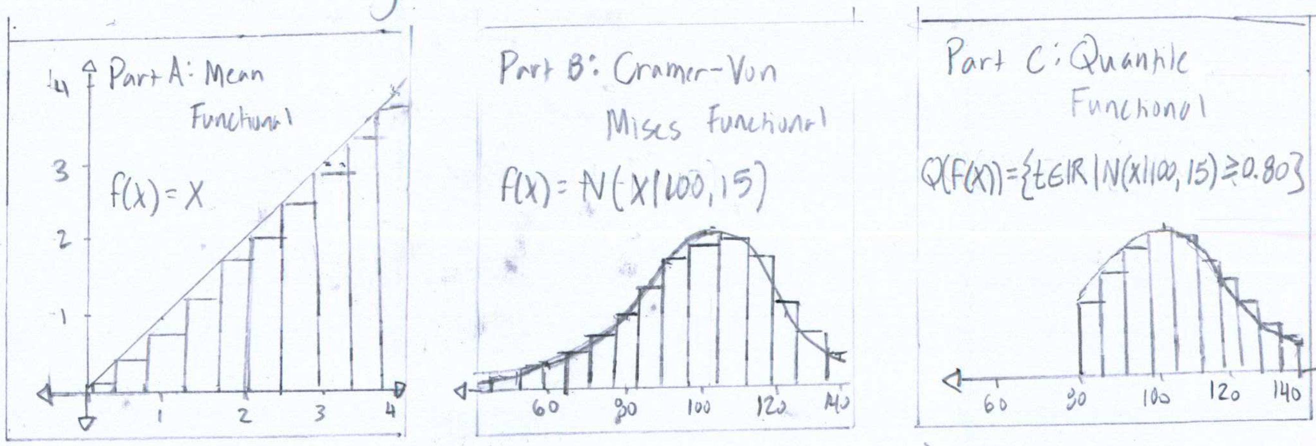
$$F = \frac{1}{n}\int_1^n (F(X_i) - F_0(X))^2 dX = \frac{1}{n}\int_1^n F(X)^2 dX - \frac{1}{n}\left[\int_1^n F_0(X)\,dX\right]$$

Cramer-von Mises asymptotically approaches zero as n rises. Cramer (1928), (Plot B)

iii) The Quantile Functional: $Q_\alpha(F) = \inf\{t \in \mathbb{R} \mid F(t) \geq \alpha\}$

$$F(t) = F(Q_\alpha(F) + \varepsilon) > \alpha \text{ and } F(t) = F(Q_\alpha(F) - \varepsilon) < \alpha$$

The quantile functional is continuous from the left and the right. (Plot C)



(Theorem 4.10: Rademacher Lower and Upper Bounds)

$\|P_n - P\| \leq 2 \cdot R_n(F) + \sigma$ unless $R_n(F) = o(1)$, then $\|P_n - P\| \xrightarrow{a.s} 0$

where $F = \{1 \mid s \in S\}$

4.2 $R_n(S) = \mathbb{E}\left[\sup_{s \in S} \left| \frac{1}{n} \sum_{i=1}^{n} \varepsilon_i [X_i] \right|\right] \geq 1/2$ at a lower bound

$2R_n(S) + \sigma \geq 1$ so the function does not asymptotically converge.

$R_n(S)$ is one at 100% accumulative probability or 100% convergent probababil.ty in a distribution.

Note: A possible typo exists for a(i) in the denominator as by Bernoulli

4.3 a) i) Bernoulli: $p_\theta(x) = \frac{e^{\theta x}}{1 + e^{\theta x}}$ for $x \in \{0, 1\}$

$$R(\theta, \theta^*) = \mathbb{E}\left[\log \frac{P_{\theta}^*(x)}{P_\theta(x)}\right] = \sum_0^1 \theta^* \circ \log \frac{P_\theta^*(x)}{P_\theta(x)}$$

$$= \sum_0^1 \theta^* \log\left(\frac{e^{\theta^* x}}{1 + e^{\theta x}} \frac{1 + e^{\theta x}}{e^{\theta x}}\right)$$

$$= \log\left(\frac{e^{x(1-\theta)} + e^{x}}{e^{x} + 1}\right)$$

ii) Poisson: $p_\theta(x) = \frac{e^{\theta x} e^{-\exp(\theta)}}{x!}$ for $x \in \{0, 1, 2, \ldots\}$

$$R(\theta,\theta^*) = \mathbb{E}_{\theta^*}\left[\log \frac{P_{\theta^*}(x)}{P_\theta(x)}\right]$$

$$= \sum_{\theta=0}^{\infty} \theta^* \circ \log\left(\frac{e^{\theta^* x} e^{-\exp(\theta^*)}}{x!} \circ \frac{x!}{e^{\theta x} \circ e^{-\exp(\theta)}}\right)$$

$$= \sum_{\theta=0}^{\infty} \theta^* \left[X(\theta^*-\theta) - e^{\theta^*} + e^{\theta}\right]$$

