باسمه تعالی

دانشگاه صنعتی شریف

دانشکده مهندسی برق

گروه دکتر یاسایی – آنالیز احتمالاتی در ابعاد بالا
نیم سال دوم ۱۴۰۲–۱۴۰۱

تمرین سری دوم

نگارنده: امیرحسین باقری ۹۸۱۰۵۶۲۱

فهرست مطالب

۱ سوال ۱ ۳
۱.۱ الف ۳
۲.۱ ب ۳
۳.۱ ج ۴
۴.۱ د ۵

۲ سوال ۲ ۶
۱.۲ الف ۶
۲.۲ ب ۶
۳.۲ ج ۶
۴.۲ د ۶
۵.۲ ذ ۸
۶.۲ و ۸

۳ سوال ۳ ۸
۱.۳ الف ۸
۲.۳ ب ۱۰
۳.۳ ج ۱۰
۴.۳ د ۱۰
۵.۳ ذ ۱۰
۶.۳ ه ۱۱
۷.۳ و ۱۱
۸.۳ ی ۱۱

۴ سوال ۴ ۱۲

۵ سوال ۵ ۱۲

۶ سوال ۷ ۱۳
۱.۶ الف ۱۳
۲.۶ ب ۱۳

۷ **سوال ۸** **۱۳**
۱.۷ الف ۱۳
۲.۷ ب ۱۴

۸ **سوال ۱۰** **۱۴**
۱.۸ الف ۱۴
۲.۸ ب ۱۵
۳.۸ ج ۱۵

۱ سوال ۱

۱.۱ الف

$$N^{ext}(k,d,\epsilon) \leq N(k,d,\epsilon) \leq N^{ext}(k,d,\frac{\epsilon}{2}) \tag{1}$$

The part $N^{ext}(k,d,\epsilon) \leq N(k,d,\epsilon)$ is trivial to prove, By definition we have

$$N(d,k,\epsilon) = inf_n\{(x_1,...,x_n)|x_i \in set\ ,\forall p \in se\ \exists x_i \in set\ such\ d(x_i,p) \leq \epsilon\} \tag{2}$$

so by 2 if we can choose x_i from outside of set the number of point needed for covering the set should not exceed number needed when we were forced to choose from inside the set only because the infimum of set should be less than infimum of the subset of the set. so we take left inequality in 1 for granted.

now we want to prove right side of ineq 1 this comes from triangular inequality imagine we have $N^{ext}(k,d,\frac{\epsilon}{2})$ we have two claims

1. every point in covering set has a shared area with set k with radius of $\frac{\epsilon}{2}$
2. every $N^{ext}(k,d,\frac{\epsilon}{2})$ can be $N^{notoptimum}(k,d,\epsilon)$

It is very trivial that claim 2 will results in

$$N(k,d,\epsilon) \leq N^{ext}(k,d,\frac{\epsilon}{2}) \tag{3}$$

now let's prove our claims.

1. if claim one is not true this means there is a point out of our target set that there is no point in radius of $\frac{\epsilon}{2}$ which means we don't need that point in our covering set can reduce the number n to n-1 which is contradiction with definition
2. pick every point in $N^{ext}(k,d,\frac{\epsilon}{2})$ which is not in our target set by triangular ineq we know that every point in the ball of radius $\frac{\epsilon}{2}$ has the maximum distance of ϵ from each other so for every point out of target set pick any one random point is the shared area with the target set. now every point in the covering this new set is in the target set but it is not no more $N^{ext}(k,d,\frac{\epsilon}{2})$ but if we increase the radius to ϵ this set become non optimal $N^{notoptimum}(k,d,\epsilon)$ set. because for those points in the covering that were in the target set increasing does not change anything but for those outside of the set now we picked one point inside the set and by triangular ineq we now every point in that ball is at most at distance ϵ from this point so with radius ϵ this is going to be our new $N^{notoptimum}(k,d,\epsilon)$ which obviously follows

$$N(k,d,\epsilon) \leq N^{notoptimum}(k,d,\epsilon) = N^{ext}(k,d,\frac{\epsilon}{2}) \tag{4}$$

۲.۱ ب

The Counterexample is easily achievable in Figuer 1:

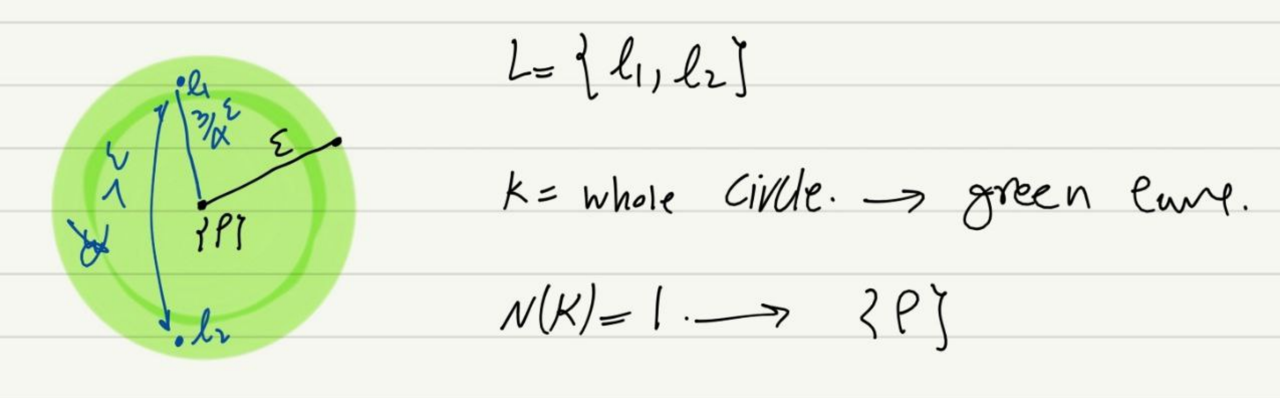


Figure 1: Counter example for problem 1.2

In Figure 1 The covering number for K is one but for L is 2

for proving

$$L \subset K \longrightarrow N(L, d, \epsilon) \leq N(K, d, \frac{\epsilon}{2}) \tag{5}$$

