3. Concentration of Measures

Danilo Jr Dela Cruz

Questions set by: Professor Xue-Mei-Li

If A is a set in $\mathbb{R}^n$ we denote by $|A|$ or $\mathrm{Vol}(A)$ its volume.

$$\mathrm{Vol}(A) := \int_{-\infty}^{\infty} \cdots \int_{-\infty}^{\infty} 1_A(x)dx$$

The (Lebesgue) volume of the cube $[-1,1]^n$ is 2^n, its normalised volume is 1. A subset of it has volume

$$\mathrm{Vol}(A \cap [-1,1]^n) = \int_{-1}^{1} \cdots \int_{-1}^{1} 1_A(x)dx$$

If $x = (x_1, \ldots, x_n) \in \mathbb{R}^n$, its norm is

$$|x| = \sqrt{\sum_{i=1}^{n} (x_i)^2}$$

The ball centered at 0 with radius r is denoted by $B(r)$, its boundary by $\partial B(r)$:

$$B(r) := \{x : |x| \le r\}, \quad \partial B(r) = \{x : |x| = r\}$$

Theorem 0.1. Most of the volume of the cube $[-1,1]^n$ is concentrated withing the boundary of the ball $B(\sqrt{\frac{n}{3}})$.In other words show that for any $\varepsilon > 0$ that the normalised volume of the set of points from the cube, with norm between $(1-\varepsilon)\sqrt{n/3}$ and $(1+\varepsilon)\sqrt{n/3}$ tends to 1:

$$\frac{1}{2^n}\mathrm{Vol}\left(\left\{x : (1-\varepsilon)\sqrt{n/3} < |x| < (1+\varepsilon)\sqrt{n/3}\right\} \cap [-1,1]^n\right) \to 1$$

1. (The right terminology for the functions I use below is ‘Borel measurable’ which we do not learn until year 2 (Analysis) / year 3 (Measure Theory). Both indicator functions of intervals and continuous functions are Borel measurable. So we try work with the last two classes of functions.)

   A random variable $(X_1, \ldots, X_n)$ on $[-1,1]^n$ is uniformly distributed if for $f : \mathbb{R}^n \to \mathbb{R}$ ‘Borel measurable function’,

$$\mathbb{E}(f(X_1, \ldots, X_n)) = \frac{1}{2^n} \int_{-1}^{1} \cdots \int_{-1}^{1} f(x_1, \ldots, x_n)dx_1 \ldots dx_n$$

If f takes on a product form, say $f(x_1,\dots,x_n) = f_1(x_1)\dots f_n(x_n)$ then this means

$$\mathbb{E}(f(X_1,\dots,X_n)) = \frac{1}{2^n}\int_{-1}^{1}\cdots\int_{-1}^{1} f_1(x_1)\dots f_n(x_n)dx_1\dots dx_n$$

Show that if $(X_1,\dots,X_n)$ are distributed uniformly over $[-1,1]^n$, then $\{X_i\}$ are iid on $[-1,1]$.

First we need to state some definitions and lemmas that will prove useful in proving the statement above.

Definition: A set of random variables $\{X_i\}$ are said to be iid if they have the same distribution and are mutually independent.

Definition: Random variables $X : \Omega_x \to \mathbb{R}, Y : \Omega_y \to \mathbb{R}$ are said to have the same distribution if

$$\forall x \in \mathbb{R},\ F_X(x) := \mathbb{P}(X \le x) = \mathbb{P}(Y \le x) =: F_Y(x)$$

Definition: A collection of random variables $\{X_i\}$ are said to be mutually independent if their joint CDF takes the following product form

$$\forall x_1,\dots,x_n \in \mathbb{R},\ F_{X_1,\dots,X_n}(x_1,\dots,x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_1)$$

Lemma: The expectation of the indicator function of a subset A of $\mathbb{R}$ with respect to a random variable X gives us the probability that $X \in A$.

$$\mathbb{E}[1_A(X)] := \mathbb{P}(X \in A)$$

Lemma: Suppose that we have a collection of sets $\{S_i\}$ and an indicator function given by

$$1_{S_1\times\cdots\times S_n}(x_1,\dots x_n) = \begin{cases} 1 & \forall i, x_i \in S_i \\ 0 & \text{otherwise} \end{cases}$$

It is clear to see that our indicator function takes the following product form because the RHS is 1 if and only if $\forall i, 1_{S_i}(x_i) = 1 \Rightarrow x_i \in S_i$

$$1_{S_1\times\cdots\times S_n}(x_1,\dots x_n) = 1_{S_1}(x_i)\cdots 1_{S_n}(x_n),\quad 1_{S_i}(x_i) = \begin{cases} 1 & x_i \in S_i \\ 0 & \text{otherwise} \end{cases}$$

Lemma: By integrating from the innermost layer and observing that it is a constant with respect to the other variables then factorising we deduce

$$\int \cdots \int f_1(x_1) \cdots f_n(x_n) d_{x_1} \ldots d_{x_n} = \left(\int f(x_1) d_{x_1} \right) \cdots \left(\int f(x_n) d_{x_n} \right)$$