= Divergent risk!!!

ii) Multi-variate Gaussian: $P_\Theta(x) = N(\Theta, \Sigma^{-1})$

$$R(\theta,\theta^*) = \mathbb{E}_{\theta^*}\left[\log \frac{P_{\theta^*}(x)}{P_\theta(x)}\right]$$

$$= \sum_{\theta=-\infty}^{\infty} \theta^* \left[\log \frac{N(\theta^*, \Sigma^{-1})}{N(\theta, \Sigma^{-1})}\right]$$

$$= \sum_{\theta=-\infty}^{\infty} \theta^* \left[-\frac{(x-\theta^*)^2}{2\Sigma^{-1}\Sigma} + \frac{(x-\theta)^2}{2\Sigma^{-1}\Sigma}\right]$$

= Divergent risk!!!

The expectation wants an integral or summation about θ^* and not X.

b) i) $\mathbb{E}(\hat{\theta},\theta^*) = R(\hat{\theta},\theta^*) - \inf_{\theta\in\Omega} R(\theta,\theta^*)$

Bernoulli: $= R\left(\underset{\theta}{\mathrm{argmin}}\{R(\theta,\theta^*)\}, \theta^*\right) - \inf_{\theta\in\Omega} R(\theta,\theta^*)$

Where $R(\theta,\theta^*) = \log\left(\frac{e^{x(1-\theta)} + e^{x}}{e^{x}+1}\right)$

Poisson: $R\left(\underset{\theta}{\mathrm{argmin}}\{R(\theta,\theta^*)\}, \theta^*\right) - \inf_{\theta\in\Omega} R(\theta,\theta^*)$

Where $R(\theta,\theta^*) = \sum_{\theta=0}^{\infty} \theta^*\left[x(\theta^*-\theta) - e^{\theta^*} + e^{\theta}\right]$

Multivariate Gaussian: $R\left(\underset{\theta}{\mathrm{argmin}}\{R(\theta,\theta^*)\}, \theta^*\right) - \inf_{\theta\in\Omega} R(\theta,\theta^*)$

Where $R(\theta,\theta^*) = \sum_{\theta=-\infty}^{\infty} \theta^*\left[-\frac{(x-\theta^*)^2}{2\Sigma^{-1}\Sigma} + \frac{(x-\theta)^2}{2\Sigma^{-1}\Sigma}\right]$

The solutions describe cases when $\theta^* \in \Omega_0$ and not a subset of Ω.

ii) The upper limit on excess risk is not a minimal likelihood estimate. A Rademacher complexity contains no minimizing, and in risk decomposition, not the minimal (infinum), nor expectation, but likely a supremum.

4.4

a) Equation 4.17: $\mathbb{E}[\|P_n - P\|] \leq \mathbb{E}_{X,Y}\left[\sup\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - f(Y_i)\right|\right]$

$\sup \mathbb{E}[g(X)] \leq \mathbb{E}[\sup|g(X)|]$, when the supremum and expectation do not commute. A real world case depends on acceptable error or experimental error. Such that $\sup \mathbb{E}[g(X)] - \mathbb{E}[\sup|g(X)|] \leq$ Error.

Jensen's proof about inequalities also justifies the statement when $g(X)$, a convex function.

b) If Φ, is convex and non-decreasing, then Jensen's inequality, $\phi(E[X]) \leq E[\phi(X)]$ relates. $\sup_{g \in G} \Phi(\mathbb{E}[|g(X)|]) \leq \mathbb{E}[\Phi(\sup_{g \in G}|g(X)|)]$ because the inequality generalizes a secant above: a concave function as reasonable argument for "less than." Proposition 4.11 applies Jensen's theory by the assumption, "Expected law of large numbers value is a minumum with upper and lower bounds about the local convexity."

