Chapter 5: Metric Entropy and its uses:

(Bounded Set)

A set covered by finitely same-sized subsets

5.1 $C([0,1], b)$ where $f^j \le \|b\|$ is not a bounded set for all convex functions and $\|f^j - f^{j+1}\| \ge 1/2$.

If $f^j = \frac{1}{2^j}$, $\|f^0 - f^1\| = 1/2$

$\|f^1 - f^2\| = 1/4$

$\|f^2 - f^3\| = 1/8$

$\vdots$

$\lim_{j \to \infty} \|f^j - f^{j+1}\| = \|f^\infty\| = 0$

The convex set $f_j^j \le \|b\|$ lowers below $1/2$.

(Covering Number)

A δ-cover of a set $\mathbb{T}$ with a metric set $\rho = \{\theta^1, \dots \theta^N\} \subset \mathbb{T}$, is $\rho(\theta, \theta^i) \le \delta$. This δ-covering number $N(\delta : \mathbb{T}, \rho)$ is a cardinality of the smallest δ-cover.

(Packing Number)

A δ-pack of a set $\mathbb{T}$ with a metric set $\rho = \{\theta^1, \dots, \theta^N\} \subset \mathbb{T}$ is $\rho(\theta, \theta^i) > \delta$. This δ-packing number $M(\delta ; \mathbb{T}, \rho)$ is the cardinality of the largest δ-packing.

5.2. Proof of Lemma 5.5: $M(2\delta : \mathbb{T}, \rho) \le N(\delta ; \mathbb{T}, \rho) \le M(\delta ; \mathbb{T}, \rho)$

An example of unit cubes on an interval $[-1, 1]$ in $\mathbb{R}$ where intervals $L = \lfloor \frac{1}{\delta} \rfloor + 1$ divide the large interval for $L = \{1, 2, \dots, L\}$ by at most 2δ.

The covering number $N(\delta : [-1, 1], |\cdot|) \le \frac{1}{\delta} + 1$

The packing number $M(2\delta : [-1, 1], |\cdot|) \le \frac{1}{2\delta} + 1$

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

packing number at 2δ $\le$ covering number at δ $\le$ packing number at δ.

(Hamming Metric)

$$\rho_H(\theta,\tilde{\theta}) = \frac{1}{d}\sum_{j=1}^{d} \mathbb{1}[\theta_j \ne \theta_i]$$

5.3. From the covering of a binary hypercube:

$$\frac{\log N_H(\delta;\mathbb{H}^d)}{\log 2} \le d(1-\delta)$$

$$\frac{\log M_H(\delta;\mathbb{H}^d)}{\log 2} \le d(1-\delta/2)$$

$$M_H(\delta;\mathbb{H}^d) \le 2^{d(1-\delta/2)}$$

$$\le \frac{2^d}{\sum_{k=1}^{d\delta/2}\binom{d}{k}}$$

$$\frac{1}{M_H(\delta;\mathbb{H}^d)} \le \frac{\sum_{k=1}^{d\delta/2}\binom{d}{k}}{2^d}$$

$$\le P\left[\sum_i X_i \le \frac{d\delta}{2}\right]$$

$$\le (d+1)\, P\left[\sum X_i \le \frac{d\delta}{2}\right]$$

$$\le (d+1)\, e^{-2d(1-\delta/2)^2}$$

Problem 2.9/2.10
$\frac{1}{d}\log P[\sum X_i \le s] = -D(s\|1/2)$
Upper bound has a
$\frac{1}{d}\log (d+1) P[\sum X_i \le s] = -D(s\|1/2)$

$$-\log M_H(\delta;\mathbb{H}^d) \le \log P\left[\sum X_i \le \frac{d\delta}{2}\right] + \log(d+1)$$

$$\frac{\log M_H(\delta;\mathbb{H}^d)}{d} \ge D(\delta/2 \,\|\, 1/2) + \frac{1}{d}\log(d+1)$$

5.4.

a) $X_i = \{X_1, \ldots, X_n\} \geq \delta$ from Definition 5.4 about packing numbers

$$P[X \in S_i \cap S_j]^c = 1 - P[X \in S_i \cap S_j]$$
$$\geq 1 - \binom{N}{2}(1-\delta)^n$$

b) If $N \geq 2$ and $n = \frac{3\log N}{\delta}$, then

$$P[X \in S_i \cap S_j] \geq 1 - \binom{N}{2}(1-\delta)^n$$
$$\geq 1 - \binom{N}{2}e^{-n\delta}$$
$$\geq 1 - \binom{N}{2}e^{-3\log N}$$
$$\geq 1 - \frac{N(N-1)}{2N^3}$$