Proof: To begin the proof we will try to show that $\{X_i\}$ have the same distribution so let us fix i and try to find the distribution of X_i

$$F_{X_i}(x_i) := P(X_i \leq x_i) = E(1_{(-\infty, x_i]}(X_i))$$

By lemma the indicator variable below takes on the product form

$$1_{\mathbb{R} \times \cdots \times (-\infty, x_i] \times \cdots \mathbb{R}}(X_1, \ldots, X_n) = 1_{\mathbb{R}}(X_1) \cdots 1_{(-\infty, x_i]}(X_i) \cdots 1_{\mathbb{R}}(X_n)$$

By definition of random variables we know that $X_j : \Omega_j \to \mathbb{R} \Rightarrow \forall \omega \in \Omega,\ X_j(\omega) \in \mathbb{R} \Rightarrow (j \neq i \Rightarrow 1_{\mathbb{R}}(X_j) = 1)$. Hence we deduce that:

$$1_{\mathbb{R} \times \cdots \times (-\infty, x_i] \times \cdots \mathbb{R}}(X_1, \ldots, X_n) = 1_{(-\infty, x_i]}(X_i)$$

Therefore as indicator functions are Borel Measurable and $(X_1, \cdots, X_n)$ is uniformly distributed on $[-1, 1]^n$ it follows that

$$\begin{aligned}
&E(1_{(-\infty, x_i]}(X_i)) = E(1_{\mathbb{R} \times \cdots \times (-\infty, x_i] \times \cdots \mathbb{R}}(X_1, \ldots, X_n)) \\
&= \frac{1}{2^n} \int_{-1}^{1} \cdots \int_{-1}^{1} 1_{\mathbb{R} \times \cdots \times (-\infty, x_i] \times \cdots \mathbb{R}}(t_1, \ldots, t_n) dt_1 \ldots dt_n \\
&= \frac{1}{2^n} \int_{-1}^{1} \cdots \int_{-1}^{1} 1_{(-\infty, x_i]}(t_i) dt_1 \ldots dt_n \\
&= \frac{1}{2^n} \left(\int_{-1}^{1} 1 d_{t_1} \right) \cdots \left(\int_{-1}^{1} 1_{(-\infty, x_i]}(t_i) dt_i \right) \cdots \left(\int_{-1}^{1} 1 \, d_{t_n} \right) \\
&= \frac{1}{2} \left(\int_{-1}^{1} 1_{(-\infty, x_i]}(t_i) dt_i \right)
\end{aligned}$$

Therefore we deduce that all $\forall i, X_i \sim \mathcal{U}(-1, 1)$

$$F_{X_i}(x_i) = \begin{cases} 0, & x_i \leq -1 \\ \frac{1+x_i}{2}, & -1 < x_i < 1 \\ 1, & x_i \geq 1 \end{cases}$$

To demonstrate mutual independence, we use the lemma on product forms of indicator variables to deduce

$$\begin{aligned}
&\mathbb{P}(X_1 \leq x_1, \ldots, X_n \leq x_n) = \mathbb{E}(1_{(-\infty,x_1]\times\cdots\times(-\infty,x_n]}(X_1,\ldots,X_n)) \\
&= \frac{1}{2^n}\int_{-1}^{1}\cdots\int_{-1}^{1} 1_{(-\infty,x_1]\times\cdots\times(-\infty,x_n]}(t_1,\ldots,t_n)dt_1\ldots dt_n \\
&= \frac{1}{2^n}\int_{-1}^{1}\cdots\int_{-1}^{1} 1_{(-\infty,x_1])}(t_1)\cdots 1_{(-\infty,x_n])}(t_n)dt_1\ldots dt_n \\
&= \left(\frac{1}{2}\int_{-1}^{1} 1_{(-\infty,x_1]}(t_1)dt_1\right)\cdots\left(\frac{1}{2}\int_{-1}^{1} 1_{(-\infty,x_n]}(t_n)dt_n\right) \\
&= \mathbb{P}(X_1 \leq x_1)\cdots\mathbb{P}(X_n \leq x_n)
\end{aligned}$$

As $x_1, \ldots, x_n$ were arbitrary we have shown that

$$\forall x_1, \ldots, x_n \in \mathbb{R},\ F_{X_1,\ldots,X_n}(x_1,\ldots,x_n) = F_{X_1}(x_1)\cdots F_{X_n}(x_1)$$

Therefore $\{X_i\}$ are iid uniform distributions on $[-1,1]$.

