

0.1 ☕☕ (Two variance formulas)

- (a) Recall that the variance of a random variable  $X$  satisfies  $\text{Var}(X) = \mathbb{E}(X - \mathbb{E}X)^2 = \mathbb{E}X^2 - (\mathbb{E}X)^2$ . Let us prove a higher-dimensional version of this identity. Check that any random vector  $Z$  in  $\mathbb{R}^n$  satisfies

$$\mathbb{E}\|Z - \mathbb{E}Z\|_2^2 = \mathbb{E}\|Z\|_2^2 - \|\mathbb{E}Z\|_2^2.$$

- (b) Let  $Z$  be a random vector in  $\mathbb{R}^n$ , and  $Z'$  be an independent copy of  $Z$ , i.e. a random vector independent of  $Z$  and with the same distribution as  $Z$ . Check that

$$\mathbb{E}\|Z - \mathbb{E}Z\|_2^2 = \frac{1}{2} \mathbb{E}\|Z - Z'\|_2^2.$$

$$\begin{aligned} \text{(a)} \quad \mathbb{E}\|Z - \mathbb{E}Z\|_2^2 &= \mathbb{E} \sum_{i=1}^n (z_i - \mathbb{E}z_i)^2 \\ &= \mathbb{E} \sum_{i=1}^n (z_i^2 - 2z_i \mathbb{E}z_i + \mathbb{E}^2 z_i) \\ &= \mathbb{E} \sum_{i=1}^n z_i^2 - \mathbb{E} \sum_{i=1}^n 2z_i \mathbb{E}z_i + \mathbb{E} \sum_{i=1}^n \mathbb{E}^2 z_i \\ &= \mathbb{E}\|Z\|_2^2 - 2 \sum_{i=1}^n \mathbb{E}^2 z_i + \mathbb{E} \sum_{i=1}^n \mathbb{E}^2 z_i \\ &= \mathbb{E}\|Z\|_2^2 - \|\mathbb{E}Z\|_2^2 \end{aligned}$$

$$\begin{aligned} \text{(b)} \quad \mathbb{E}\|Z - Z'\|_2^2 &= \mathbb{E} \sum_{i=1}^n (z_i - z'_i)^2 \\ &= \mathbb{E} \sum_{i=1}^n [z_i^2 - 2z_i z'_i + (z'_i)^2] \\ &= \mathbb{E} \sum_{i=1}^n z_i^2 - 2 \sum_{i=1}^n \mathbb{E}(z_i z'_i) + \sum_{i=1}^n \mathbb{E}(z'_i)^2 \\ &= 2\mathbb{E}\|Z\|_2^2 - 2 \sum_{i=1}^n \mathbb{E}^2 z_i \\ &= 2\mathbb{E}\|Z\|_2^2 - 2\|\mathbb{E}Z\|_2^2, \end{aligned}$$

WHICH, BY PART (a), COMPLETES THE PROOF.

0.2 ☕☕☕ (Expectation minimizes the mean squared error) The variance of a random variable  $X$  has the following extremal property:

$$\text{Var}(X) = \mathbb{E} \min_{a \in \mathbb{R}} (X - a)^2.$$

Let us prove a more general, high-dimensional version of this fact. Check that a random vector  $Z$  in  $\mathbb{R}^n$  with finite  $\mathbb{E} \|Z\|_2^2$  satisfies

$$\mathbb{E} \|Z - \mathbb{E} Z\|_2^2 = \min_{a \in \mathbb{R}^n} \mathbb{E} \|Z - a\|_2^2.$$