The proof for this part is very similar to the proof for right ineq in 1 we proof we can convert every $N(k, d, \frac{\epsilon}{2})$ to $N^{notoptimum}(L, d, \epsilon)$

consider a set of $\frac{\epsilon}{2}$-covering for K now first by triangular inequality we now every point in $\frac{\epsilon}{2}$ radius of each converging point will be at most distance of ϵ from each other. now consider set L for those covering points in the L set increase the radius to ϵ now set all points that are not in this converging set $\hat{L}$ so every point in this set is in the ball of radius $\frac{\epsilon}{2}$ to the initial covering set of K and there is a point in covering set of K that cover this point in $\hat{L}$ and it is not in L because if it was in L the point would not have been in $\hat{L}$ so now change that covering point to the point in $\hat{L}$ which is covered by that point and remove all point in radius of ϵ in set $\hat{L}$ and continue this until $\hat{L}$ is empty. we claim that no point will be left in our $\hat{L}$ before we run out of covering set of K.

if there still is point in $\hat{L}$ but no covering point in K is left which is not in L , this means that there was a point in K which was not in $\frac{\epsilon}{2}$-cover of K because of triangular inequality so this contradicts with definition of N and it is not possible so we have.

we proved that every $N(K, d, \frac{\epsilon}{2}$ is convertible to covering set(not optimal) for L with radius ϵ so our proof is complete and inequality 5 holds.

۳.۱ ج

The total number of all points in set is 2^n every point in covering with distance m has distance of 1,2,3,4,5,6,..., m with the points in covering so we we calculate number of points with distance 1,2,3,...,m we will have $S = \sum_{i=1}^{m} \binom{n}{i}$ so the N should cover all this points so number of covering set is at least the number of points that can cover points with distance of 1,2,3,4,...m with each other so we have : By applying union bond we have

$$total \leq \cup B_m \longrightarrow N \times S \longrightarrow \frac{2^n}{S} N(K, d_H, m) \tag{6}$$

$$\frac{2^n}{\sum_{i=1}^{m} \binom{n}{i}} \leq N(K, d_H, m) \checkmark \tag{7}$$

we claim that any maximal ϵ-packing is an ϵ -covering. Otherwise, there exists some $x \in K$ such that $d(x, x_i) > \epsilon$, so one can add x into the the packing and the resulting larger set is still an ϵ-packing so we have

$$N(K, d_H, m) \leq M(K, d_H, m) \checkmark \tag{8}$$

Consider M ball so in each ball there exist $\sum_{i=1}^{[\frac{m}{2}]}$ points which by definition are separated from other balls so we have

$$M \times \textit{point in each ball} \leq \textit{total points} \tag{9}$$

$$M \times (\sum_{i=1}^{[\frac{m}{2}]}) \leq 2^n \longrightarrow M \leq \frac{2^n}{\sum_{i=1}^{[\frac{m}{2}]}} \checkmark \tag{10}$$

۴.۱ د

first we reference question nine in assignments one which required proof for

$$P[Z_n \leq \delta n] \geq \frac{1}{n+1} e^{-nD(\delta, \alpha)} \tag{11}$$

where Zn is binomial with parameters n, α

for proving above we consider m to be the number of one and n-m number of zeros (why it is bigger because the m-1 , m-2 , ...) can be the case too.

$$P[Z_n \leq n\delta] \geq \binom{n}{m} \times \alpha^m (1-\alpha)^{n-m} \tag{12}$$

so we will have for $\frac{1}{n} log P[Z_n \leq n\delta]$

$$\delta = \frac{m}{n} \tag{13}$$

$$\frac{1}{n} log(\binom{n}{m}) + \delta log(\alpha) + (1-\delta) log(1-\alpha) \tag{14}$$

by assumption of that the maximum occurs at $m = \frac{n}{2}$ and bunch of derivative and writing! we will result in

$$\frac{1}{n} log P \geq \frac{1}{n} log(\binom{n}{m}) + \delta log(\alpha) + (1-\delta) log(1-\alpha) \tag{15}$$

$$\geq -\delta log(\delta) - (1-\delta) log(1-\delta) - \frac{log(n+1)}{n} + \delta log(\alpha) + (1-\delta) log(1-\alpha) \tag{16}$$

$$= \frac{-log(n+1)}{n} + \delta log(\frac{\alpha}{\delta}) + (1-\delta) log(\frac{1-\alpha}{1-\delta}) = -\frac{log(n+1)}{n} - D(\delta || \alpha) \tag{17}$$

which will result in

$$P[Z_n \leq \delta n] \geq \frac{1}{n+1} e^{-nD(\delta || \alpha)} \tag{18}$$

By definition of question we can consider d_H as sum of d Bernoulli random variable and use 18 we would have to prove :

$$\frac{1}{d+1} e^{-dD(\frac{\delta}{2} || \frac{1}{2})} \leq \frac{1}{M} \tag{19}$$

for proving above we address the fact that every point in space would have uniform probability to assign to one the points in M which will be $\frac{1}{M}$ now By definition of M we have that of noise which is distance of a point from its mapping is less than $\frac{\delta d}{2}$ we have the mapping so

$$\frac{1}{M} \geq P[Z_n \leq \frac{\delta d}{2}] \tag{20}$$

By applying 18 proof is complete ✓

۲ سوال ۲

۱.۲ الف

from Union bound it is trivial that

$$P(\cup A_k) \leq \sum P(A_k) \tag{21}$$

but we now probability is always less than one so

$$P(\cup A_k) \leq min(1, \sum P(A_k)) \tag{22}$$

By applying supplement rule for independent events and inequality given in question we have

$$P(\cup A_k) = 1 - \prod(1 - P(A_k)) \geq 1 - e^{-\sum P(A_k)} \geq (1 - e^{-1}) \times 1 \wedge \sum P(A_k) \tag{23}$$