Notes: $e^{-xn} \simeq (1-x)^n$

Why? Binomial Expansion:

$$(a+b)^n = {}^nC_0 a^n + {}^nC_1 a^{n-1} b^1 + \cdots + {}^nC_{n-1} a^1 b^{n-1} + {}^nC_n b^n$$

$$(1-x)^n = {}^nC_0 1^n - {}^nC_1 1^{n-1} x + {}^nC_2 1^{n-2} x^2$$
$$= 1 - nx + \frac{nx^2}{2} + \cdots$$

$$e^{-x} \simeq 1 - nx + \frac{nx^2}{2} + \cdots$$

$e^{-x} \simeq (1-x)^n$ up till last binomial term.

$$N(\delta; X \in S_i \cap S_j; X) \leq card(S(x^n)) \leq \left(\frac{en}{v}\right)^v \text{ from exercise } 4.18$$
$$\leq e^v \left(\frac{3\log N}{\delta v}\right)^v$$

c) $N(\delta; X \in S_i \cap S_j; X) \leq e^v \left(\frac{3\log N}{\delta v}\right)^v$

$$\leq (2v)^{2v-1}\left(\frac{3}{\delta}\right)^{2v}$$

This part, (c) remains without a correct relationship between N, v, and δ. The inequality isn't always true, but supposedly.

5.5.

a. $T \subseteq \mathbb{R}^d$; $G(T) = \frac{1}{\sqrt{2\pi}} e^{-T^2/2}$; $R(T) = P[\sum_{i=1}^n \varepsilon_i \cdot t_i > \| t_i \|]$

$$G(T) = \mathbb{E}[\textstyle\sum \varepsilon_i t_i \cdot g] \leq \mathbb{E}[g] \circ \mathbb{E}[\textstyle\sum \varepsilon_i t]$$
$$\leq \frac{2}{\sqrt{2\pi}} \circ R(T)$$

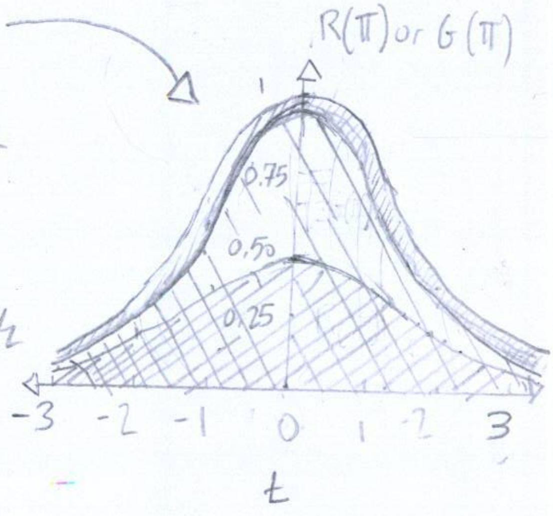
$$R(T) \leq \sqrt{\frac{\pi}{2}}\, G(T)$$

b) $R(\Pi) \le \sqrt{\frac{\pi}{2}}\, G(\Pi) \le \sqrt{\frac{\pi}{2}} \circ 2\sqrt{\log d}\, R(\Pi)$

because $R(\Pi) = P[\sum \varepsilon_i t > \|t\|] \le e^{-t^2/2}$

Similarly,

$$e^{-t^2/2} \le \sqrt{\frac{\pi}{2}} \circ \frac{1}{\sqrt{2\pi}} e^{-t^2/2} \le \sqrt{\frac{\pi}{2}} \circ (2\sqrt{\log d})\, e^{-t^2/2}$$



The Rademacher complexity assigns a negative one(-1) or positive one (+1) about a system. The probability outcome is about Gaussion.