TO REFRESH, LET'S PROVE FIRST THAT  $\text{V}(X) = \mathbb{E} \min_{a \in \mathbb{R}} (X - a)^2$ :

$$\begin{aligned} \min_{a \in \mathbb{R}} \mathbb{E} (X - a)^2 &= \min_{a \in \mathbb{R}} \mathbb{E} (X - \mathbb{E} X + \mathbb{E} X - a)^2 \\ &= \min_{a \in \mathbb{R}} \mathbb{E} (X - \mathbb{E} X)^2 - 2 \mathbb{E} (X - \mathbb{E} X)(\mathbb{E} X - a) + \mathbb{E} (\mathbb{E} X - a)^2 \\ &= \min_{a \in \mathbb{R}} \mathbb{E} (X - \mathbb{E} X)^2 + (\mathbb{E} X - a)^2, \\ &= \mathbb{E} (X - \mathbb{E} X)^2, \end{aligned}$$

WHICH IS ACHIEVED WHEN  $a = \mathbb{E} X$ . THEN

$$\begin{aligned} \min_{a \in \mathbb{R}^n} \mathbb{E} \|Z - a\|_2^2 &= \min_{a \in \mathbb{R}^n} \mathbb{E} \|Z - \mathbb{E} Z + \mathbb{E} Z - a\|_2^2 \\ &= \min_{a \in \mathbb{R}^n} \mathbb{E} \sum_{i=1}^n (z_i - \mathbb{E} z_i + \mathbb{E} z_i - a_i)^2 \quad \text{WHERE } a = (a_1, \dots, a_n) \\ &= \min_{a \in \mathbb{R}^n} \left[ \mathbb{E} \sum_{i=1}^n (z_i - \mathbb{E} z_i)^2 + 2 \mathbb{E} \sum_{i=1}^n (z_i - \mathbb{E} z_i)(\mathbb{E} z_i - a_i) + \mathbb{E} \sum_{i=1}^n (\mathbb{E} z_i - a_i)^2 \right] \\ &= \min_{a \in \mathbb{R}^n} \|\mathbb{E}(Z - \mathbb{E} Z)\|_2^2 + \|\mathbb{E} Z - a\|_2^2 \\ &= \|\mathbb{E}(Z - \mathbb{E} Z)\|_2^2. \end{aligned}$$

0.3 ☕☕ (Variance of a sum) Recall that the variance of a sum of independent random variables equals the sum of variances. Let us prove a higher-dimensional version of this identity. Check that any independent mean-zero random vectors  $Z_1, \dots, Z_k$  in  $\mathbb{R}^n$  satisfy

$$\mathbb{E} \left\| \sum_{j=1}^k Z_j \right\|_2^2 = \sum_{j=1}^k \mathbb{E} \|Z_j\|_2^2.$$

$$\mathbb{E} \left\| \sum_{j=1}^k Z_j \right\|_2^2 = \mathbb{E} \sum_{i=1}^n \left( \sum_{j=1}^k z_{ji} \right)^2 \quad \text{WHERE } X_j = (X_{j1}, \dots, X_{jn})$$

$$\begin{aligned}
&= \mathbb{E} \sum_{i=1}^n \left( z_{1i} + \dots + z_{ki} \right)^2 \\
&= \mathbb{E} \sum_{i=1}^n \sum_{j=1}^k z_{ji} z_{ji} \\
&= \sum_{i=1}^n \sum_{j=1}^k \mathbb{E}(z_{ji} z_{ji}) \\
&= \sum_{i=1}^n \left( \sum_{j=1}^k \mathbb{E} z_{ji}^2 + \sum_{j \neq l} (\mathbb{E} z_{ji} z_{li}) (\mathbb{E} z_{li} z_{ji}) \right) \\
&= \sum_{i=1}^n \sum_{j=1}^k \mathbb{E} z_{ji}^2 \\
&= \sum_{j=1}^k \mathbb{E} \sum_{i=1}^n z_{ji}^2 \\
&= \sum_{j=1}^k \mathbb{E} \|z_j\|_2^2
\end{aligned}$$

BY INDEP.

BY THE MEAN-ZERO ASSUMPTION

0.4 ☹☹ (Balancing vectors) Let  $x_1, \dots, x_n$  be vectors in  $\mathbb{R}^n$  that lie within the unit Euclidean ball centered at the origin.