4.5. (Proposition 4.12)

$$\|P_n - P\| \geq \frac{1}{2}R_n(F) - \frac{\sup|\mathbb{E}[f]|}{2\sqrt{n}} - \delta \quad \text{with } \mathbb{P} \text{ at least } 1 - e^{-\frac{n\delta^2}{2b^2}}$$

a) A recentered function class: $\bar{\mathcal{F}} = \{f - \mathbb{E}[f] \mid f \in \mathcal{F}\}$

$$\|S_n\| = \sup\left|\frac{1}{n}\sum_i \varepsilon_i f(x) - \frac{1}{n}\sum_i \varepsilon_i \mathbb{E}[f(x)]\right| \geq \sup\left|\frac{1}{n}\sum_i \varepsilon_i f(x)\right| - \left|\sum_i \varepsilon_i\right| \frac{\sup \mathbb{E}[f(x)]}{n}$$

$$\mathbb{E}[\|S_n\|] \geq \mathbb{E}[\|S_n\|_F] - \mathbb{E}\left[\left|\sum_i \varepsilon_i\right|\right] \frac{\sup \mathbb{E}[f(x)]}{n}$$

With Chebyshev's inequality,

$$\mathbb{E}[\|S_n\|] \geq \mathbb{E}[\|S_n\|_F] - \sqrt{n}\,\frac{\sup \mathbb{E}[f(x)]}{\sqrt{n}}$$

$$\geq \mathbb{E}[\|S_n\|_F] - \frac{\sup \mathbb{E}[f(x)]}{\sqrt{n}}$$

b) Equation 4.21: $\frac{1}{2}\mathbb{E}[\|S_n\|] \leq \mathbb{E}[\|P_n - P\|] \leq 2\mathbb{E}[\|S_n\|]_F$

$$\frac{1}{2}\left(R(f) - \frac{\sup \mathbb{E}[f(x)]}{\sqrt{n}}\right) \leq \frac{1}{2}\mathbb{E}[\|S_n\|] \leq \mathbb{E}[\|P_n - P\|]$$

$$\frac{1}{2} \circ R(f) - \frac{\sup \mathbb{E}[f(x)]}{2\sqrt{n}} \leq \|P_n - P\|$$

4.6. $\mathcal{F} := [x \longmapsto \text{sign}(\langle \theta, x\rangle) \mid \theta \in \mathbb{R}^n, \|\theta\|_2 = 1]$

$R(F(x_1^n)|n) = \mathbb{E}[\sup|\frac{1}{n}\sum_i \varepsilon_i \text{sign}(\varepsilon_i)|] = 1$ because a Rademacher complex above 1/2 is fit for empirical risk minimization, an overfit at 1.0

(Converse Inequality)

$a < b \longleftrightarrow b > a$ and $a \leq b \longleftrightarrow b \geq a$ for $a, b \in \mathbb{R}$

(Triangle Inequality)

$\|x + y\| = \|x\| + \|y\|$ for $a, b \in \mathbb{R}$

4.7a) If (F and $\bar{F}$) $\in \text{conv}(F)$, then

$$\frac{\left|\sum_i \varepsilon_i \bar{F}(x_i)\right|}{n} \leq \frac{\left|\sum_i \varepsilon_i F(x_i)\right|}{n} \leq \sup\left|\frac{\sum_i \varepsilon_i f(x_i)}{n}\right|$$

$$R_n\left[\frac{\sum_i \varepsilon_i \bar{F}(x_i)}{n}\right] \leq R_n\left[\frac{\sum_i \varepsilon_i F(x_i)}{n}\right] \leq \sup R_n\left[\frac{\sum_i \varepsilon_i f(x_i)}{n}\right]$$

$$R_n[\bar{F}] \leq R_n[F] \leq R_n[\text{conv}(F)]$$

b) $\|F+G\| \leq \|F\| + \|G\|$

$$R[\|F+G\|] \leq R[\|F\|] + R[\|G\|]$$

c) $\|\frac{1}{n}\sum_i \varepsilon_i f(x) + \frac{1}{n}\sum_i \varepsilon_i g(x)\| \leq \|\frac{1}{n}\sum_i \varepsilon f(x)\| + \|\frac{1}{n}\sum_i \varepsilon g(x)\|$

$$R[\|F+g\|] \leq R[\|F\|] + \|\sum_i \varepsilon_i\| \frac{g(x)}{n}$$

$$\leq R[\|F\|] + \frac{g(x)}{\sqrt{n}}$$ by Cauchy-Shwartz inequality

(Vapnik-Chervonenski Dimension)

"VC"; a measure about the maximum points on algorithm handles without "shatter."

A sets sizes growth by 2^n "shatters" an algorithm.

