• **DISCLAIMER:** These notes are not necessarily an accurate representation of what I said during the class. They are mostly what I intend to say, and have not been carefully edited.

• **Summary**
  - Why concentration bounds?
  - Quick proof of Chernoff bound.
  - The matrix exponential.
  - Proof of matrix Chernoff bound.

• **Matrix Concentration (Theorem 1.1. of [http://arxiv.org/pdf/1004.4389.pdf](http://arxiv.org/pdf/1004.4389.pdf), with $R = 0.1$ and $\mu_{\text{min}} = \mu_{\text{max}} = \mu = O(\log n)$):**

  **Theorem 0.1.** Consider a finite sequence $X_1 \ldots X_m$ of independent, random symmetric matrices with dimension $n$ with $0 \preceq X_i \preceq 0.1I$. Let their sum be $X = \sum_i X_i$. Then with high probability $X \approx I$.

• **Where do concentration bounds fit into the bigger picture**
  - Immediate use: sparsification
    * Chernoff bounds: tight up to $\log n \epsilon^{-1}$. Can reverse engineer the ‘right’ sampling probabilities.
    * Full dimensional case: intrinsic dimension.
    * Often in practice: low dimensional data.
  - More sophisticated uses of sparsifiers:
    * Iterative methods: given $A \approx B$, can solve problems in $B$ using problems in $A$.
    * Proof of iterative methods: potential function reduction.
    * (coincidental / remarkable) similarity between these potentials and ones that we use to prove concentration bounds.

• **Single variate case:**
  - Potential function: $\exp(x)$. 
Probability of big: $E \left[ \exp(x) \right] / \exp(2\mu)$.

Independence: $E \left[ \exp(x) \right] = \prod_{i=1}^{m} E_{x_i} \left[ \exp \left( x_i \right) \right]$.

Algebraic manipulation: $E_{x_i} \left[ \exp \left( x_i \right) \right] \leq \exp \left( 1.5 \ E \left[ x_i \right] \right)$.

* Extreme case (via convexity of $\exp$):
  $x_i = \begin{cases} 
  0 & \text{w.p. } 1 - p \\
  0.1 & \text{w.p. } p 
  \end{cases}$

  * $E \left[ x_i \right] = 0.1p$.
  * $E_{x_i} \left[ \exp \left( x_i \right) \right] = (1 - p) + p \exp \left( 0.1 \right)$.
  * Plot, or do algebra using $| \exp(p) - 1 - p | \leq 0.2p$ when $p \leq 0.1$.

Combine the product gives w.h.p.:

$$\leq \prod_{i=1}^{m} \exp \left( 1.5 \ E \left[ x_i \right] \right) / \exp(2\mu) = \exp \left( 1.5\mu \right) / \exp(2\mu) = \exp \left( -O \left( \log n \right) \right)$$

** Doing this for matrices: the matrix exponential.

* Exponentiate each eigenvalue: if $X = \sum_i \lambda_i u_i u_i^T$, $\exp \left( X \right) = \sum_i \exp \left( \lambda_i \right) u_i u_i^T$.

  * Potential function: $\text{tr} \left[ \exp \left( X \right) \right] = \sum_i \exp \left( \lambda_i \left( X \right) \right) \geq \exp \left( \lambda_{\text{max}} \left( X \right) \right)$.

  * Major difference: $\exp \left( A + B \right) \neq \exp \left( A \right) \exp \left( B \right)$.

** Fix for non-commutivity:

  * Lieb’s inequality: for a fixed matrix $H$ and a random matrix $X$,
    $$E_X \left[ \text{tr} \left[ \exp \left( H + X \right) \right] \right] \leq \text{tr} \left[ \exp \left( H + \log \left( E_X \left[ \exp \left( X \right) \right] \right) \right) \right]$$

  * Repeatedly applying this allows us to move each expectation over $X_i$ inside the matrix exponential:
    $$E_{X_1 \ldots X_k} \left[ \text{tr} \left[ \exp \left( \sum_{i=1}^{k} X_i \right) \right] \right] \leq \text{tr} \left[ \exp \left( \sum_{i=1}^{k} \log \left( E_{X_i} \left[ \exp \left( X_i \right) \right] \right) \right) \right]$$

** Finishing off in the Bernoulli case:

  * Let $Y_i = \log \left( E_{X_i} \left[ \exp \left( X_i \right) \right] \right)$.
  * $\exp \left( X_i \right)$, $I$ commute, so can work with the scalar case.
  * Matrix version of $E \left[ \exp \left( x \right) \right] \leq \exp(1.5 E \left[ x \right])$: $Y_i \leq 1.5 E \left[ X_i \right]$.
  * Maximum eigenvalue of $\sum_i Y_i$ is most $\lambda_{\text{max}} \left( \sum_i E \left[ X_i \right] \right) = 1.5\mu$.
  * $\text{tr} \left[ \exp(\sum_i Z_i) \right] \leq n \exp \left( 1.5\mu \right)$.
  * Extra $n$ absorbed by $\exp(-\mu)$ again.