- Prove that it is possible to assign a sign  $\pm$  to each vector such that the sum  $\pm x_1 \pm x_2 \pm \dots \pm x_n$  lies within a Euclidean ball of radius  $\sqrt{n}$  centered at the origin.
- Explain why the value  $\sqrt{n}$  cannot be reduced in general.

a) LET

$$y_i = \begin{cases} 1 & \text{w.p. } 1/2 \\ -1 & \text{w.p. } 1/2 \end{cases} \quad i=1, \dots, n$$

BE IID SIGNED R.V.'S. SINCE  $\mathbb{E} \sum_{i=1}^n y_i x_i = 0$ ,

$$\mathbb{E} \left\| \sum_{i=1}^n y_i x_i \right\|_2^2 = \sum_{i=1}^n \mathbb{E} \|y_i x_i\|_2^2 = \sum_{i=1}^n x_i^2 = \sum_{i=1}^n \|x_i\|_2^2 \leq n.$$

THEREFORE, THERE EXISTS A CHOICE OF  $\{y_i\}_{i=1}^n$  S.T.  $\left\| \sum_{i=1}^n y_i x_i \right\|_2^2 \leq n$ . THIS, IN TURN, IMPLIES THAT  $\left\| \sum_{i=1}^n y_i x_i \right\|_2 \leq \sqrt{n}$ .

b) LET  $\|x_i\|_2 = 1$ ,  $x_1 = \dots = x_n$ , AND  $y_1 = \dots = y_n$ . THEN  $\left\| \sum_{i=1}^n y_i x_i \right\|_2 = \sqrt{n}$ , SO THAT THE UPPER BOUND IS ACHIEVED FOR SOME SELECTION OF  $y_1 x_1 + \dots + y_n x_n$ .

0.5 ☕☕☕ (Approximate Caratheodory is asymptotically tight) Demonstrate by example that the bound in Theorem 0.0.2 is almost tight. Specifically, for every  $n$  find a set  $T \subset \mathbb{R}^n$  such that for any convex combination  $\sum_{j=1}^k \lambda_j x_j$  of any  $k$  points  $x_1, \dots, x_k \in T$ , one has

$$\left\| x - \sum_{j=1}^k \lambda_j x_j \right\|_2 \geq \sqrt{\frac{1}{k} - \frac{1}{n}}.$$

Let  $n \rightarrow \infty$  while keeping  $k$  fixed to see that Theorem 0.0.2 is asymptotically tight in high dimensions.

LET  $T := \{e_1, \dots, e_n\}$  SO THAT  $\text{CONV}(T)$  IS AN  $(n-1)$ -DIMENSIONAL SIMPLEX, THE CENTER OF WHICH IS  $x = \frac{1}{n} \sum_{i=1}^n e_i$ . LET  $y$  BE THE CENTER OF A  $(k-1)$ -DIMENSIONAL FACE OF THE SIMPLEX, SAY  $T_k := \text{CONV}(\{e_1, \dots, e_k\})$ . THEN THE DISTANCE FROM  $x$  TO  $T_k$  IS

$$\begin{aligned} \|x - y\|_2 &= \left\| \frac{1}{k} \sum_{i=1}^k e_i - \frac{1}{n} \sum_{i=1}^n e_i \right\|_2 \\ &= \left\| \left( \frac{1}{k} - \frac{1}{n} \right) \sum_{i=1}^k e_i - \frac{1}{n} \sum_{i=k+1}^n e_i \right\|_2 \\ &= \sqrt{\frac{1}{k} - \frac{1}{n}}. \end{aligned}$$

THEN THE DISTANCE FROM  $x$  TO ANY POINT IN  $T_k$  (WHICH CAN BE EXPRESSED AS A CONVEX COMBINATION OF  $e_1, \dots, e_k$ ) IS AT DISTANCE BIGGER THAN  $\sqrt{\frac{1}{k} - \frac{1}{n}}$ , AS REQUIRED.