4.8a) $S^c := \{S^c | S^c \in S\}$

$T \cup S^c = \{T : S^c | T : S^c \in T \text{ or } T : S^c \in S\}$

$T \cup S^c = U$ and $VC(T \cup S) = VC(U)$

b) $S \cap T = [S \cap T | S \in S, T \in T]$

$S \cap T = U$ and $VC(S \cap T) = VC(U)$

$$\text{card}(S \cap T(x_i^n)) \leq (n+1)^{VC(S \cap T)}$$

c) $S \cup T = \{S \cup T | S \in S, T \in T\}$

$S \cup T = U$ and $VC(S \cup T) = VC(U)$

$$\text{card}(S \cup T(x_i^n)) \leq (n+1)^{VC(S \cup T)}$$

Notes: A set growth by 2^n is positively above a set growth at $n, n^2, n^3,$ or n^4

4.9. $R(F(X_1^n)|n) = \mathbb{E}\left[\sup\left|\frac{1}{n}\sum \varepsilon_i f(x)\right|\right] \le 4\,\mathbb{E}\left[\sup\left|\frac{1}{n}\sum \varepsilon_i f(x)\right|\right]\log(n+1)^{v}$

by the Cauchy-Shwarz

$$R(F(X_1^n)|n) \le 4\sup\sqrt{\frac{\sum f^2(x)}{n}}\sqrt{\frac{v\log(n+1)}{n}}$$

where v is VC's number.

4.10: $$\binom{n-1}{k} + \binom{n-1}{k-1} = \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-1-k+1)!}$$

$$= \frac{(n-1)!}{k!(n-1-k)!}\frac{(n-1-k+1)}{(n-1-k+1)} + \frac{k}{k}\frac{(n-1)!}{(k-1)!(n-1-k+1)!}$$

$$= \frac{(n-1)!(n-k) + k(n-1)!}{k(n-k)!}$$

$$= \frac{n!}{k(n-k)!}$$

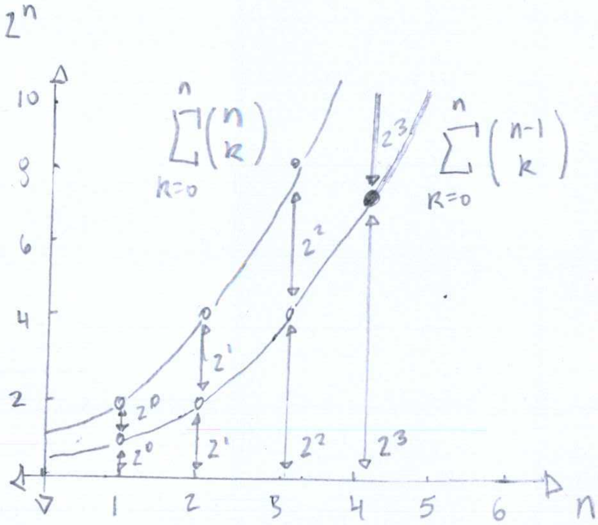
$$= \binom{n}{k}$$

$$\mathrm{Card}(S(P)) \le \mathcal{Z}_k(n-1) + \mathcal{Z}_{k-1}(n-1)$$

$$\le \sum_{k=0}^{n}\binom{n-1}{k} + \sum_{k=0}^{n}\binom{n-1}{k-1}$$

$$\le 2^{n-1} + 2^{n-1}$$

$$\le 2^n = \mathcal{Z}_k(n)$$

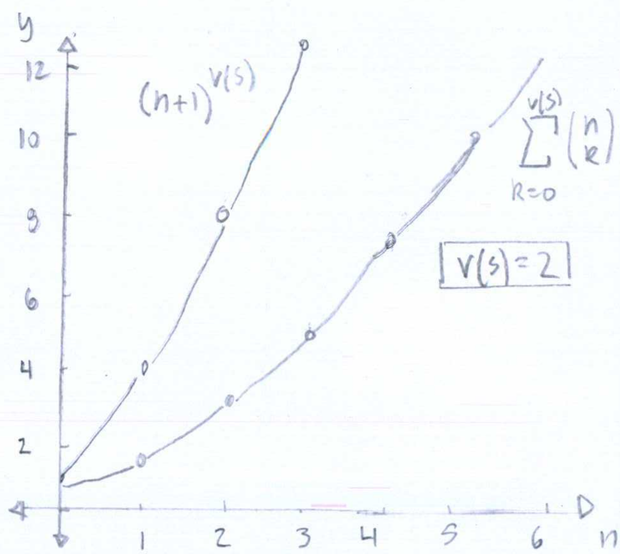


(Vapnik-Chervonenks, Sauer and Shelah)

$$\mathrm{Card}(S(P)) \le \sum_{k=0}^{v(S)}\binom{n}{k} \le (n+1)^{v(S)}, \quad n \ge v(S)$$

