Theory: estimating the mean

Theorem

$\mathbb{E}\,\hat{\bar{x}} = \bar{x}$ and $\|\hat{\bar{x}} - \bar{x}\|_\infty \le t$ with probability greater than

$$1 - 2p\,\mathit{exp}\left(\frac{-N\gamma t^2/2}{\|X\|^2_{\textit{max-row}} + t/3\|X\|_{\textit{max-entry}}}\right)$$

where $\gamma = p_{small}/p$ *(simplifying to $p_{small} \ll p \ll N$)*

If X has normalized columns, then $\|X\|_{\text{max-entry}} = 1$ and $\|X\|_{\text{max-row}} = \sqrt{N}$ are possible, which is bad.

Lemma

If X is preconditioned, then *(for $Np = 10^{10}$, $\sqrt{\log(2Np) + 1000} = 32$)*

$$\mathbb{P}\left\{\|X\|_{\textit{max-entry}} \ge \frac{\sqrt{2}}{\sqrt{p}} \cdot \sqrt{\log(2Np) + 1000}\right\} \le .001$$

$$\mathbb{P}\left\{\|X\|_{\textit{max-row}} \ge \frac{\sqrt{2N}}{\sqrt{p}} \cdot \sqrt{\log(2Np) + 1000}\right\} \le .001$$