Problem Set 1

[Your name] and [student ID]
MAT1801-2020

Problem 1 [BHK 2.4] (10 points). Show that for any $c \geq 1$, there exist distributions for which Chebyshev's inequality is tight, in other words, $Prob(|x - \mathbb{E}(x)| \geq c) = Var(x)/c^2$.

Problem 2 [BHK 2.20] (10 points). Consider a unit ball A centered at the origin and a unit ball B whose center is at distance s from the origin. Suppose that a random point x is drawn from the mixture distribution: "with probability 1/2, draw at random from A; with probability 1/2, draw at random from B". Show that a separation $s \gg 1/\sqrt{d-1}$ is sufficient so that $Prob(x \in A \cap B) = o(1)$; i.e., for any $\epsilon > 0$, there exists c such that if $s \geq c/\sqrt{d-1}$, then $Prob(x \in A \cap B) < \epsilon$. In other words, this extent of separation means that nearly all of the mixture distribution is identifiable.

Problem 3 [BHK 2.39] (10 points). 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:

1. Begin with $1,000$ orthonormal $1,000$-dimensional vectors, and then project them to a random 100-dimensional space.
2. Generate 1000 100-dimensional random Gaussian vectors.

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