(Holder's Inequality)

$\frac{1}{p} + \frac{1}{q} = 1$ for $p \ge 1$ and $q < \infty$

5.6. $G(\mathbb{B}_q^d(1)) = \mathbb{E}[\mathbb{B}_q^d(1)] = \mathbb{E}\left[\frac{\sum_{j=1}^{d} |\theta_j|^p}{d^q}\right]^{1/p} = \frac{1}{d^{1-1/q}} \mathbb{E}\left[\sum_{j=1}^{d} |\theta_j|^p\right]^{1/p} \le C_q$

$G(\mathbb{B}_q^d(1)) = d\,\mathbb{E}[\mathbb{B}_q^d(1)] = \frac{d}{d^{1/p}} \mathbb{E}\left[\sum_{j=1}^{d} |\theta_j|^p\right]^{1/p} = \sqrt{\frac{2}{\pi}}$ when $p = 1$

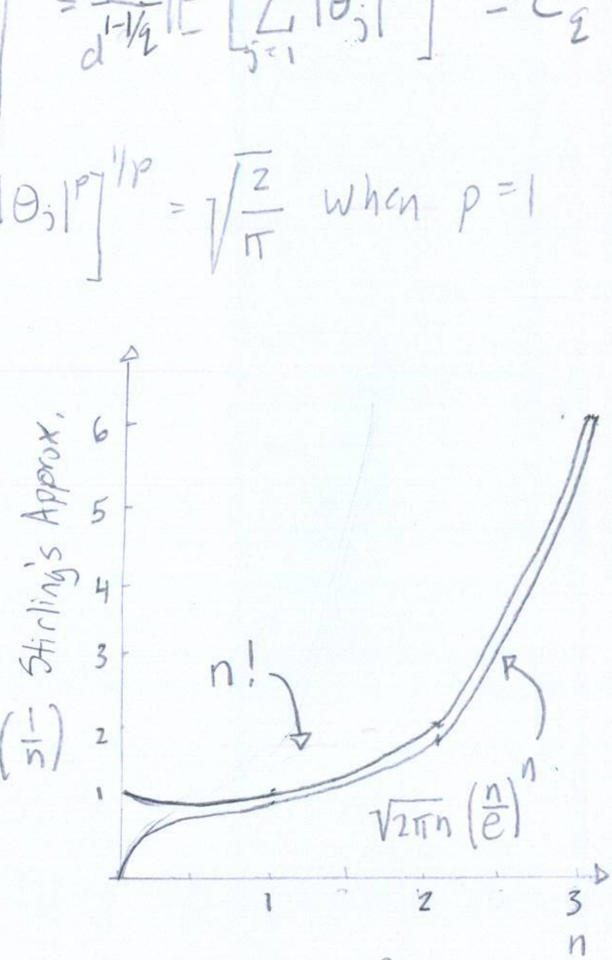
$$\sqrt{\frac{2}{\pi}} \le \frac{G(\mathbb{B}_q^d(1))}{d^{1-1/q}} \le C_q$$

(Sterling Approximation)

$\ln n! = n \ln n - n + O(\ln n)$

$\log_2 n! = n \log_2 n - n \log_2 e + \frac{1}{2}\log(2\pi n) + O(\frac{1}{n})$

$n! = \sqrt{2\pi n}\left(\frac{n}{e}\right)^n + O(2^{1/n})$



(Binomial and Gaussion)

$$Bin(n, p, X) = \sum_{k=1}^{n} \binom{n}{k} p^k (1-p)^{n-k} = N(X \mid np, np(1-p)) = \frac{1}{\sqrt{2\pi np(1-p)}} e^{\frac{-(X-np)^2}{2np(1-p)}}$$ for $p = 1/2$

5.7

a) $\Pi^d(s) := \{\theta \in \mathbb{R}^d \mid \|\theta\|_0 \leq s, \|\theta\|_2 \leq 1\}$

$$G(\Pi^d(s)) = \mathbb{E}[\{\theta \in \mathbb{R}^d \mid \|\theta\|_0 \leq s, \|\theta\|_2 \leq 1\}] = \mathbb{E}\left[\left(\sum_{j=1}^{d} |\theta_j|^q\right)^{1/q}\right]$$

$$= \mathbb{E}\left[\max_{|S|=1..d} \|\theta_S\|\right] = \mathbb{E}\left[\max_{\|S\|=s} \|w\|\right]$$

b) $$P[\|w\|_2 \geq s+\delta] = P\left[\|w\| \geq \mathbb{E}\left[\max_{\|S\|=s} \|w\|\right] + \delta\right]$$

$$= P[\|w\| \geq \sqrt{s} + \delta]$$

$$\leq e^{-\delta^2/2}$$

c) $\mathbb{G}(\Pi(0)) = Bin(0, p, x) = \underline{1} \leftarrow$ A Gaussian and Binomial relationship

