作业 2 参考答案

2019 年 7 月 4 日

由于本次习题相对较为开放，因此评分也会比较灵活，言之有理即可，故给出的参考评分标准也比较粗略，还请见谅。

Exercise 2.25

Almost all of the volume of a ball in high dimensions lies in a narrow slice of the ball at the equator. However, the narrow slice is determined by the point on the surface of the ball that is designated the North Pole. Explain how this can be true if several different locations are selected for the location of the North Pole giving rise to different equators.

参考解答： （15 分）

使用 Union Bound 说明在所有的赤道之外的点是非常少的，或者用类似书中 2.7 的方法使用积分来估计重叠的 “赤道部分” 的界，下面提供 Union Bound 做法：

我们已经分析过了一个单位球的情况，即至少有 $1-\frac{2}{c}e^{-\frac{c^2}{2}}$ 的体积在 $|x_1| \leq \frac{c}{\sqrt{d-1}}$ 的赤道环附近。

考虑单位球上的均匀分布，上述定理告诉我们 $P(|x_i| \geq \frac{c}{\sqrt{d-1}}) \leq \frac{2}{c}e^{-\frac{c^2}{2}}$。

那么当两个赤道面相交呢？这时候就可以用 Union Bound：

$$P(|x_i| \geq \frac{c}{\sqrt{d-1}}, |x_j| \geq \frac{c}{\sqrt{d-1}}) \leq P(|x_i| \geq \frac{c}{\sqrt{d-1}}) + P(|x_j| \geq \frac{c}{\sqrt{d-1}}) \leq \frac{4}{c}e^{-\frac{c^2}{2}}$$

上述推导虽然是两个赤道面 “正交” 的情况，但对于不 “正交” 的情况，推导过程是一模一样的。

从概率的角度回到均匀球体的体积，我们得到球体的大部分体积在两个赤道切片的相交处。

评分标准：

解答中正确体现了有关两个赤道面相交的体积的计算 8 分

解答过程得到了与 d 相关的界，如参考解答中的 $P(|x_i|) \geq \frac{c}{\sqrt{d-1}}$，或者 $P(X) \leq f(d)$ 等 5 分

得到如参考答案这样指数级别的界，或者内含指数级别的界 2 分

其他角度的回答，只要涉及 “多个赤道相交” 的概念的都会酌情给分。

Exercise 2.36

Define the equator of a d-dimensional unit cube to be the hyperplane $\left\{\mathbf{x} | \sum_{i=1}^{d} x_i = \frac{d}{2}\right\}$.

1. *Are the vertices of a unit cube concentrated close to the equator?*

2. *Is the volume of a unit cube concentrated close to the equator?*

3. *Is the surface area of a unit cube concentrated close to the equator?*

参考解答： （30 分）

d 维空间当中的一个点 $\mathbf{x} = \{x_1, x_2, \ldots, x_d\}$ 到题中所定义赤道的距离为 $\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}|$。

分别对 $x_i \sim B(\frac{1}{2}), U[0,1]$ 的情形使用各类 Tail Bound 工具。

1. 可以把 x_i 的取值两种取值等价为有 $\frac{1}{2}$ 的概率取 1，$\frac{1}{2}$ 的概率取 0 的伯努利随机变量。

   则关于顶点是否集中在赤道面附近的讨论，转换为了 $P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq \epsilon)$ 的界的判定。

   这里有两种思路，第一是用 Chebyshev's inequity，得到 $P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq \epsilon) \leq \frac{1}{4\epsilon^2}$。

   第二种思路是使用 Chernoff bound，得到 $P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq \epsilon) \leq 2e^{-2\epsilon^2}$。

   这里得到了两个不等式。注意到这里的不等式与我们之前研究的有相当的区别。此处我们得到的界并不会随着 d 的增大而变换。事实上，当 $d \to \infty$ 时，由中心极限定理，$P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq t)$ 会变成正态分布，即最终的分布会趋于一个不为零的定值。

2. 做法与上一题相同，只不过此时 x_i 被视为 $[0,1]$ 上的均匀分布。

   Chebyshev's inequity: $P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq \epsilon) \leq \frac{1}{12\epsilon^2}$

   Chernoff bound: $P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq \epsilon) \leq 2e^{-2\epsilon^2}$

3. 第 3 问略有不同。首先，表面的定义为 $\{(x)|x_i = t, 0 \leq x_j \leq 1, j \neq i\}$，$t = 0, 1$。即需要固定一个维度上的值即可。

   对应到概率，这等价于 $\mathbf{x}$ 有相同的概率落到每个 $2d$ 个表面上，而每个表面上都是均匀分布的。

   通过条件概率公式，以及对称关系，得到

   $$\begin{aligned} & P(\tfrac{1}{\sqrt{d}}|\textstyle\sum_{i=1}^{d} x_i - \tfrac{d}{2}| \geq \epsilon) \\ = \ & \tfrac{1}{2}P(|0 + \textstyle\sum_{i=2}^{d} x_i - \tfrac{d}{2}| \geq \epsilon) + \tfrac{1}{2}P(|1 + \textstyle\sum_{i=2}^{d} x_i - \tfrac{d}{2}| \geq \epsilon) \\ = \ & \tfrac{1}{2}P(\tfrac{1}{\sqrt{d}}|(\textstyle\sum_{i=2}^{d} x_i - \tfrac{d-1}{2}) - \tfrac{1}{2}| \geq \epsilon) + \tfrac{1}{2}P(\tfrac{1}{\sqrt{d}}|(\textstyle\sum_{i=2}^{d} x_i - \tfrac{d-1}{2}) + \tfrac{1}{2}| \geq \epsilon) \\ = \ & P(\tfrac{1}{\sqrt{d}}|(\textstyle\sum_{i=2}^{d} x_i - \tfrac{d-1}{2}) - \tfrac{1}{2}| \geq \epsilon) \end{aligned}$$

