ORFE 525: Statistical Learning and Nonparametric Estimation
Homework 6

Zachary Hervieux-Moore

Friday 12^{th} May, 2017

Exercise 1:

1) Determine the VC dimension of the class of all polygons with k vertices in the two-dimensional plane.

2) Consider the function class

$$\mathcal{F} := \{\{x : \sin(\theta x) \geq 0\} : \theta \in \mathbb{R}\}$$

Show that the VC-dimension of $\mathcal{F}$, denoted as $d_{VC}(\mathcal{F})$, is infinite.

3) Let $\mathcal{C}$ be a class of convex subsets of $\mathbb{R}^d$. Show that if $I \subset \mathbb{R}^d$ is shattered by $\mathbb{R}^d$, then every $x \in I$ must be an extreme point of the convex hull of I (denoted as conv I).

4) Let $\mathcal{C}$ be of finite VC dimension $d \geq 1$. Consider the class $\mathcal{C}_k = \{\cup_{i=1}^k c_i : c_i \in \mathcal{C}\}$ for some $k \geq 1$.

   a) Let $\mathbb{X}_{1:n}$ be a sample of size n and let $\Pi_K(n)$ denote the set of all labelings of $\mathbb{X}_{1:n}$ using class K. Show that $|\Pi_{\mathcal{C}_k}(n)| \leq |\Pi_{\mathcal{C}}(n)|^k$.

   b) Prove that $d_{VC}(\mathcal{C}_k) = O(dk \log k)$. (Hint: you can use the fact that $\log_2(3k) < 9k/(2e)$ for $k \geq 2$).

Answer:

1) We first show that it is possible to shatter $2k + 1$ points. Suppose the points are equally spaced around the unit sphere. If the number of positive points (the ones labeled $+1$) is less than or equal to k, then simply draw a k-polygon that connects the k points inside the circle. This clearly does not contain the -1 labeled points as the vertices of the polygon lie on the other points and the edges are contained in the circle. Now suppose the number of $+1$ labeled points is strictly larger than k. That means the number of -1 labeled points is less than or equal to k. Then, we draw a k-polygon such that the circle is inscribed and the sides are tangent to the circle at the negative points. Thus, all $+1$ labeled points are inside the polygon and the negative points are on the edge. We can then just shrink the polygon by a little such that the negative points are just outside but the positive points stay inside the polygon. This shows that we can shatter at least $2k + 1$ points.

Now, suppose there are $2k+2$ points. First, consider them scattered in the plane. In this orientation, consider the polar representation, by labeling the points with alternating +1 and -1 by according to their angle, one sees that the problem is the most difficult when the points are all radially the same distance away from the origin (or if a shattering exists in the scattered version, a shattering exists in the circular version). Thus, we can only consider the case when the points are on a circle. We can normalize the circle and so we only have to prove this for the case of $2k+2$ points on the unit circle. Now, label the points alternating from +1 and -1 again. In this configuration, any solution that shatters the points will admit a convex solution. This is because the points lie on a circle and so non-convexity does not help with "getting around" the negative points. Now that we are can only consider the convex k-polygons, it is easy to see that it is impossible to shatter the $k+1$ points in the alternating labeling because the negative points between every positive point forces you to to draw a line that connects the adjacent positive points while being below the negative point. Thus, one needs a $k+1$-polygon. Which proves the result.

2) Suppose the points $\{x_i\}_{i=1}^n$ are located at $x_i = 2^{-i}$ and have labels $\{y_i\}_{i=1}^n$. Then we will show that picking

$$\theta = \pi(1 + \sum_{i=1}^{n}(1-y_i)2^{i-1})$$

Classifies these points correctly. We have

$$\begin{aligned}
&\sin(x_j\pi(1 + \sum_{i=1}^{n}(1-y_i)2^{i-1})) \\
&= \sin(2^{-j}\pi(1 + \sum_{i=1}^{n}(1-y_i)2^{i-1})) \\
&= \sin(2^{-j}\pi + \pi\sum_{i=1}^{n}(1-y_i)2^{i-j-1}))
\end{aligned}$$