$\leq \sqrt{s} \leftarrow \mathbb{E}[\max\|w\|]$

$\leq 2\sqrt{\log d} + \sqrt{s}$

$\leq \sqrt{s \log\left(\frac{de}{s}\right)} \leftarrow$ Sterling's Approximation

(Gilbert-Varshamov Lemma)

The maximum size of a q-ary code with n-length and distance q is:

$$\frac{\text{partial combination}}{\text{total combinations}} = \frac{q^n}{\sum_{k=0}^{d-1} \binom{n}{k}(q-1)^j} \leq A_q(n,d)$$

When $2 \leq d \leq n$ $\quad \leq \frac{q^n}{q^{n-k}} \leq q^k \leq A_q(n,d)$

(Sudokov's Minimization)

(scaled)

$X^n = \{X_1, \ldots, X_n\}$: $F(X^n) = \{f(X_1), \ldots f(X_n) \mid f \in \mathcal{F}\} \subseteq \mathbb{R}^d$; $F(X^n)/\sqrt{n}$

5.8 a) $\frac{1}{\sqrt{n}} G\left(\frac{q^n}{\sum_{k=0}^{d-2}\binom{n}{k}(q-1)^5}\right) = \frac{G(q^{n=n+k})}{\sqrt{n}} = \frac{1}{\sqrt{n}} = \frac{1}{\sqrt{2}}$ at lowest bound of Gilbert-Varshamov lemma.

b) See problem 5.7b

5.9. $\mathcal{E} := \{(\theta_j)_{j=1}^{\infty} \| \sum_{j=1}^{\infty} \frac{\theta_j^2}{\mu_j^2} \leq 1\}$

a) Lower Bound $\leq$ $\mathbb{E}$xpected $\leq$ Upper Bound

$$\inf_{\theta\in\mathcal{E}} G(\mathcal{E}) \leq G(\mathcal{E}) \leq \sup_{\theta\in\mathcal{E}} G(\mathcal{E})$$

$$\mathbb{E}\left[\sum_{j=1}^{\infty}[\varepsilon_i(\mu_i^2)]\right]^{1/2} \leq G(\mathcal{E}) \leq \mathbb{E}\left[\sum_{j=1}^{\infty}(\mu_i^2)\right]^{1/2}$$

$$\mathbb{E}[\varepsilon_i]\,\mathbb{E}\left[\sum_{j=1}^{\infty}(\mu_i^2)\right]^{1/2} \leq G(\mathcal{E}) \leq \left(\sum_{j=1}^{\infty}\mu_j^2\right)^{1/2}$$

$$\sqrt{\frac{2}{\pi}}\left(\sum_{j=1}^{\infty}\mu_j\right)^{1/2} \leq G(\mathcal{E}) \leq \left(\sum_{j=1}^{\infty}\mu_j^2\right)^{1/2}$$

b) $\tilde{\mathcal{E}}(r) := \mathcal{E} \cap \{(\theta_j)_{j=1}^{\infty} \mid \sum_{j=1}^{\infty}\theta_j^2 \leq r^2\}$

Method #1: Lower Bound $\leq$ $\mathbb{E}$xpected $\leq$ Upper Bound

$$\sqrt{\frac{2}{\pi}}\,\sigma \leq G(\tilde{\mathcal{E}}(r)) \leq \sigma$$

$$\sqrt{\frac{2}{\pi}}\left[\sum_{j=1}^{\infty}\theta_j^2\left(\frac{1}{r^2}+\frac{1}{\mu_i^2}\right)\right]^{1/2} \leq G(\tilde{\mathcal{E}}(r)) \leq \left[\sum_{j=1}^{\infty}\theta_j\left(\frac{1}{r}+\frac{1}{\mu_i}\right)\right]^{1/2}$$

Method #2: Lower Bound $\leq$ $\mathbb{E}$xpected $\leq$ Upper Bound

$$G(\inf_{r\in\mathcal{E}}\tilde{\mathcal{E}}(r)) \leq G(\tilde{\mathcal{E}}(r)) \leq G(\sup_{r\in\mathcal{E}}\tilde{\mathcal{E}}(r))$$

$$\mathbb{E}[\varepsilon_i]\,\mathbb{E}[\tilde{\mathcal{E}}(r)]^{2\,1/2} \leq G(\tilde{\mathcal{E}}(r)) \leq \mathbb{E}[\tilde{\mathcal{E}}(r)]^{2\,1/2}$$

$$\sqrt{\frac{2}{\pi}}\left[\sum_{j=1}^{\infty}\theta_j^2\left(\frac{1}{r^2}+\frac{1}{\mu_i^2}\right)\right]^{1/2} \leq G(\tilde{\mathcal{E}}(r)) \leq \left[\sum_{j=1}^{\infty}\theta_j^2\left(\frac{1}{r^2}+\frac{1}{\mu_i^2}\right)\right]^{1/2}$$