0.6 ☕☕ (Bounds on binomial coefficients) Prove the inequalities

$$\left(\frac{n}{k}\right)^k \leq \binom{n}{k} \leq \sum_{j=0}^k \binom{n}{j} \leq \left(\frac{en}{k}\right)^k$$

for any integers  $1 \leq k \leq n$ .

• LOWER BOUND

$$\frac{n}{k} \dots \frac{n}{k} \geq \frac{n}{k} \frac{n-1}{k-1} \dots \frac{n-k+1}{1} = \binom{n}{k}$$

THIS CAN ALSO BE OBTAINED BY USING THE RECIPROCAL AS INSTANCES OF SAMPLING  $k$  TIMES FROM A POPULATION OF SIZE  $n$  PARTITIONED IN TWO CLASSES, ONE OF WHICH HAS  $k$  ELEMENTS.

- THE SECOND INEQ. IS TRIVIAL

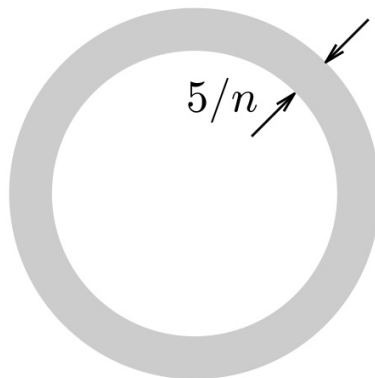
- UPPER BOUND:

$$\sum_{j=0}^k \binom{n}{j} \left(\frac{k}{n}\right)^j \leq \sum_{j=0}^k \binom{n}{j} \left(\frac{k}{n}\right)^j \leq \sum_{j=0}^n \binom{n}{j} \left(\frac{k}{n}\right)^j = \left(1 + \frac{k}{n}\right)^n \leq e^k$$

THEREFORE,

$$\sum_{j=0}^k \binom{n}{j} \leq \left(\frac{en}{k}\right)^k.$$

0.7 🙄🙄 (Thin shell phenomenon) Let us prove a counterintuitive fact that most of the volume of the high-dimensional ball lies near the surface. Consider the points inside the unit Euclidean ball of  $\mathbb{R}^n$  that lie within distance  $5/n$  from the surface of the ball, see Figure 0.3. Prove that such points make up over 99% of the volume of the ball.



**Figure 0.3** Over 99% of the volume of the unit ball in  $\mathbb{R}^n$  lies within distance  $5/n$  from the surface (Exercise 0.7)

LET  $V(B)$  BE THE VOLUME OF THE UNIT BALL IN  $\mathbb{R}^n$ . THE VOLUME OF THE BALL WITH RADIUS  $1 - 5/n$  IS  $(1 - 5/n)^n V(B)$ . THEN

$$\frac{(1 - 5/n)^n V(B)}{V(B)} \approx e^{-5} \approx 0,006.$$

THEN, ABOUT 0,4% OF THE VOLUME LIES WITHIN DISTANCE  $5/n$  OF THE SURFACE.

0.8 ☕☕☕ (Thin shell phenomenon, continued) Let  $X$  be a random vector that is uniformly distributed<sup>2</sup> in the Euclidean unit ball of  $\mathbb{R}^n$ . Prove that

$$\mathbb{E}\|X\|_2 = \frac{n}{n+1}.$$

LET  $R = \|X\|_2$ . THEN, USING THE SAME IDEA OF THE PREVIOUS EXERCISE,

$$F_R(r) = \frac{r^n V(B)}{V(B)} = r^n, \quad 0 \leq r \leq 1.$$

THUS,

$$f_R(r) = nr^{n-1}, \quad 0 \leq r \leq 1.$$

THEREFORE,

$$\mathbb{E}\|X\|_2 = \mathbb{E}R = \int_0^1 nr^n dr = n/(n+1).$$

0.9 ☕☕☕ (Carl-Pajor theorem) Let's improve Theorem 0.0.4 by replacing  $N$  with  $N/n$ . Let  $P$  be a polytope with  $N \geq n$  vertices, which is contained in the unit Euclidean ball of  $\mathbb{R}^n$ , denoted by  $B$ . Prove that