۲.۲ ب

from question assumption we have

$$P[X_t \geq f^{-1}(log(|T|) + u)] \geq e^{f(f^{-1}(log(|T|)+u))} = e^{-(log(|T|)+u)} \tag{24}$$

now assume event E to be $X_t \geq f^{-1}(log(|T|) + u)$ from part one we can conclude that

$$P[\cup E] \geq (1 - e^{-1})(1 \wedge \sum P(E)) \geq (1 - e^{-1})(1 \wedge \sum_{j=1}^{|T|} e^{-(log(|T|)+u)}) \geq (1 - e^{-1})e^{-u} \tag{25}$$

when we want to calculate $P[\sup_{t \in T} X_t \geq f^{-1}(log(|T|) + u)]$ we can have union bound that when this event occurs means that all other events holds so

$$P[\sup_{t \in T} X_t \geq f^{-1}(log(|T|) + u)] \geq P[\cup E] \geq (1 - e^{-1})e^{-u} \tag{26}$$

to be continued

۳.۲ ج

by assumption we know

$$P[X_t \geq \frac{f^{-1}(2log(|T|)) + 2x}{2}] \geq e^{-f(\frac{f^{-1}(2log(|T|))+2x}{2})} \tag{27}$$

because f is convex we can conclude

$$-f(\frac{f^{-1}(2log(|T|)) + 2x}{2}) \leq \frac{1}{2}(f(f^{-1}(2log(|T|))) + f(2x)) \longrightarrow e^{-\cdots} \geq e^{-log(|T|) - \frac{1}{2}f(2x)} \tag{28}$$

following exactly like 26 25 we have

$$P[\sup_{t \in T} X_t \geq f^{-1}(log(|T|) + u)] \geq P[\cup E] \geq (1 - e^{-1})e^{-\frac{1}{2}f(2x)} \tag{29}$$

۴.۲ د

lower bound follows

$$E[\sup_{t \in T} X_t] = E[\sup_{t \in T} min(X_t, 0)] + E[\sup_{t \in T} max(X_t, 0)] \tag{30}$$

expectation of suprimum is always more than suprimum of expectation so

$$E[\sup_{t\in T} min(X_t,0)] + E[\sup_{t\in T} max(X_t,0)] \geq \sup_{t\in T} E[min(X_t,0)] + E[\sup_{t\in T} max(X_t,0)] \tag{31}$$

$$\geq \sup_{t\in T} E[min(X_t,0)] + \int_0^\infty P[\sup_{t\in T} max(X_t,0) \geq t]dt \geq \tag{32}$$

$$\sup_{t\in T} E[min(X_t,0)] + \int_0^{f^{-1}(log(|T))} P[\sup_{t\in T} max(X_t,0) \geq t]dt \tag{33}$$

now this part is tricky $\int_0^{f^{-1}(log(|T))} P[\sup_{t\in T} max(X_t,0) \geq t]dt$ if instead of t we set $P[\sup_{t\in T} max(X_t,0) \geq f^{-1}(log(|T|))]$ inequality below holds because for t less than $f^{-1}(log(|T|))$ it is zero

$$\int_0^{f^{-1}(log(|T))} P[\sup_{t\in T} max(X_t,0) \geq t]dt \geq \int_0^{f^{-1}(log(|T))} P[\sup_{t\in T} max(X_t,0) \geq f^{-1}(log(|T|))]dt \tag{34}$$

so we have

$$\sup_{t\in T} E[min(X_t,0)] + \int_0^{f^{-1}(log(|T))} P[\sup_{t\in T} max(X_t,0) \geq t]dt \geq \tag{35}$$

$$\sup_{t\in T} E[min(X_t,0)] + \int_0^{f^{-1}(log(|T))} P[\sup_{t\in T} max(X_t,0) \geq f^{-1}(log(|T|))]dt \tag{36}$$

$$\geq \sup_{t\in T} E[min(X_t,0)] + (f^{-1}(log(|T|)))(1-e^{-1}) \tag{37}$$

$$\longrightarrow E[\sup_{t\in T} X_t] \geq C_1(\sup_{t\in T} E[min(X_t,0)] + f^{-1}(log(|T|)))\checkmark \tag{38}$$

for upper bound we would have following

$$E[\sup_{t\in T} X_t] \leq E[max(0, \sup_{t\in T} x_t)] = \int_0^\infty P[\sup_{t\in T} X_t \geq t]dt \tag{39}$$

$$= \int_0^{g^{-1}(log(|T|))} P[\sup_{t\in T} X_t \geq t]dt + \int_{g^{-1}(log(|T|))}^\infty P[\sup_{t\in T} X_t \geq t]dt \tag{40}$$

now we should bound both integrals for first one easily follows

$$\int_0^{g^{-1}(log(|T|))} P[\sup_{t\in T} X_t \geq t]dt \leq \int_0^{g^{-1}(log(|T|))} 1dt = g^{-1}(log(|T|)) \tag{41}$$

and for second one we have by changing variable

$$\int_{g^{-1}(log(|T|))}^\infty P[\sup_{t\in T} X_t \geq t]dt = \int_0^\infty P[\sup_{t\in T} X_t \geq log(|T|)+t]dt \tag{42}$$

with rational assumption that size of T is at least one we have and by question assumption

$$\int_0^\infty P[\sup_{t\in T} X_t \geq log(|T|)+t]dt \leq \int_0^\infty e^{-g(g^{-1}(log(|T|))+t)}dt \tag{43}$$

multiply by |T|

$$\int_0^\infty e^{-g(g^{-1}(log(|T|))+t)}dt \leq \int_0^\infty |T|e^{-g(g^{-1}(log(|T|))+t)}dt = \int_0^\infty e^{-(g(g^{-1}(log(|T|))+t)-g(g^{-1}(log(|T|))))}dt \tag{44}$$

