Set 1. Due January 30, 2017

Problem 1 Let X be a real-valued random variable with mean m, median M, and standard deviation σ. Prove that
$$|m - M| \le \sqrt{2}\sigma \ .$$

Problem 2 Write a program that compares the performance of the empirical mean and the median-of-means mean estimators. Test them on randomly generated samples drawn from different distributions, including heavy-tailed ones. Try different parameters of the median-of-means estimator and different sample sizes. Compare the estimators according to different measures, such as average deviation from the true mean, as well as worst case deviation (when the random sample is re-drawn many times). You may consider the Pareto family or Student's t-distribution (with different degrees of freedom) for heavy-tailed examples.

Problem 3 Let $X_1, \ldots, X_n$ be i.i.d. *non-negative* random variables with mean $\mathbf{E}X_1 = m$ and second moment $\mathbf{E}X_1^2 = a^2$. Use the Chernoff bound to prove that, for all $t \in (0, m)$,

$$\mathbf{P}\left\{ \frac{1}{n}\sum_{i=1}^{n} X_i < m - t \right\} \le e^{-nt^2/(2a^2)} \ .$$

Hint: use the fact that for $x > 0$, $e^{-x} \le 1 - x + x^2/2$.

Problem 4 Write a program that projects the n standard basis vectors in $\mathbb{R}^n$ to a random 2-dimensional subspace. (You may do this simply by using a $2 \times n$ matrix whose entries are i.i.d. normals.) Center the point set appropriately and re-scale such that the empirical variance of the (say) first component equals 1. Plot the obtained point set. Now generate n independent standard normal vectors on the plane and compare the two plots. Do this for a wide range of values of n. What do you see?

Repeat the same exercise but now projecting the 2^n vertices of the hypercube $\{-1, 1\}^n$ instead of the standard basis vectors. (Naturally, you can only do this for small values of n, say up to $n \approx 13$.) What do you see now?