5.10 $\mathbb{P}[|Z-\mathbb{E}[Z]| \geq \delta]$ where $Z = \sup_{\theta \in \mathbb{T}}$ and $\sigma^2 = \sup_{\theta \in \mathbb{T}} var(X_\theta)$

$$\mathbb{P}\left[\left|1 + \lambda\mathbb{E}[Z] + \frac{\lambda^2\mathbb{E}[Z^2]}{2} + \cdots - \lambda\mathbb{E}[Z]\right| \geq \lambda\delta\right]$$

$$\leq \mathbb{P}\left[\left|\frac{\lambda^2\mathbb{E}[Z^2]^2}{2}\right| \geq \lambda\delta\right]$$

$$\mathbb{P}\left[\left|\frac{\lambda^2\mathbb{E}[Z^2]^2}{2}\right| - \lambda\delta \geq 0\right]$$

$$\underset{\lambda}{\operatorname{argmin}} \left\{\left|\frac{\lambda^2\mathbb{E}[Z^2]^2}{2}\right| - \lambda\delta\right\} = 0$$

$$\lambda^* = \delta / \mathbb{E}[Z^2]^2$$

$$\mathbb{P}\left[\left|\frac{\lambda^{*2}\mathbb{E}[Z^2]^2}{2}\right| \geq \lambda^*\delta\right] = \mathbb{P}\left[\left|\frac{\delta^2\mathbb{E}[Z^2]^2}{\mathbb{E}[Z^2]^4 2}\right| \geq \frac{\delta^2}{\mathbb{E}[Z^2]^2}\right]$$

$$= \mathbb{P}\left[-\frac{\delta^2}{\mathbb{E}[Z^2]^2} \leq \frac{\delta^2}{2\mathbb{E}[Z^2]^2} \leq \frac{\delta^2}{\mathbb{E}[Z^2]^2}\right]$$

$$= 2 \cdot \mathbb{P}\left[\frac{\delta^2}{2\mathbb{E}[Z^2]^2} \geq \frac{\delta^2}{\mathbb{E}[Z^2]^2}\right]$$

$$\therefore = 2\,\mathbb{P}\left[e^{\frac{\delta^2}{2\mathbb{E}[Z^2]^2}} \geq e^{\delta^2/\mathbb{E}[Z^2]^2}\right]$$

$$\mathbb{P}[|Z-\mathbb{E}[Z] \geq \delta] \leq 2 \cdot e^{-\delta^2/2\sigma^2}$$

(Von Neumann Trace and Inequality)

$$Tr[A \cdot B] \leq \sum_{i=1}^{n} \sigma_i(A)\sigma_i(B) \quad \text{-or-} \quad Tr[A \cdot B] \leq \sum_{i=1}^{n}\sum_{j=1}^{d} \sigma_i(A) \cdot \sigma_j(B)$$

(Proposition 5.17 - One step Discretization)

$$\underbrace{\mathbb{E}[\sup(X_\theta - X_{\bar{\theta}})]}_{\text{Upper Bound}} \leq \underbrace{2 \cdot \mathbb{E}[\sup(X_\gamma - X_{\gamma'})]}_{\delta\text{-covering Number}} + \underbrace{4\sqrt{D^2 \log N(\delta ; \mathbb{T})}}_{\text{Error}}$$

Upper Bound ≤ δ-covering Number + Error

Computational problems fit a packing problem with analysis about error.