now because our function g is convex then

$$g(g^{-1}(log(|T|))+t) - g(g^{-1}(log(|T|))) \leq g\prime(g^{-1}(log|T|))t \longrightarrow \tag{45}$$

$$\int_0^\infty e^{-(g(g^{-1}(log(|T|))+t)-g(g^{-1}(log(|T|))))}dt \leq \int_0^\infty e^{-g\prime(g^{-1}(log|T|))t}dt = \frac{1}{g\prime(g^{-1}(log|T|))} \tag{46}$$

so by combining these two bound we get

$$E[\sup_{t\in T} X_t] \leq g^{-1}(log(|T|)) + \frac{1}{g\prime(g^{-1}(log|T|))} \tag{47}$$

by rational assumption that size T is at least one and g in increasing convex function which means that its derivative is also increasing we get

$$E[\sup_{t\in T} X_t] \leq g^{-1}(log(|T|)) + \frac{1}{g\prime(g^{-1}(0))} \leq C_2 g^{-1}(log(|T|))\checkmark \tag{48}$$

۵.۲ ذ

$$P[X \geq x] = \int_x^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2}{2}} dt = \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{(t+x)^2}{2}} dt \geq \int_0^\infty \frac{1}{\sqrt{2\pi}} e^{-\frac{t^2+x^2}{2}} dt \tag{49}$$

$$= \frac{1}{\sqrt{2\pi}} e^{-x^2} e^{-t^2} dt = \frac{1}{\sqrt{2\pi}} e^{-x^2} \times \frac{\sqrt{\pi}}{2} = \frac{e^{-x^2}}{2\sqrt{2}}\checkmark \tag{50}$$

۶.۲ و

we know from before that

$$P[X \geq x] \leq e^{\frac{-x^2}{2}} \longrightarrow g(x) = \frac{x^2}{2} \tag{51}$$

$$g^{-1}(x) = \sqrt{2x} \& g\prime = x \tag{52}$$

$$P[X \geq x] \geq \frac{e^{-x^2}}{2\sqrt{2}} = e^{-(x^2+log(2\sqrt{2}))} f(x) = x^2 + log2\sqrt{2} \& f^{-1}(x) = \sqrt{x - log2\sqrt{2}} \tag{53}$$

for lower bound

$$E[maxX_i] \geq (1-e^{-1})(\sup E[min(X_i,0)] + \sqrt{logn - log2\sqrt{2}}) \tag{54}$$

$$E[min(X_i,0)] \geq E[min(X,0)] = \int_{-\infty}^0 \frac{xe^{\frac{-x^2}{2}}}{\sqrt{2\pi}} = \frac{-1}{\sqrt{2\pi}} \longrightarrow \tag{55}$$

$$E[maxX_i] \geq \frac{1-e^{-1}}{2}(2\sqrt{log2^{\frac{-3}{4}}g}) - \frac{1}{\sqrt{2\pi}}\checkmark \tag{56}$$

for upper bound we have

$$E[MAXX_i] = \frac{1}{a}E[maxloge^{ax_i}] \geq \frac{1}{a}E[log\sum e^{aX_i}] \tag{57}$$

$$\leq \frac{1}{a}log(\sum E[e^{aX_i}]) = \frac{1}{a}log(nE[e^{aX}]) = \frac{1}{a}(logn + \frac{a^2}{2}) \tag{58}$$

now set $a = \sqrt{2logn}$ then $E[MAXX_i] = \sqrt{2logn}\checkmark$

۳ سوال ۳

۱.۳ الف

For lower bound we know that $|X_k - X_l| \geq \min_{t\neq k}|X_k - X_l|$ so it will results in

$$L_n \geq \sum_{k=1}^n \min_{t\neq k}|X_k - X_l| \tag{59}$$

$$E[L_n] \geq \sum_{k=1}^n E[\min_{t\neq k}|X_k - X_l|] \tag{60}$$

$$\tag{61}$$

we need to compute $E[\min_{t\neq k}|X_k - X_l|]$ we use Lebesgue for computing Expectation.

$$E[\min_{t\neq k}|X_k - X_l|] = \int_0^\infty P(\min_{t\neq k}|X_k - X_l| \geq t)dt \tag{62}$$

we can see that $P[\min_{t\neq k}|X_k - X_l| \geq t] = P(|X_k - X_l| \geq t)^{n-1}$ which would be $(1-\pi t^2)^{n-1}$ and because probability is always on the bound on t would become $1-\pi t^2 \geq 0 \longrightarrow t \leq \frac{1}{\sqrt{\pi}}$ so notice that t is always less than one

$$E[\min_{t\neq k}|X_k - X_l|] = \int_0^{\frac{1}{\sqrt{\pi}}} (1-\pi t^2)^{n-1} \tag{63}$$

now we do some reformation $E[\min_{t\neq k}|X_k - X_l|] = E[\sqrt{(\min_{t\neq k}|X_k - X_l|)^2}] \leq \sqrt{E[\min^2_{t\neq k}|X_k - X_l|]}$