Now, if $y_i = 1$, then the term in the sum is 0 so we drop those. Also, if $i > j$, this will result in integral multiples of 2π which does nothing to sine ($\sin(x) = \sin(x + 2\pi)$). Thus, we get

$$= \sin(2^{-j}\pi + \pi \sum_{i:i \geq j, y_i = -1}^{n} (1 - y_i)2^{i-j-1}))$$

$$= \sin((1 - y_j)\frac{\pi}{2} + 2^{-j}\pi + \pi \sum_{i:i<j, y_i = -1}^{n} 2^{i-j}))$$

Now, it is an elementary real analysis fact that $\sum_{i=1}^{\infty} 2^{-i} = 1$ and so a subset is strictly less than one. Thus, we have

$$= \begin{cases} \sin(\pi c) & \text{if } y_i = 1 \\ \sin(\pi + \pi c) & \text{if } y_i = -1 \end{cases}$$

where $c \in [0, 1]$. Thus, if $y_i = 1$ then $\sin(\theta x_i) = 1$ and if $y_i = -1$ then $\sin(\theta x_i) = -1$.

3) By contradiction, suppose that there exists $x \in I$ that is not an extreme point. Then, it is a convex combination of two other points in I, say y and z. That is, $x = \lambda y + (1 - \lambda)z$. Thus, picking the labels

$$x = -1, y = 1, z = 1$$

then it is impossible to shatter this with a convex set because if y and z are in this set, then convexity enforces that x is also in this set. However, $x = -1$ and so this set doesn't classify these points. Thus, we must have x be an extreme point

4)

a) We have that every set of labeling $Y \in \Pi_{\mathcal{C}_k}(n)$ is of the form $Y = \cup_{i=1}^{k} Y_i$ with $Y_i \in \Pi_{\mathcal{C}}(n)$. Thus, every Y_i has $|\Pi_{\mathcal{C}}(n)|$ different labelings. We conclude that

$$|\Pi_{\mathcal{C}_k}(n)| \leq |\Pi_{\mathcal{C}}(n)|^k$$

b) By Sauer's lemma and the previous part

$$|\Pi_{\mathcal{C}_k}(n)| \leq \left(\frac{en}{d}\right)^{dk}$$

Now, by definition of VC, we need

$$\left(\frac{en}{d}\right)^{dk} \geq 2^n$$

Otherwise, $\mathbb{X}_{1:n}$ cannot be shattered by $\mathcal{C}_k$. Equivalently, if we find n such that,

$$\left(\frac{en}{d}\right)^{dk} < 2^n$$

Then $d_{VC}(\mathcal{C}_k) = O(n)$. Now, if $n = 2dk\log(3k)$, we have that

$$\left(\frac{e2dk\log(3k)}{d}\right)^{dk} < (3k)^{2dk}$$
$$\log(3k) < \frac{9k}{2e}$$

Which is true for $k \geq 2$ as per the hint. Thus, we conclude that $d_{VC}(\mathcal{C}_k) = O(2dk\log(3k)) = O(dk\log(k))$.

Exercise 2: We use the same notations in the lecture note. Let ϵ_i be a Rademacher sequence, $\mathcal{F} = \{f : \mathcal{X} \to [0,1]\}$, $X_1, \ldots, X_n$ be i.i.d. variables. Please modify the proofs in the lecture notes to prove the following problems.

1) Prove that there exists a universal constant C such that

$$\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right|\right] \leq \inf_{a>0}\left\{C\alpha + C\int_{\alpha}^{\infty}\sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon\right\}$$

Is this inequality better than the Dudley's inequality proved in this class? Why?