5.11 a) $W \in \mathbb{R}^d \mid W_{ij} = N(0,1) \cong W_{ij} \{-1, +1\}$; $\|W\|_2 = \sup_{v \in S^{d-1}} \|W_v\|_2$

Where $S^{d-1} = \{ v \in \mathbb{R}^d \mid \|v\| = 1 \}$

$$\mathbb{M}^{n,d}(1) = \{ \Theta \in \mathbb{R}^{n \times d} \mid \text{rank}(\Theta) = 1, \|\Theta\|_F, 1 \}$$

$$|||W||| = \sup \sum_{i=1}^{n} \sum_{j=1}^{d} W_{ij} \, \Theta_{ij}$$

$$= \sup \sum X_{ij} \quad \text{where} \quad X_{ij} = Tr[W_{ij} \Theta_{ij}]$$

b) If $X_0 = \langle\langle W, \Theta \rangle\rangle$, then $X - X_{\bar{\theta}} = \langle\langle W, \Theta - \Theta_{\theta} \rangle\rangle$

c) $$\langle T - T', W \rangle = \frac{\sum_{i=1}^{1} \sum_{j=1}^{2} \sigma_j (T - T') \sigma_i (W)}{\sqrt{n \circ d}} \leq \delta \, \mathbb{E}[\|W\|]$$

$$\mathbb{E}\left[\sup \langle T - T', W \rangle\right] \leq \sqrt{2} \, \delta \, \mathbb{E}[\|W\|]$$

d) From example 5.2 and 5.6: $N(\delta, \mathbb{M}^{n \times d}(1)) \leq \left(1 + \frac{1}{\delta}\right)^{(n+d)}$

$$\log N(\delta; \mathbb{M}^{n \times d}(1)) \leq (n+d) \log\left(1 + \frac{2}{\delta}\right)$$

(Sudakov - Fernigic Inequality)

If $\mathbb{E}[(X_i - X_j)^2] \leq \mathbb{E}[(Y_i - Y_j)^2]$ for all i, j

then $\mathbb{E}[\max_{j=1..N} X_j] \leq \mathbb{E}[\max_{j=1..N} Y_i]$

(Gaussian Contraction Inequality)

For any set $\mathbb{T} \subseteq \mathbb{R}^d$ and a family centered 1-Lipshitz

$$\mathbb{E}\left[\sup_{\theta \in \mathbb{T}} \sum_{j=1}^{d} W_j \, \phi_j(\theta_j)\right] \leq \mathbb{E}\left[\sup \sum_{j=1}^{d} W_j \theta_j\right]$$

$$G(\phi(\mathbb{T})) \leq G(\mathbb{T})$$

5.12. $\mathbb{E}[(X-X_0)^2] \leq \mathbb{E}[(Y-Y_0)^2]$

$\mathbb{E}[X^2-2XX_0+X_0^2] \leq \mathbb{E}[Y^2-2YY_0+Y_0^2]$ because means at zero $(X_0=Y_0=0)$

$\mathbb{E}[(\sum w_i \phi_i(\theta))^2] \leq \mathbb{E}[(\sum w_i \theta_i)^2]$

$G(\Phi(\mathbb{T})) \leq G(\mathbb{T})$

5.13 $\mathbb{E}[|||W|||]_2 = \mathbb{E}[\sup_{\theta \in W} \langle\langle W,\theta\rangle\rangle]$

$\log M(\delta; M^{n,d}(1), |||\cdot|||) = (n+d)\log\frac{1}{\delta}$ from Example 5.2/5.6.

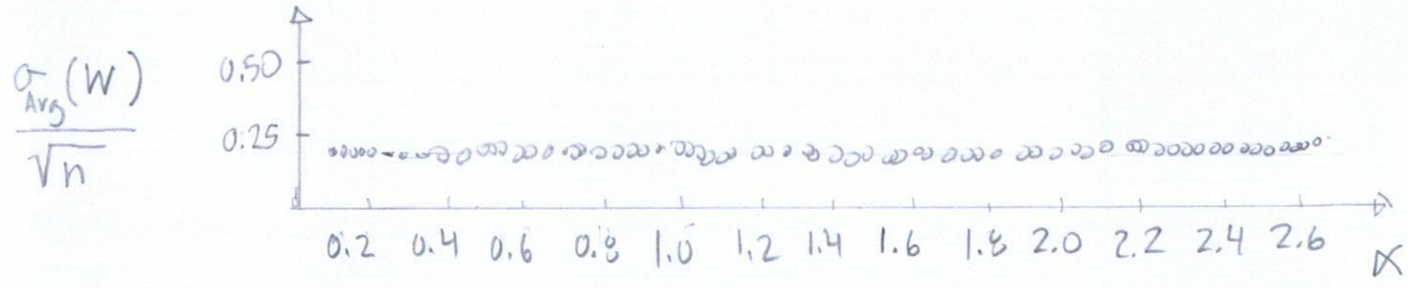
5.14a) Gaussian random values as a matrix columns grow in size.