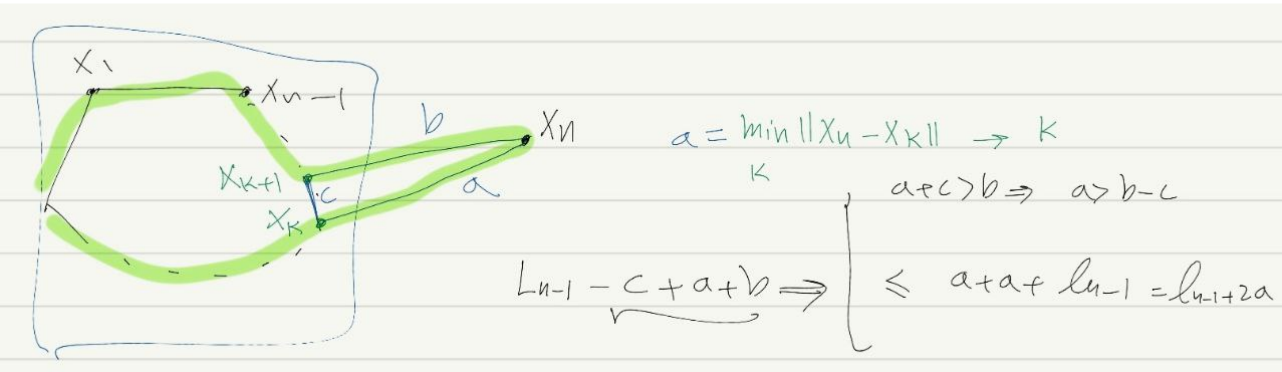
$$E[\min_{t\neq k}^{2}|X_k - X_l|] = \int_0^{\infty} P(\min_{t\neq k}|X_k - X_l|^2 \geq t)dt = \int_0^{\infty} P(\min_{t\neq k}|X_k - X_l| \geq \sqrt{t})dt \tag{64}$$

$$\int_0^{\infty} P(\min_{t\neq k}|X_k - X_l| \geq u)2udu = \int_0^{\frac{1}{\sqrt{\pi}}} (1-\pi u^2)^{n-1}2udu = O(\frac{1}{n}) \tag{65}$$

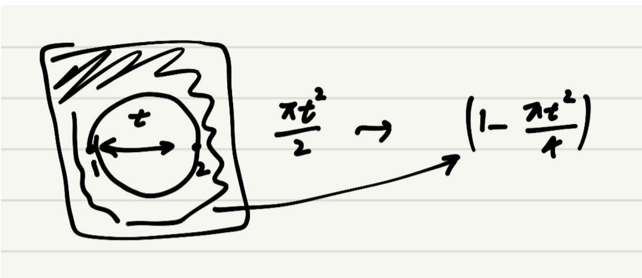
so we have $E[\min_{t\neq k}|X_k - X_l|] = E[\sqrt{(\min_{t\neq k}|X_k - X_l|)^2}] \leq \sqrt{E[\min^2_{t\neq k}|X_k - X_l|]} = O(\frac{1}{\sqrt{n}})$ then this would lead to

$$E[L_n] \geq \sum_{k=1}^{n} E[\min_{t\neq k}|X_k - X_l|] = nO(\frac{1}{\sqrt{n}}) = O(\sqrt{n}) \tag{66}$$

first we proof the upper bound inequality by triangular inequality



$$L_n \leq L_{n-1} + 2min|X_n - X_k| \longrightarrow E[L_n] \leq 2E[\sum_{k=1}^{n} \min_{l\leq k}|X_k - X_l|] \tag{67}$$



$$E[\min_{l\leq k}|X_k - X_l|] = \int_0^{\infty} P[\min_{l\leq k}|X_k - X_l| \geq t]dt \leq \int_0^{\frac{2}{\sqrt{\pi}}} (1-\frac{\pi t^2}{4})^{k-1} \longrightarrow \tag{68}$$

$$E[L_n] \leq 2\sum_{k=1}^{n} \int_0^{\frac{2}{\sqrt{\pi}}} (1-\frac{\pi t^2}{4})^{k-1}dt \tag{69}$$

just like part one we can assume the integral is order of $\frac{1}{\sqrt{n}}$ then it will gives us

$$E[L_n] \leq 2 * n * O(\frac{1}{\sqrt{n}}) = O(\sqrt{n}) \tag{70}$$

۲.۳ ب

changing a one point in direction would cause at most 2 × radius of square different because at worst case the point would force us to go $\sqrt{2}$ and return $\sqrt{2}$ so the function is differential bounded and by mcdiarmid it is $\sigma^2 \frac{n}{4}$ subg

$$|f(x1 : x_K : x_n) - f(x_1 : x_{k\prime} : x_n)| \leq 2\sqrt{2} \longrightarrow \tag{71}$$

$$(2\sqrt{2})^2 \frac{n}{4} = 2n \longrightarrow subg(2n) \tag{72}$$

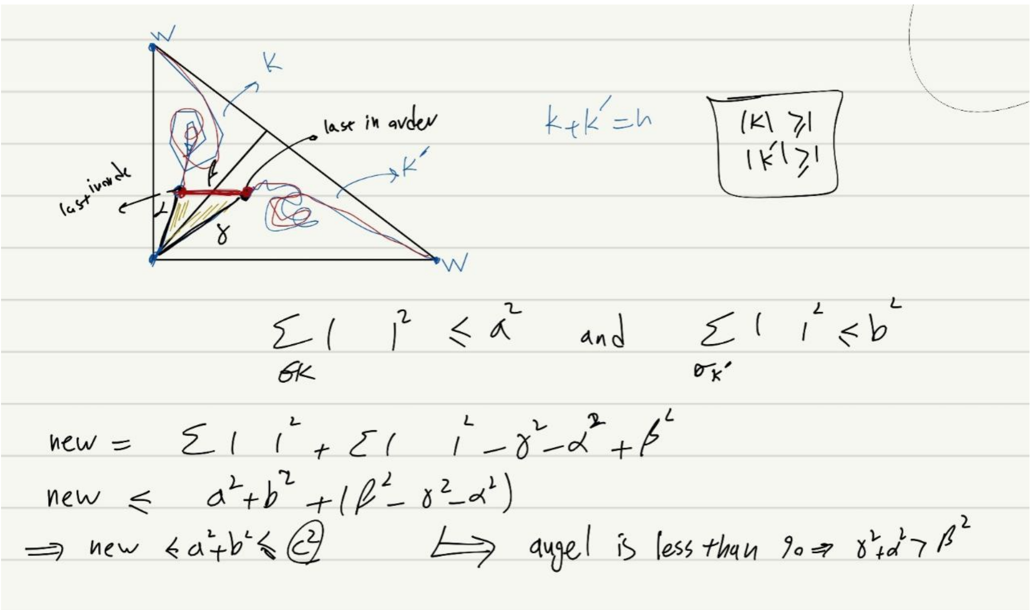
۳.۳ ج

have we law cosine by so ۹۰ than more is side two of angular the

$$c^۲ = a^۲ + b^۲ - ۲abcos(\theta) \longrightarrow c^۲ \geq a^۲ + b^۲ \checkmark \tag{۷۳}$$