2) Let $\phi(\cdot)$ be a convex function. Prove that

$$\mathbb{E}\left[\phi\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}f(X_i) - \mathbb{E}[f(X_i)]\right|\right)\right] \leq \mathbb{E}\left[\phi\left(2\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right|\right)\right]$$

3) Use the chaining proof of the Dudley's inequality to prove that there exists a universal constant C such that

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i) \geq C\int_{0}^{\infty}\sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + x\right] \leq Ce^{-x^2/C}$$

Answer:

1) Let

$$\alpha_j = 2^{-j}\sup_{f\in\mathcal{F}}\sqrt{\mathbb{E}[f(X_i)^2]}$$

Now, for every j, create a α_j-net of $\mathcal{F}$ and denote it by T_j. Then create the chaining

$$f = f - \hat{f}_N + \sum_{i=1}^{N}(\hat{f}_i - \hat{f}_{i-1})$$

where $\hat{f}_0 = 0$ and $\hat{f}_i$ is in the α_j-net. Now we get for any N,

$$\begin{aligned}
&\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right|\right] \\
&= \frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\epsilon_i\left(f(X_i) - \hat{f}_N(X_i) + \sum_{j=1}^{N}(\hat{f}_j(X_i) - \hat{f}_{j-1}(X_i))\right)\right|\right] \\
&\leq \frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\epsilon_i\left(f(X_i) - \hat{f}_N(X_i)\right)\right|\right] \\
&\quad + \sum_{j=1}^{N}\frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\epsilon_i(\hat{f}_j(X_i) - \hat{f}_{j-1}(X_i))\right|\right] \\
&\leq \sup_{f\in\mathcal{F}}|f - \hat{f}_N| + \sum_{j=1}^{N}\frac{1}{n}\mathbb{E}\left[\sup_{f\in\mathcal{F}}\left|\sum_{i=1}^{n}\epsilon_i(\hat{f}_j(X_i) - \hat{f}_{j-1}(X_i))\right|\right]
\end{aligned}$$

Where the last line is due to repeatedly applying triangle inequality and the fact that X_i's are i.i.d. Note that $\sup_{f\in\mathcal{F}}|f - \hat{f}_N| \leq \alpha_N$ so we have

$$\leq \alpha_N + \sum_{j=1}^{N}\frac{1}{n}\mathbb{E}\left[\sup_{\hat{f}_j\in T_j, \hat{f}_{j-1}\in T_{j-1}}\left|\sum_{i=1}^{n}\epsilon_i(\hat{f}_j(X_i) - \hat{f}_{j-1}(X_i))\right|\right]$$

Now, we use finite class lemma to get

$$\leq \alpha_N + \sum_{j=1}^{N}\sqrt{\frac{2R^2\log(|T_j||T_{j-1}|)}{n}}$$

We have that $|T_j| \geq |T_{j-1}|$. For this class, we have that $\|\hat{f}_j - \hat{f}_{j-1}\|^2 \leq 3\alpha_j^2$ by triangle inequality and because $\alpha_{j-1} = 2\alpha_j$. Putting this together we get

$$\leq \alpha_N + 6\sum_{j=1}^{N} \alpha_j \sqrt{\frac{\log(|T_j|)}{n}}$$

$$\leq \alpha_N + 12\sum_{j=1}^{N} (\alpha_j - \alpha_{j+1}) \sqrt{\frac{\log N(\alpha_j, \mathcal{F}, d_n))}{n}}$$

$$\leq \alpha_N + 12\int_{\alpha_{N+1}}^{\alpha_0} \sqrt{\frac{\log N(\epsilon, \mathcal{F}, d_n))}{n}} d\epsilon$$

$$\leq \alpha_N + 12\int_{\alpha_{N+1}}^{\infty} \sqrt{\frac{\log N(\epsilon, \mathcal{F}, d_n))}{n}} d\epsilon$$

Now we pick α such that $\alpha_{N+1} < 2\alpha$ and this means that $\alpha_{N+1} > \alpha$. Then we get

$$\leq 4\epsilon + 12\int_{\epsilon}^{\infty} \sqrt{\frac{\log N(\alpha_j, \mathcal{F}, d_n))}{n}} d\epsilon$$

Now, we pick the smallest such ϵ as it was arbitrary

$$\leq \inf_{\alpha > 0} \left\{ 4\alpha + 12\int_{\alpha}^{\infty} \sqrt{\frac{\log N(\epsilon, \mathcal{F}, d_n))}{n}} d\epsilon \right\}$$

Thus, the universal constant here is $C = 12$.