Notes: Precision become important with the random Gaussian. "Box-Muller" method is a common simulation for a random normal distribution. Also, the large matrix sizes crushed memory at 55mb and 100% CPU capability with an 8-core, 3.2MHz central processor.

```
//gcc randomGaussianMatrix.c -std=c99 -lm
#include <stdio.h>
#include <math.h>
#include <stdlib.h>
#define M_PI = 3.14159265358

int main(){
  int n=1000, m=100, T=20;
  int t,i,j,k,d;
  float alpha;
```

```
float means[100]
float **W = (float**)malloc(2700*sizeof(float*));
for(i=0;i<2700;i++){
  W[i]=(float*)malloc(2700*sizeof(float));
}
for(int t=0;t<T;t++){
  for(k=0;k<100;k++){
    alpha=0.1+k*0.025;
    d=(int)(alpha x n);
    max=0;
    for(i=0;i<n;i++){
      for(j=0;j<d;j++){
        W[i][j]=sqrt(-2log((double)rand()/(double)RAND_MAX))    \\
                x(sin(2*M_PI*(double)rand()/(double)RAND_MAX));
        if(W[i][j]>max){
          max=W[i][j];
        }
      }
    }
    max/=sqrt(n);
    means[k]+=max;
  }
}
for(k=0;k<100;k++){
  means[k]/=T;
  printf("(%f,%f)\n",0.1+k*0.025,means[k]);
}
for(i=0;i<2700;i++){
  free(W[i]);
}
return 0;
}
```



- A plot about an average random Gaussian value as matrix $(n,d)=(n,\alpha n)$ grows. with mean zero and Standard deviation one.
- Alternative investigations are about non-zero Gaussian means, non-symmetric standard deviation, and bivariate parameters.
- "White noise" has a baseline, average intensity, or background-level, no matter the size.

b) $\sigma_{max}(W) = \sup_{u \in S^{n-1}} \sup_{v \in S^{d-1}} u^T \cdot W \cdot v$ is an "S.V.D." or singular value decomposition.

where $U_{100 \times 100}$, $W_{100 \times 1}$, and $V_{1 \times 1}$.

An analysis with $\sigma_{max}(W)_{100 \times 1}$ shows the components:

$W_{100 \times 1} = [47.52 \ 0 \cdots 0 \ 0 \ 0]$

$V_{1 \times 1} = [1]$

c) $\mathbb{E}[\sigma_{max}(W)] = \mathbb{E}[\sup U^T W v]$

$= \mathbb{E}\left[\sum_{i=1}^{n} g_i u + \sum_{j=1}^{d} h_j v\right]$

$\leq \sqrt{n} + \sqrt{d}$

A "plug-and-chug" attempt with part (a, b) data - a Gaussian value in a random Gaussian matrix is below $\sqrt{n} + \sqrt{d}$.

d) $P\left[|\sigma_{max}(W)/\sqrt{n}| \geq 1 + \sqrt{\frac{d}{n}} + t\right]$

$$= P\left[|1 + \lambda \mathbb{E}[\sigma_{max}(W)] + \frac{\lambda^2 \mathbb{E}[\sigma_{max}(W)^2]^2}{2}| \geq \lambda \mathbb{E}[\sigma_{max}(W)] + \lambda t\right]$$

$$\leq P\left[|\frac{\lambda^2 \mathbb{E}[\sigma_{max}(W)^2]^2}{2}| \geq \lambda t\right]$$

$$\leq 2 \cdot P\left[\frac{\lambda^2 \mathbb{E}[\sigma_{max}(W)^2]^2}{2} \geq \lambda t\right]$$

$$\arg\min_{\lambda} \left\{ \frac{\lambda^2 \mathbb{E}[\sigma_{max}(W)^2]^2}{2} - \lambda t \right\} = 0$$

$$\lambda^* = \frac{t}{\mathbb{E}[\sigma_{max}(W)^2]^2}$$

$$2 \cdot P\left[|\sigma_{max}(W)/\sqrt{n}| \geq 1 + \sqrt{\frac{d}{n}} + t\right] := 2 \cdot P\left[e^{|\sigma_{max}(W)/\sqrt{n}|} \geq e^{1 + \sqrt{\frac{d}{n}} + t}\right]$$

$$\leq 2 \cdot e^{-\frac{nt^2}{2}}$$