۴.۳ د

by induction we have



$$\longrightarrow |v - X_{\sigma(1)}|^2 + \sum_{i=1}^{i=n-1} |X_{\sigma(i)} - X_{\sigma(i+1)}|^2 + |w - X_{\sigma(n)}|^2 \leq |v - w|^2 \checkmark \tag{74}$$

۵.۳ ذ

part previous using and triangle Right two to square the divide to need just we

$$|X_{\sigma_۱} - (۱, ۱)|^۲ + \sum_{i=۱}^{m-۱} |X_{\sigma_i} - X_{\sigma_{i+۱}}|^۲ + |X_{\sigma_m} - (۰, ۰)|^۲ \leq ۲ \tag{۷۵}$$

$$|X_{\sigma_{m+۱}} - (۰, ۰)|^۲ + \sum_{i=m+۱}^{n} |X_{\sigma_i} - X_{\sigma_{i+۱}}|^۲ + |X_{\sigma_n} - (۱, ۱)|^۲ \leq ۲ \tag{۷۶}$$

because the angel for connection point is less than ۹۰ degree

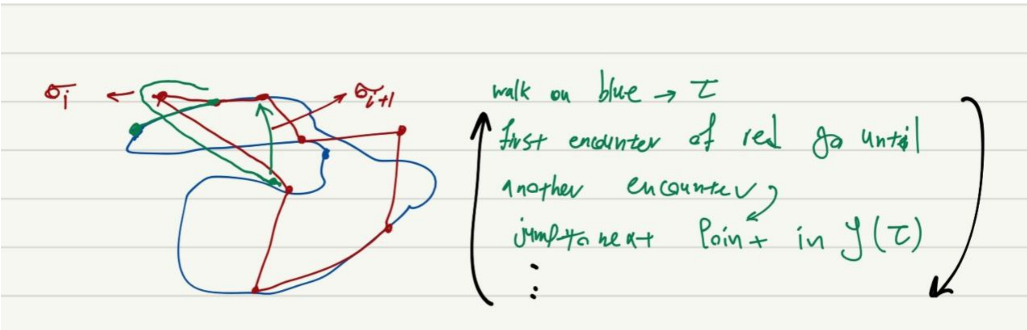
$$|X_{\sigma_۱} - X_{\sigma_n}|^۲ \leq |X_{\sigma_۱} - (۱,۱)|^۲ + |(۱,۱) - X_{\sigma_n}|^۲ \tag{۷۷}$$

so

$$\sum_{i=۱}^{n} |X_{\sigma_i} - X_{\sigma_{i+۱}}|^۲ \leq ۴ \checkmark \tag{۷۸}$$

۶.۳ ه

If we can prove that the path in y can be converted to path of union x,y we can prove inequality.



we walk on the t which involves y points to the encounter to the path σ we walk on σ and then when we arrive to the next encounter(we don't step on the encountering point) or end of path we jump to the next previous point in t path, remember we jump to the next point compare to the point we started to walk on σ so with this algorithm it is still a cycle and also it is obvious by triangular inequality that the extra path of t is less than sum of commute and commute is sum of point that are not in t and are in σ so we have

$$l_{2n}(x \cup y, p) \leq l_n(y, t) + 2\sum_{i=1}^{n} 1_{x \notin y} d_i(x, \sigma) \checkmark \tag{79}$$

۷.۳ و

this is quit easy because we already can use the result from previous part. first notice that

$$\min_p l_{2n}(x \cup y, p) \leq l_{2n}(x \cup y) \tag{80}$$

$$\min_\sigma l_n(x, \sigma) \leq \min_p l_{2n}(x \cup y, p) \tag{81}$$

no imagine the two path than are optimal answers for σ and t now exactly like previous part we can convert optimal path to union path and have

$$\min_\sigma l_n(x, \sigma) \leq \min_\sigma l_n(y, \sigma) + \sum_{i=1}^{n} 2d_i(x, \sigma_x) 1_{x_i \neq y_i} \tag{82}$$

۸.۳ ی

$$\min_\sigma l_n(x, \sigma) - \min_\sigma l_n(y, \sigma) \leq \sum_{i=1}^{n} 2d_i(x, \sigma_x) 1_{x_i \neq y_i} \tag{83}$$

by talagrand we have $\sum c_i^2 = 4 \times \sum d_i^2 = 4 \times 4 = 16$ so f is 16-subg.

۴ سوال ۴

this problem is solved in chapter 4.6 of vershynin but here we try different approach.

$$X = \frac{1}{m}A^T A - I \tag{84}$$
$$X^T = X \tag{85}$$
$$\tag{86}$$

consider ϵ-cover V with elements $v_i||v_i| = 1$ then

$$\forall v : v^T X v = (v_i + \delta)^T X (v_i + \delta) = v_i^T X v_i + v_i^T X \delta + \delta^T X v_i + \delta^T X \delta \tag{87}$$
$$\leq |v_i^T X v_i| + 2|X||\delta| + |X||\delta|^2 = |v_i^T X v_i| + |X|(\epsilon^2 + 2\epsilon) \longrightarrow |X| \leq \frac{\max_{v_i} v_i^T X v_i}{1 - (\epsilon^2 + 2\epsilon)} \tag{88}$$
$$v_i^T X v_i = \frac{1}{m}(\sum_j (A_j^T v_i)^T (A_j^T v_i) - 1) \tag{89}$$