   记 $S = \frac{1}{\sqrt{d-1}}(\sum_{i=2}^{d} x_i - \frac{d-1}{2})$，

   $$P(|\sqrt{1-\tfrac{1}{d}}S - \tfrac{1}{2\sqrt{d}}| \geq \epsilon) = P(S \geq \tfrac{\frac{1}{2\sqrt{d}}+\epsilon}{\sqrt{1-\frac{1}{d}}}) + P(S \leq \tfrac{\frac{1}{2\sqrt{d}}-\epsilon}{\sqrt{1-\frac{1}{d}}}) = P(S \geq \epsilon + O(1)) + P(S \leq -\epsilon + O(1))$$

   由此可以沿用第二问的结果：

   Chebyshev's inequity: $P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq \epsilon) \leq \frac{1}{12\epsilon^2} + O(1)$

   Chernoff bound: $P(\frac{1}{\sqrt{d}}|\sum_{i=1}^{d} x_i - \frac{d}{2}| \geq \epsilon) \leq 2e^{-2\epsilon^2} + O(1)$

关于 concentrate 的标准问题。在第一问已经提到了，这里的分布和球上的分布有所不同。有些同学认为，这里的 bound 应该随着 d 的增大趋于零，才叫集中，而这里最终是一个正态分布，所以不能叫集中。而另外有一些同学认为这里的 bound 虽然为常数，但与两点的最长距离 $\sqrt{d}$ 的比值却不断趋于零，也能叫集中。对于两种观点，批改作业时均接受。

关于点到赤道的距离的计算（答案不要求）：

$$s(\mathbf{x}^*)^2 = \min_{\mathbf{x}\in\{\mathbf{x}|\sum_{i=1}^{d} x_i = d/2\}} \sum_{i=1}^{d} (\mathbf{x_i^*} - \mathbf{x_i})^2$$

利用拉格朗日乘子法即可得到。

评分标准：（每小问分值平均分配）

涉及距离或距离公式 3 分

根据距离公式给出正确的表达式（尤其注意第三问不能 “显然” 地套用第二问低一维度的结果）12 分

对表达式进行合理的放缩，并得到正确的结果 9 分

给出合理的 concentrate 判断标准，并在每小问做出判断 6 分

Exercise 2.37

Consider a non-orthogonal basis $\mathbf{e_1}, \mathbf{e_2}, \ldots, \mathbf{e_d}$. *The* $\mathbf{e_i}$ *are a set of linearly independent unit vectors that span the space.*

参考解答： （25 分）

1. 利用线性无关性反证。

2. $\frac{1}{2}$

3. 不是。反例如 $\mathbf{e_1} = (-2, 1), \mathbf{e_2} = (1, 0), \mathbf{a} = (0, 1)_e, \mathbf{b} = (1, 2)_e$。

4. (a) $(1, \sqrt{2})_e$

   (b) $(\sqrt{2}, 1)_e$

   (c) $(3, 2\sqrt{2})_e$

评分标准：

第 1 问 10 分

第 2 问 4 分（距离的平方或者距离皆可）

第 3 问 5 分

第 4 问每题 2 分

Exercise 2.40

In d-dimensions there are exactly d-unit vectors that are pairwise orthogonal. However, if you wanted a set of vectors that were almost orthogonal you might squeeze in a few more. For example, in 2-dimensions if almost orthogonal meant at least 45 degrees apart, you could fit in three almost orthogonal vectors. Suppose you wanted to find 1000 almost orthogonal vectors in 100 dimensions. Here are two ways you could do it:

- *Begin with 1,000 orthonormal 1,000-dimensional vectors, and then project them to a random 100-dimensional space.*

- *Generate 1,000 100-dimensional random Gaussian vectors.*

Implement both ideas and compare them to see which does a better job.

参考解答： （25 分）

当第一种方法使用高斯随机向量张成 100 维空间时，两种方法在效果上没有区别。

粗糙的理论证明思路如下：

首先，若第一种方法取单位矩阵 I，很显然两者等价。

当第一种方法取正交矩阵 A 时，对任意的高斯随机向量，其分布密度函数满足 $p(A\mathbf{x}) = p(\mathbf{x})$，所以我们得到 $A\mathbf{x_1}, A\mathbf{x_2}, ..., A\mathbf{x_n}$ 等价于直接生成 100 个 1000 维的高斯随机向量，而生成 100 个 1000 维的高斯随机向量和生成 1000 个 100 维的高斯随机向量是没有区别的。

评分标准：

实现代码 10 分

给出实验的数据结果，且符合代码运行结果 7 分

给出结论，结论符合数据 7 分

理论分析 1 分