2) We duplicate the steps done in class. We also make the additional assumption that $\phi(\cdot)$ is increasing. Applying the Tower property and Jensen's, we get

$$\mathbb{E}\left[\phi\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[f(X_i)]\right|\right)\right]$$

$$= \mathbb{E}\left[\phi\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - \mathbb{E}[\sum_{i=1}^{n} f(X_i')]\right|\right)\right]$$

$$\leq \mathbb{E}\left[\phi\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n} f(X_i) - f(X_i')\right|\right)\right]$$

Where the ϵ_i are symmetric Bernoulli's. Now, we know that $\epsilon_i(f(X_i) - f(X_i')) \stackrel{d}{=} f(X_i) - f(X_i')$ as ϵ_i is symmetric. So

$$\leq \mathbb{E}\left[\phi\left(\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i) - \frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i')\right|\right)\right]$$

$$\leq \mathbb{E}\left[\phi\left(\frac{2}{2}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right| + \frac{2}{2}\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i')\right|\right)\right]$$

Where we used the triangle inequality. Now we apply convexity to this convex combintation and use the fact that $X \stackrel{d}{=} X$ to conclude.

$$\leq \mathbb{E}\left[\frac{1}{2}\phi\left(2\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right|\right) + \frac{1}{2}\phi\left(2\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i')\right|\right)\right]$$

$$\leq \mathbb{E}\left[\phi\left(2\sup_{f\in\mathcal{F}}\left|\frac{1}{n}\sum_{i=1}^{n}\epsilon_i f(X_i)\right|\right)\right]$$

3) Consider the decomposition

$$f = \sum_{j=1}^{\infty}(\hat{f}_j - \hat{f}_{j-1})$$

Now, we bound

$$P = \mathbb{P}\left[\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i(\hat{f}_j - \hat{f}_{j-1})(X_i) \geq C\int_{\alpha_{j+1}}^{\alpha_j}\sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + \frac{x}{2}\frac{2}{3}^j\right]$$

To achieve this bound, we compute

$$\star = \frac{\mathbb{E}[e^{\lambda\sup_{f\in\mathcal{F}}\frac{1}{n}\sum_{i=1}^{n}\epsilon_i(\hat{f}_j-\hat{f}_{j-1})(X_i)}]}{e^{\lambda(C\int_{\alpha_{j+1}}^{\alpha_j}\sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon+\frac{x}{2}\frac{2}{3}^j)}}$$

Bounding the numerator, we get

$$\begin{aligned}
&\mathbb{E}[e^{\lambda \sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \epsilon_i(\hat{f}_j-\hat{f}_{j-1})(X_i)}] \\
&= \mathbb{E}[\sup_{f\in\mathcal{F}} e^{\lambda\frac{1}{n}\sum_{i=1}^n \epsilon_i(\hat{f}_j-\hat{f}_{j-1})(X_i)}] \\
&\le \sum_{\hat{f}_j\in T_j, \hat{f_{j-1}}\in T_{j-1}} \mathbb{E}[e^{\lambda\frac{1}{n}\sum_{i=1}^n \epsilon_i(\hat{f}_j-\hat{f}_{j-1})(X_i)}] \\
&\le \sum_{\hat{f}_j\in T_j, \hat{f_{j-1}}\in T_{j-1}} e^{\lambda^2\frac{3}{2}\frac{\alpha_j^2}{n}}
\end{aligned}$$