so as A were subg and |v| are one then the result is subexp$(\frac{\sigma^2}{\sqrt{m}}, \frac{\sigma^2}{m})$ so we can conclude by Bernstein that

$$p[v_i^T X v_i \geq 16\sigma^2 u] \leq 2e^{-min(\frac{mu}{2}, \frac{m^2u^2}{2})} \tag{90}$$

so using previous definition we have

$$P[|X| \geq 16\sigma^2 u] \leq P[v_i^X v_i \geq 16\sigma^2(1 - 2\epsilon - \epsilon^2)u] \tag{91}$$

by setting $u = \frac{\sqrt{n}+t}{\sqrt{m}}$ and $\epsilon = \frac{1}{4}$ we can have

$$P[|B| \geq 16\sigma^2 max(\delta, \delta^2)] \leq e^{-t^2} \checkmark \tag{92}$$
$$\delta = \frac{\sqrt{n} + t}{\sqrt{m}} \tag{93}$$

in vershynin the similar approach will result in

$$P[|\frac{1}{m}|AX|^2 - 1| \geq \frac{\epsilon}{2}] \leq 2e^{-c1m\min(\frac{\epsilon^2}{k_1}, \frac{\epsilon}{k_2})} \tag{94}$$

since $max(\delta, \delta^2) = \frac{\epsilon}{k_2}$ then

$$\leq 2e^{-c1C^2(n+t^2)} \tag{95}$$

By uniosn bound we reach the proof

$$P[\max|\frac{1}{m}|AX|^2 - 1| \geq \frac{\epsilon}{2}] \leq 2e^{-t^2} \checkmark \tag{96}$$

۵ سوال ۵

this is van Handel 2.8 example we use Poincare and we have:

$$Var(f(Y)) \leq E[|\nabla f|^2] \tag{97}$$
$$f(Y) = \max_i \Sigma^{\frac{1}{2}} Y \tag{98}$$

by calculating gradient we have

$$E[|\nabla f(Y)|^2] = E[\sum_{i=0}^{n} (\Sigma_{jmax,i}^{\frac{1}{2}})^2] = E[\sum_{i=0}^{n} \Sigma_{jmax,i}] \tag{99}$$

By inequality that expectation is less than maximum we have

$$E[\sum_{i=0}^{n} \Sigma_{jmax,i}] \leq \max_j \sum_{i=0}^{n} \Sigma_{j,i} = \max_j Var(X_i) \tag{100}$$

$$\longrightarrow Var(\max_i X_i) \leq \max_j Var(X_i)\checkmark \tag{101}$$

۶ سوال ۷

۱.۶ الف

Efron–Stein inequality says :

consider X $= (X_1, ...X_n)$ and $X^i = (X_1, ..., X_i', X_n$ then $VAR(f(X)) \leq \frac{1}{2}\sum_{i=1}^{n} E[(f(X) - f(X^i))^2]$

$$\frac{1}{2}\sum_{i=1}^{n} E[((f(X) - \mu) - (f(X^i) - \mu))^2] = \frac{1}{2}\sum_{i=1}^{n} 2var_i(f) \longrightarrow \tag{102}$$

$$VAR(f(X)) \leq E[\sum_{i=1}^{n} Var_{x_i}(f(X))] \tag{103}$$

now we only need to prove that var of f when all input are fixed except i is not more than $\frac{1}{4}(b-a)^2$ which a is infimum and b is suprimum, We refrence the proof of subgussian in the class where we proved that the variance would not be more than that. so by this we have now

$$\frac{1}{2}\sum_{i=1}^{n} E[((f(X) - \mu) - (f(X^i) - \mu))^2] = \frac{1}{2}\sum_{i=1}^{n} 2var_i(f) \longrightarrow \tag{104}$$

$$VAR(f(X)) \leq E[\sum_{i=1}^{n} \frac{1}{4}(D_i f(X))^2] = \frac{1}{4}E[\sum_{i=1}^{n}(D_i f(X))^2]\checkmark \tag{105}$$

۲.۶ ب

By calculation is it easy that $D_i \leq 1$ so $Var(f) \leq \frac{1}{4}E[\sum 1] = \frac{n}{4}\checkmark$

for avg part we know that $B_n \geq \sum X_i$ By applying expectation we have $E[B_n] \geq n \times \frac{1}{2}\checkmark$

what this means is that we have concentration around $\frac{n}{2}$ and for example if we use $\frac{3n}{4}$then with probability approx 0.975 we are good to go!.

۷ سوال ۸

۱.۷ الف

Proof. By Jensen's inequality, we have for any $\lambda > 0$

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \leq \frac{1}{\lambda}\log \mathbf{E}\left[e^{\lambda \sup_{t\in T} X_t}\right] \leq \frac{1}{\lambda}\log\sum_{t\in T}\mathbf{E}\left[e^{\lambda X_t}\right] = \frac{1}{\lambda}log|T| * e^{\psi(\lambda)} \leq \frac{\log|T| + \psi(\lambda)}{\lambda}. \tag{106}$$

As $\lambda > 0$ is arbitrary, we can now optimize over λ on the right hand side. In the general casewe have to evaluate the infimum in

$$\mathbf{E}\left[\sup_{t\in T} X_t\right] \leq \inf_{\lambda>0}\frac{\log|T| + \psi(\lambda)}{\lambda} = \psi^{*-1}(\log|T|). \tag{107}$$

Suppose ψ to be invertible. Observe that, by the definition of ψ, $\psi^(z) + \psi(\lambda)\ /\lambda \geq z$ holds true for all $\lambda > 0$, and the inequality is achieved if we select λ as the optimizer in the definition of ψ. Setting $\psi^*(z) = \log|T|$ leads us to the conclusion. We must now demonstrate that ψ is indeed invertible. As ψ represents the supremum of linear functions, the mapping $x \mapsto \psi^(x)$ is convex and strictly increasing, save for those values of x where the maximum in the definition of ψ is reached at $\lambda = 0$. In other words, when $\lambda x - \psi(\lambda) \leq -\psi(0)$ for all $\lambda \geq 0$. By the first-order condition for convexity, this occurs if and only if $x \leq \psi'(0) = 0$. Furthermore, given that $\psi^(0) = 0$, we deduce that $x \mapsto \psi^(x)$ is convex, strictly increasing, and nonnegative for $x \geq 0$. Consequently, the inverse $\psi^{*-1}(x)$ is well defined for $x \geq 0$.

