

Due: Friday 2/7, at 11:59pm on Gradescope

Please follow the homework policies on the course website.

1. (8 pt.) Poisson Approximation

Suppose that n balls are thrown into n bins independently and uniformly at random. Let the random variable X_i denote the number of balls that end up in the i -th bin.

- (a) **(1 pt.)** Consider $\Pr[X_1 = \dots = X_n = 1]$, the probability that each bin receives exactly one ball. Explain why this is equal to $n!/n^n$.
- (b) **(3 pt.)** Alternatively, we can approximate the above probability by replacing X_1, \dots, X_n with independent Poisson random variables $Y_1, \dots, Y_n \sim \text{Poi}(1)$. Find $\Pr[Y_1 = \dots = Y_n = 1]$, and prove that it is smaller than the actual probability by a factor of $\Theta(\sqrt{n})$. That is, show that there are constants $C_1, C_2 > 0$ so that:

$$C_1\sqrt{n} \leq \frac{\Pr[X_1 = \dots = X_n = 1]}{\Pr[Y_1 = \dots = Y_n = 1]} \leq C_2\sqrt{n}. \tag{1}$$

[**HINT:** *Stirling's approximation* $\sqrt{2\pi}n^{n+\frac{1}{2}}e^{-n} \leq n! \leq en^{n+\frac{1}{2}}e^{-n}$ *might be handy.*]

While the above approximation is off by a $\Theta(\sqrt{n})$ factor, we will show in the remainder of this problem that the previous example is essentially the worst case: informally, $\mathbb{E}[\text{any function of } X_1, \dots, X_n]$ is within a $\Theta(\sqrt{n})$ factor of the corresponding expression for Y_1, \dots, Y_n .

We'll state that formally (and you'll prove it!) in the next part, but we'll need the following two facts about $(X_i)_{i=1}^n$ and $(Y_i)_{i=1}^n$:

- $\Pr[Y_1 + \dots + Y_n = n] \geq \frac{1}{e\sqrt{n}}$.
- Conditioning on $Y_1 + \dots + Y_n = n$, the distribution of (Y_1, \dots, Y_n) is the same as that of (X_1, \dots, X_n) .

You may assume these facts without proof for the next part.

- (c) **(4 pt.)** Let \mathbb{N} denote $\{0, 1, 2, \dots\}$ and $f : \mathbb{N}^n \rightarrow [0, +\infty)$ be an arbitrary function that maps n nonnegative integers to a nonnegative real number. Define the random variables $(X_i)_{i=1}^n$ and $(Y_i)_{i=1}^n$ as above. Prove the following inequality:

$$\frac{\mathbb{E}[f(X_1, \dots, X_n)]}{\mathbb{E}[f(Y_1, \dots, Y_n)]} \leq e\sqrt{n}. \tag{2}$$

[**NOTE:** *This states that the Poisson approximation* $\mathbb{E}[f(Y)]$ *underestimates* $\mathbb{E}[f(X)]$ *by a factor of at most* $O(\sqrt{n})$. *When the function* $f(t_1, \dots, t_n)$ *is chosen to be the indicator of* $t_1 = \dots = t_n = 1$, *$\mathbb{E}[f(X)]$ and* $\mathbb{E}[f(Y)]$ *are exactly* $\Pr[X_1 = \dots = X_n = 1]$ *and* $\Pr[Y_1 = \dots = Y_n = 1]$, *and the bound (2) is consistent with inequality (1).*]

- (d) **(0 pt.) [Optional: this won't be graded]** Prove the facts listed before part (c).
2. **(4 pt.)** We showed that Bourgain's embedding allows us to embed an arbitrary metric space (X, d) with $|X| = n$ into (\mathbb{R}^k, ℓ_1) with target dimension k being $O((\log n)^2)$ and distortion being $O(\log n)$. Moreover, the embedding can be computed efficiently using a randomized algorithm. Prove that the exact same embedding computed by the randomized algorithm also achieves $O(\log n)$ distortion with high probability when the target metric is ℓ_p for $p > 1$. We encourage you to emphasize only the differences from the proof in the lecture notes rather than copying the entire proof.
[HINT: Let $f : X \rightarrow \mathbb{R}^k$ denote the relevant embedding. For any two points $x, y \in X$, we showed that $\|f(x) - f(y)\|_1 \leq k \cdot d(x, y)$. Can we say something similar about $\|f(x) - f(y)\|_p$?]
[HINT: For any two points $a, b \in \mathbb{R}^k$ and $p > 1$, it holds that $\|a - b\|_p \geq k^{(1/p)-1} \|a - b\|_1$. This is a special case of Hölder's inequality.]
3. **(10 pt.)** The *equilateral dimension* of a metric space is the maximum number of points in the space that are all at the same distance from each other. In this problem, we will determine the equilateral dimension of the d -dimensional Euclidean space $\mathcal{E}^d = (\mathbb{R}^d, \ell_2)$, and also explore a relaxed version of the equilateral dimension where the pairwise distances are only required to be *approximately* the same.
- (a) **(2 pt.)** Let X be a set of $d + 1$ points in \mathcal{E}^m for some arbitrary positive integer m . Show that (X, ℓ_2) can be isometrically embedded into \mathcal{E}^d .
[HINT: Suppose $X = \{x_0, \dots, x_d\}$. Consider the linear subspace spanned by the vectors $x_i - x_0$ for $i = 1, \dots, d$. How large can the dimension of the subspace be?]
- (b) **(2 pt.)** For all positive integers d , use Part (a) to show that there exist $d + 1$ points in \mathcal{E}^d that are all at distance 1 from each other. This shows that the equilateral dimension of \mathcal{E}^d is at least $d + 1$.
- (c) **(0 pt.) [Optional: this won't be graded.]** For all positive integers d , show that one cannot find $d + 2$ points in \mathcal{E}^d that are all at distance 1 from each other. Together with Part (b), this shows that the equilateral dimension of \mathcal{E}^d is exactly $d + 1$.
[HINT: Prove by contradiction. Suppose there are $d + 2$ such points x_0, x_1, \dots, x_{d+1} . Consider the vectors $v_i = x_i - x_0$ and the matrix $A = [v_1, \dots, v_{d+1}]$ (with the v_i as columns) of size $d \times (d + 1)$. Explicitly compute the inner products $v_i^\top v_j$ and the matrix $A^\top A$. Use the fact that $\text{rank}(A^\top A) = \text{rank}(A)$ to derive a contradiction.]
- (d) **(3 pt.)** Show that there exists a constant $c > 0$ such that for all positive integers d , one can find at least 2^{cd} points in \mathcal{E}^d that are all at distances between 1 and 1.1 from each other.
- (e) **(3 pt.)** Show that there exists a constant $C > 0$ such that for all positive integers d , one cannot find more than 2^{Cd} points in \mathcal{E}^d that are all at distances between 1 and 1.1 from each other.
[HINT: You can use the fact that the volume of a d -dimensional ball of radius r is equal to $\gamma_d r^d$ for some constant γ_d that depends on d but not r . It may be helpful if you understand the proof of Lemma 3 in lecture notes #9, but that is not required for working on this problem.]