Where we used the fact that these are sub-Gaussian variables and we showed that $\|\hat{f}_j - \hat{f}_{j-1}\|_2^2 \le 3\alpha_j^2$. Now, let $N(\alpha_j, \mathcal{F}, d_n)$ be the number of $\hat{f}_j$. We get that

$$\star \le e^{2\log(N(\alpha_j,\mathcal{F},d_n)) - \lambda\frac{1}{2}C\alpha_j\sqrt{\frac{\log N(\alpha_j,\mathcal{F},d_n)}{n}}d\epsilon + \lambda^2\frac{3}{2}\frac{\alpha_j^2}{n} - \lambda\frac{x}{2}\frac{2}{3}^j)}$$

As this is true for all $\lambda > 0$, we take the minimum

$$\lambda = \frac{\frac{1}{2}C\alpha_j\sqrt{\frac{\log N(\alpha_j,\mathcal{F},d_n)}{n}} + \frac{x}{2}\frac{2}{3}^j)}{\frac{3}{2}\frac{\alpha_j^2}{n}}$$

Which means that

$$\star \le e^{-\frac{1}{24}(\frac{4}{3})^j x^2 n}$$

Now, as in part 1, take $C \ge 12$ and we get

$$P \le \star \le e^{-\frac{1}{24}(\frac{4}{3})^j x^2 n}$$

We wish to apply the union bound on our chaining. We note that

$$C\int_0^\infty \sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + x \ge \sum_{j=1}^\infty C\int_{\alpha_{j+1}}^{\alpha_j} \sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + (2/3)^j\frac{x}{2}$$

Now, we can apply the union bound

$$\star\star = \mathbb{P}(\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \geq C\int_0^\infty \sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + x)$$

$$\leq \sum_{j=1}^n \mathbb{P}(\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \geq C\int_{\alpha_{j+1}}^{\alpha_j} \sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + (2/3)^j\frac{x}{2})$$

$$\leq \sum_{j=1}^n e^{-\frac{1}{24}(4/3)^j x^2 n}$$

Taking C_n big enough, we have

$$-\frac{1}{24}(4/3)^j x^2 n \leq -\frac{x^2}{C_n}j$$

Thus, we have

$$\star\star \leq e^{-\frac{x^2}{C_n}}\frac{1}{1-e^{-\frac{x^2}{C_n}}}$$

Now, taking C_n to be sufficiently large, we get that

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \geq C_n\int_0^\infty \sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + x\right] \leq C_n e^{-x^2/C_n}$$

To remove the dependence on n, we note that if $x' = \frac{x}{\sqrt{n}}$, then we have

$$-\frac{1}{24}(4/3)^j x'^2 \leq -\frac{x'^2}{C}j$$

Thus, we conclude that there exists a universal C

$$\mathbb{P}\left[\sup_{f\in\mathcal{F}} \frac{1}{n}\sum_{i=1}^n \epsilon_i f(X_i) \geq C\int_0^\infty \sqrt{\frac{\log N(\epsilon,\mathcal{F},d_n)}{n}}d\epsilon + x'\right] \leq Ce^{-x^2/C}$$

Exercise 3: 1) In the class, we know that if $X_1, \ldots, X_n$ are independent subgaussian variables with X_i satisfying $\mathbb{E}[X_i] = 0$ and variance-proxy σ_i^2 for any $1 \le i \le n$, then we have

$$\mathbb{P}\left[\sum_{i=1}^n X_i \ge t \left\{\sum_{i=1}^n \sigma_i^2\right\}^{1/2}\right] \le e^{-t^2/2} \text{ for all } t \ge 0$$

However, for general random variables, there is no hope to obtain such an inequality. Indeed, if the variables X_i have heavy tails, for example, then clearly the sum cannot have a Gaussian tail for large t.

Remarkably, there is a method to obtain Gaussian inequalities of this type that works without any tail assumption on the random variables! The key idea is to choose a random normalization that plays the role of the sum of the variaces in the Gaussian case. We then say the sum is self-normalized.