۲.۷ ب

By applying Union bound we get

$$\mathbf{P}\left[\sup_{t\in T} X_t \geq x\right] = \mathbf{P}\left[\bigcup_{t\in T}\{X_t \geq x\}\right] \tag{108}$$

continuing union bound we get

$$\mathbf{P}\left[\bigcup_{t\in T}\{X_t \geq x\}\right] \leq \sum_{t\in T} \mathbf{P}\left[X_t \geq x\right] \tag{109}$$

Recall from lemma 3.1 van Handel if $\log \mathbf{E}\left[e^{\lambda X_t}\right] \leq \psi(\lambda)$ for all $\lambda \geq 0$ and $t \in T$, then

$$\mathbf{P}\left[X_t \geq x\right] \leq e^{-\psi^*(x)} \quad \text{for all } x \geq 0, t \in T.$$

so

$$\sum_{t\in T} \mathbf{P}\left[X_t \geq x\right] \leq |T| e^{-\psi^*(x)} = e^{\log|T| - \psi^*(x)} \tag{110}$$

now just by subtituion we can have

$$\tag{111}$$

$$x = \psi^{*-1}(\log|T| + u) \tag{112}$$

$$\leq e^{\log|T| - \psi^*(\psi^{*-1}(\log|T|+u))} = e^{\log|T| - \psi^*(\psi^{*-1}(\log|T|+u))} = e^{-u} \longrightarrow \tag{113}$$

$$\mathbf{P}\left[\sup_{t\in T} X_t \geq \psi^{*-1}(\log|T| + u)\right] \leq e^{-u} \quad \forall u \geq 0 \checkmark \tag{114}$$

$$\tag{115}$$

۸ سوال ۱۰

۱.۸ الف

$$Z(\theta) = X_k \sin(\theta) + Y_k \cos(\theta) \tag{116}$$

$$Z\prime(\theta) = X_k \cos(\theta) - Y_k \sin(\theta) \tag{117}$$

First notice that those both $Z(\theta), Z\prime(\theta)$ are Gaussian random variable because they are linear combination of two Gaussian random variable. for proving independence of these two random variable we should prove that Cov of those two is equal to zero, remember that both X and Y are zero mean.

$$E[Z^T Z\prime] = E[X_k^2 \sin(\theta)\cos(\theta)] - E[X_k Y_k \sin^2\theta] + E[Y_k X_k \cos^2\theta] - E[Y_k^2 \sin(\theta)\cos(\theta)] = \tag{118}$$

$$\sin(\theta)\cos(\theta)(E[X_k^2] - E[Y_k^2]) = 0\checkmark \tag{119}$$

۲.۸ ب

this is from Wainwright lemma 2.27

Using the convexity of ϕ, we can establish the inequality

$$\mathbb{E}X\left[\phi\left(f(X)-\mathbb{E}Y[f(Y)]\right)\right] \leq \mathbb{E}X,Y[\phi(f(X)-f(Y))].$$

To further bound the right-hand side, we substitute the integral representation

$$f(X)-f(Y)=\int_0^{\pi/2}\frac{d}{d\theta}f(Z(\theta))d\theta=\int_0^{\pi/2}\left\langle\nabla f(Z(\theta)),Z'(\theta)\right\rangle d\theta,$$

where $Z'(\theta)\in\mathbb{R}^n$ denotes the elementwise derivative, a vector with the components $Z_k'(\theta)=X_k\cos\theta-Y_k\sin\theta$. This allows us to write

$$\mathbb{E}X\left[\phi\left(f(X)-\mathbb{E}Y[f(Y)]\right)\right] \leq \mathbb{E}X,Y\left[\phi\left(\int_0^{\pi/2}\left\langle\nabla f(Z(\theta)),Z'(\theta)\right\rangle d\theta\right)\right] \quad =\mathbb{E}X,Y\left[\phi\left(\frac{1}{\pi/2}\int_0^{\pi/2}\frac{\pi}{2}\left\langle\nabla f(Z(\theta)),Z'(\theta)\right\rangle d\theta\right)\right]$$

where the final step again uses the convexity of ϕ. By the rotation invariance of the Gaussian distribution, for each $\theta\in[0,\pi/2]$, the pair $(Z_k(\theta),Z_k'(\theta))$ is a jointly Gaussian vector, with zero mean and identity covariance. Therefore, the expectation inside the integral does not depend on θ, and we have

$$\frac{1}{\pi/2}\int_0^{\pi/2}\mathbb{E}_{X,Y}\left[\phi\left(\frac{\pi}{2}\left\langle\nabla f(Z(\theta)),Z'(\theta)\right\rangle\right)\right]d\theta=\mathbb{E}\left[\phi\left(\frac{\pi}{2}\langle\nabla f(\widetilde{X}),\widetilde{Y}\rangle\right)\right],$$

where $(\widetilde{X},\widetilde{Y})$ are independent standard Gaussian n-vectors. This completes the proof of the bound. ✓

۳.۸ ج

this is easily achievable by substitution and using L-Lipschitz properties

$$E[e^{\frac{\pi}{2}<\nabla f,Y>}]\leq E[e^{\frac{\pi}{2}L|Y|}] \tag{120}$$

so because Y and X have same distribution we have

$$E[e^{f(X)-E[f(X)]}]\leq E[e^{\frac{\pi}{2}L|Y|}]=E[e^{\frac{\pi}{2}L|X|}] \tag{121}$$

so we can see that it is $\frac{\pi L}{2}$-subg ✓