2. Write

$$U_n := \{x : (1-\varepsilon)\sqrt{n/3} < |x| < (1+\varepsilon)\sqrt{n/3}\}$$
$$\mathrm{Vol}(U_n \cap [-1,1]^n) = \int_{-1}^{1}\cdots\int_{-1}^{1} 1_{U_n}(x_1,\ldots,x_n)dx_1\ldots dx_n$$

Explain how does $\mathrm{Vol}(U_n \cap [-1,1]^n)$ compare with the following probability

$$\mathbb{P}\left(\left\{\omega : (1-\varepsilon)\sqrt{n/3} < \sqrt{X_1(\omega)^2 + \ldots X_n(\omega)^2} < (1+\varepsilon)\sqrt{n/3}\right\}\right)$$

Intuitively, the probability is equal to $\mathrm{Vol}(U_n \cap [-1,1]^n)/\mathrm{Vol}([-1,1]^n)$ as it represents the proportion that falls inside the region U_n.

But this follows when we consider the expectation of an indicator variable

$$\begin{aligned}
&\mathbb{P}\left(\left\{\omega : (1-\varepsilon)\sqrt{n/3} < \sqrt{X_1(\omega)^2 + \ldots X_n(\omega)^2} < (1+\varepsilon)\sqrt{n/3}\right\}\right) \\
&= E(1_{U_n}(X_1,\ldots,X_n)) = \frac{1}{2^n}\int_{-1}^{1}\cdots\int_{-1}^{1} 1_{U_n}(x_1,\ldots,x_n)dx_1\ldots dx_n \\
&= \mathrm{Vol}(U_n \cap [-1,1]^n)/2^n = \mathrm{Vol}(U_n \cap [-1,1]^n)/\mathrm{Vol}([-1,1]^n)
\end{aligned}$$

3. Define $Y_i = (X_i)^2$. State a law of large numbers for Y_i.

The Weak Law of Large Numbers states that given a sequence of iid random variables Y_i which have a finite expectation μ then:

$$\forall \varepsilon > 0, \lim_{n\to\infty} \mathbb{P}(|\overline{Y}_n - \mu| < \varepsilon) = 1$$

4. Prove the statement.

   First we rearrange the following expression for

$$(1-\varepsilon)\sqrt{\frac{n}{3}} < \sqrt{X_1(\omega)^2 + \ldots X_n(\omega)^2} < (1+\varepsilon)\sqrt{\frac{n}{3}} \Leftrightarrow \left|\sqrt{\frac{3\overline{Y}_n}{n}} - 1\right| < \varepsilon$$

   Observe that we can bound the following by

$$\left|\sqrt{\frac{3\overline{Y}_n}{n}} - 1\right| = \left|\frac{\frac{3\overline{Y}_n}{n} - 1}{\sqrt{\frac{3\overline{Y}_n}{n}} + 1}\right| < \left|\frac{3\overline{Y}_n}{n} - 1\right| < \varepsilon$$

$$\mathbb{P}\left(\left|\frac{3\overline{Y}_n}{n} - 1\right| < \varepsilon\right) \leq \mathbb{P}\left(\left|\sqrt{\frac{3\overline{Y}_n}{n}} - 1\right| < \varepsilon\right) \leq 1$$

   Assuming that X_i is uniformly distributed on $[-1,1]$ we deduce:

$$\mathbb{E}(X_i^2) = \int_{-1}^{1} x^2 \cdot \frac{1}{2}\, dx = \frac{1}{3} \Rightarrow E\left(\frac{\overline{Y}_n}{n}\right) = \frac{1}{3}$$

   Using the weak law of large numbers and squeeze we deduce that $\forall \varepsilon > 0$

$$\begin{aligned}\mathbb{P}\left(\left|\sqrt{\frac{3\overline{Y}_n}{n}} - 1\right| < \varepsilon\right) &= \mathbb{P}\left(\left\{\omega : (1-\varepsilon)\sqrt{n/3} < \sqrt{X_1(\omega)^2 + \ldots X_n(\omega)^2} < (1+\varepsilon)\sqrt{n/3}\right\}\right) \\ &= \frac{\mathrm{Vol}(U_n \cap [-1,1]^n)}{[-1,1]^{\mathrm{n}}} = \frac{\mathrm{Vol}(U_n \cap [-1,1]^n)}{2^n} \to 1\end{aligned}$$

5. Try to give an heuristic explanation for this concentration phenomenon. Note that the furthest away points in $[-1,1]^n$ has distance

$$\left\| \begin{vmatrix} 1 \\ \vdots \\ 1 \end{vmatrix} - \begin{vmatrix} -1 \\ \vdots \\ -1 \end{vmatrix} \right\| = 2\sqrt{n}$$

   It's quite hard to think and make sense of higher dimensions, but I think the motivation behind the current proof is quite intuitive already. The more dimensions there are the more components you have. To be far away from the origin you need to have all dimensions large which is unlikely and to get close you need to have all dimensions small which again is unlikely. Also the law of large number already tells us that such behaviour for lots of iid observations has measure zero and so we almost expect everything to cluster around the average. As we are measure radially and by the symmetry of the set up it makes sense that the clustering phenomenon manifests at the boundary of a ball centered at the origin.