Consider independent random variables $X_1, \ldots, X_n$ satisfying $\mathbb{E}[X_i] = 0$ and have symmetric distributions, i.e., X_i has the same distribution as $-X_i$ for any $1 \le i \le n$. Show that

$$\mathbb{P}\left[\sum_{i=1}^n X_i \ge t \left\{\sum_{i=1}^n X_i^2\right\}^{1/2}\right] \le e^{-t^2/2} \text{ for all } t \ge 0$$

Hint: Apply symmetrization and Hoeffding inequality conditionally.

2) Let $X_1, \ldots, X_n$ be i.i.d. random variables with values in $[0, 1]$. Each X_i represents the size of a package to be shipped. The shipping containers are bins of size 1 (so each bin can hold a set of packages whose sizes sum to at most 1). Let $B_n = f(X_1, \ldots, X_n)$ be the minimal number of bins needed to store the packages. Note that computing B_n os a hard combinatorial optimzation problem, but we can bound its probabilistic fluctuations by easy arguments.

   a) Show that $\mathbb{E}[B_n] \ge n\mathbb{E}[X_1]$
   b) Prove that $\mathbb{P}[|B_n - \mathbb{E}[B_n]| \ge t] \le 2e^{-2t^2/n}$

3) Let $X_1, X_2, \ldots, X_n$ be n points picked independently and uniformly at random from the unit square $[0, 1]^2$. Let $\tau : ([0, 1]^2)^n \to \mathbb{R}$ be the length

of the shortest traveling salesman tour on n points. Prove that

$$\mathbb{P}[|\tau - \mathbb{E}[\tau]| \geq t] \leq 2\exp(-t^2/16n)$$

Note: a traveling salesman tour is to find a shortest tour which visits each point exactly once. (Hint: use McDiarmaid's inequality).

Answer:

1) We condition X_i on X_i^2. That is, since they are symmetric,

$$\mathbb{P}(X_i|X_i^2) = \begin{cases} \sqrt{X_i^2} & \text{w.p. } 1/2 \\ -\sqrt{X_i^2} & \text{w.p. } 1/2 \end{cases}$$

Now, we apply Hoeffding to the conditional distribution

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\left(\sum_{i=1}^{n} X_i^2\right)^{1/2} |X_1^2, \ldots, X_n^2\right) =\leq e^{-t^2/2}$$

Now, take the expectation of the above and apply the tower property to conclude that

$$\mathbb{P}\left(\sum_{i=1}^{n} X_i \geq t\left(\sum_{i=1}^{n} X_i^2\right)^{1/2}\right) =\leq e^{-t^2/2}$$

2) a) We have that the numbers of bins required is at the least the size of all the packages combined. Thus, as the sizes are i.i.d.,

$$\mathbb{E}[B_n] \geq \mathbb{E}[\sum_{i=1}^{n} X_i] = \sum_{i=1}^{n} \mathbb{E}[X_i] = n\mathbb{E}[X_1]$$

b) We first note that we have

$$\sup_{x_i'} |B_n(x) - B_n(x')| \leq 1 \forall i$$

That is, the greatest difference in the number of bins by changing a single X_i is 1 as we might need to add one or remove one. Then, McDiarmaid's inequality applies and we get

$$\mathbb{P}(|B_n - \mathbb{E}[B_n]| \geq t) \leq 2e^{-2t^2/\sum_{i=1}^{n} c_i^2} = 2e^{-2t^2/n}$$

3) Again, consider the case when all the points are in one corner. Then, we move one of the points to the opposite corner. This is the largest change in the tour we can create by moving a single point. That is

$$\sup_{x_i'} |\tau(x) - \tau(x')| \leq 2\sqrt{2} \ \forall i$$

Where we must multiply by 2 since it is a tour (must return back to the original point). Then, by McDiarmaid's inequality,

$$\mathbb{P}(|\tau - \mathbb{E}[\tau]| \geq t) \leq 2e^{-t^2/2\sum_{i=1}^{n} c_i^2} = 2e^{-t^2/16n}$$