4.11

a) A plot visible toward the right shows $\sum_{k=0}^{v(S)}\binom{n}{k} \le (n+1)^{v(S)}$ where $n \ge v(S)$



b) When $n \geq v$, prove : $card(S(x_1^n)) \leq \left(\frac{en}{v}\right)^v$

$$e^v = \left(1+\frac{x}{n}\right)^n \geq \sum_{k=0}^{v} \binom{n}{k} \left(\frac{v}{n}\right)^v$$

$$\left(\frac{n}{v}\right)^v e^v \geq \sum_{k=0}^{v} \binom{n}{k} = card(S(x_1^n))$$

$$e^x = \lim_{n \to \infty} \left(1+\frac{x}{n}\right)^n$$

Proof: Taylor Expansion

$$(1+x)^n = 1 + nx + \frac{n(n-1)}{2!}x^2 + \cdots$$

$$\left(1+\frac{x}{n}\right)^n = 1 + n\left(\frac{x}{n}\right) + \frac{n(n-1)}{2!}\left(\frac{x}{n}\right)^2 + \cdots$$

$$e^x = 1 + nx + \frac{n(n-1)}{2!} \cdot x^2 + \cdots$$

$$e^x = \left(1+\frac{x}{n}\right)^n$$

4.12. $S_{left}^d := \{(-\infty, t_1] \times (-\infty, t_2] \times \cdots \times (-\infty, t_d] \mid (t_1, \ldots t_d) \in \mathbb{R}^d\}$

Each subset is countable, such as $S_{left}^1 := \{(-\infty, t_1] \mid t_1 \in \mathbb{R}^1\}$, so S_{left}^d and $VC(S_{left}^d) = d$.

4.13.

a) $S_{sphere}^2 := \{S_{a,b}\ (a,b) \in \mathbb{R}^2 \times \mathbb{R}_+\}$

where $S_{a,b} := \{x \in \mathbb{R}^2 \mid \|x-a\|_2 \leq b\}$

and $b \geq 0$ at $a = (a_1, a_2)$

$$S^2 = S \cdot S = \|x-a\|\|x-a\| = \|x-a\|^2 \cong (n+1)^d$$

Although a real sphere with three dimensions shatters growth in the VC-measurement. $(x-a)^d = (n+1)^d$

$$(x-a)^{d+1} \geq (n+1)^d$$

b) (D. Fitzpatrick, A. Iosevich, B. McDonald, E. Wyman, 2021)

Ultimately, a unique bound in shatter limit analysis. The set size $\{(n+1)^d\}$ in an algorithm remains with solutions and unbreakable limits when $d \leq 2$. A set of points growing at $d=3$, "shatters" the algorithms... but the authers propose $d \leq 2$ and $d \geq 4$ as collapsable problem sets.

4.14 $d \geq 2 : h_s : \{0,1\}^d \to \{0,1\}$ of the form $h_s(X_1, \ldots, X_d) = \begin{cases} 1 & \text{if } x_j = 1 \text{ for all } j \\ 0 & \text{otherwise} \end{cases}$

The problem has a lower and upper bound.

A lower bound is a set size of 0, because $\{0,1\}^d \to \{0,1\}$, while an upper at $VC(h_s(X_1, \ldots, X_d)) = d$ where each element a one.

(Convex Set)

If a set joins subsets by a single intersection or line segment.

(Closed set)

A set whose complement is an open set; a set containing all limit points.

4.15: C^d_{cc} of all closed and convex sets in $\mathbb{R}^n$.

The set cardinality, $card(C^d_{cc})$ is not possible with the set of closed and convex sets. A convex set joined by intersection leaves out points when also ~~clos~~ closed set.

4.16 The VC-dimension of a set of all polygons in $\mathbb{R}^2$ is $2n+1$, for a quadrilateral on a circle, $2(4)+1=9$.

4.17: $f_1(X) = sign(\sin(tx))$; $t \in \mathbb{R}$; $f_t : [-1,1] \to \mathbb{R}[t \in \mathbb{R}]$

If $x = 2^{-n}$, then a binary expansion about $\sin(tx)$ is infinite through "Cantor's diagnol argument," so VC-measure$(f_1(X)) = \infty$.