$$\frac{\text{Vol}(P)}{\text{Vol}(B)} \leq \left( C \sqrt{\frac{\log(eN/n)}{n}} \right)^n$$

where  $C > 0$  is an absolute constant.<sup>3</sup>

THE BOUND  $|W| \leq N^k$  ASSUMES THE  $k$  VERTICES SELECTED ARE ORDERED OR

DISTINGUISHABLE, WHICH IS TOO STRONG. INSTEAD, CONSIDER

$$x_1 + \dots + x_N = k, \quad (*)$$

WHERE  $x_i$  IS THE NUMBER OF TIMES VERTEX  $i$  IS SELECTED (WHICH CAN BE 0).

THE # OF NONNEGATIVE SOLUTIONS OF (\*) IS

$$\binom{k+N-1}{N-1} = \binom{k+N-1}{k}$$

(SEE, E.G., CH. 1, PROP. 6.1 OF "A FIRST COURSE IN PROBABILITY"). THIS IMPLIES

THAT EQ. (0.3) IN THE BOOK CAN BE REWRITTEN AS

$$\frac{\text{VOL}(P)}{\text{VOL}(B)} \approx \binom{k+N-1}{k} \cdot \frac{1}{k^{n/2}}$$

$$\leq \left( \frac{e(k+N-1)}{k} \right)^k \frac{1}{k^{n/2}}$$

BY EXERCISE 0.6

$$\leq \left( \frac{e2N}{k} \right)^k \frac{1}{k^{n/2}}$$

SINCE  $k+N-1 \leq 2N$

$$\leq \left[ 2 \left( \frac{eN}{n} \right) \log(eN/n) \right]^k \frac{1}{k^{n/2}} \quad \text{TAKING } k_0 = n / \log(eN/n)$$

TAKING  $k_0 = n / \log(eN/n)$

$$\leq 2^n e^{\frac{n}{\log(eN/n)}} \left[ \log(eN/n) \frac{1}{\log(eN/n)} \right]^k \left[ \frac{\log(eN/n)}{n} \right]^{n/2} \quad \text{SINCE } \frac{n}{\log(eN/n)} \leq n$$

$$\leq 2^n e^n e^n \left[ \frac{\log(eN/n)}{n} \right]^{n/2}$$

TAKE  $x = \log(eN/n)$  AND  $y = x^{1/x}$   
THEN  $\log y = 1 \Rightarrow y = e$

$$\leq C^n \left[ \frac{\log(eN/n)}{n} \right]^{n/2}$$

THIS PROOF RELIED HEAVILY ON THE HINT IN THE BOOK. ITS MOST SIGNIFICANT PART WAS THE SUGGESTION OF  $k_0$ . WHY THIS VALUE? OBSERVE THAT THE FIRST INEQ. IS MADE OF TWO FACTORS: THE COVERING NUMBER (GIVEN BY THE BINOMIAL

COEFFICIENT) AND THE VOLUMETRIC TERM ( $k^{-n/2}$ ). ONCE THE COVERING NUMBER HAS BEEN BOUNDED, THE NEED TO KEEP THE RED TERM BOUNDED DICTATES THE FORM OF  $k_0$ .

IT IS ALSO INTERESTING TO COMPARE THE BOUND OF THIS EXERCISE TO THAT OF THEOREM 0.0.4 IN THE BOOK, TRY A COUPLE WITH SMALL  $N$  AND  $n$ , AND A COUPLE WITH LARGE  $N$  AND  $n$ . YOU'LL SEE THAT THE NEW BOUND IS MUCH TIGHTER!