Concentration Inequalities Reference

Lily Chung

Concentration inequalities are inequalities which state that a random variable X is unlikely to be "too far" from its expected value $\mathbf{E}X$.

We will typically think of events like $X \geq \mathbf{E}X + t$ or $X \leq \mathbf{E}X - t$ (where $t > 0$) as "bad events", where the variable X deviates far from its mean. We would like to ensure that such bad events don't happen by writing down an inequality like

$$\Pr[X \geq \mathbf{E}X + t] \leq \delta$$

for some "small" probability δ.

Concentration inequalities give us the means to do this. This note is a reference for useful concentration inequalities which appear often in the analysis of randomized algorithms. We won't give proofs for these inequalities.

What is δ?

Two regimes for δ to keep in mind are (1) when δ is a small constant (e.g. $\delta = 1\%$) and (2) when δ is a function of n which goes to zero (e.g. $\delta = \Theta(n^{-2})$ or $\delta = \Theta(e^{-n})$).

A very common reason we might need to pick, say, $\delta = \Theta(n^{-2})$ is if we actually have n^2 random variables $X_1, \ldots, X_{n^2}$, and we want to be sure that *none* of the bad events $X_i \geq \mathbf{E}X_i + t$ occur. A straightforward way to do so is to first show that the inequality

$$\Pr[X_i \geq \mathbf{E}X_i + t] \leq \delta$$

holds for each i with $\delta = 1\% \cdot n^{-2}$, and then apply the union bound:

$$\Pr[\text{any of the } X_i \geq \mathbf{E}X_i + t \text{ occur}] \leq \sum_i \Pr[X_i \geq \mathbf{E}X_i + t] \leq n^2\delta = 1\%.$$

Markov

Theorem 1 (Markov's inequality)**.**

Let $X \geq 0$ be a *nonnegative* random variable with finite mean. Then

$$\Pr[X \geq t] \leq \frac{\mathbf{E}X}{t}$$

for any $t > 0$.

- Rephrased: For any $\delta \in (0, 1)$, the inequality $X < \mathbf{E}X \cdot \delta^{-1}$ holds with probability at least $1 - \delta$.
- Simple and applies to any nonnegative random variable.
- Disadvantage: Only a linear relationship between t and δ.

Chebyshev

Theorem 2 (Chebyshev's inequality).

Let X be a random variable with finite variance. Then

$$\Pr[|X - \mathbf{E}X| \geq t] \leq \frac{\operatorname{Var} X}{t^2}$$

for any $t > 0$.

- Rephrased: X lies in the interval $\mathbf{E}X \pm \sqrt{\operatorname{Var} X \cdot \delta^{-1}}$ with probability $1 - \delta$.
- Useful when $X = \sum_i X_i$ is a sum of pairwise independent X_i since then $\operatorname{Var} X = \sum_i \operatorname{Var} X_i$.
- Disadvantage: Still only a quadratic relationship between t and δ.

Hoeffding

Theorem 3 (Hoeffding's inequality).

Let $X = \sum_i X_i$ where $X_1, \ldots, X_k$ are *independent* random variables such that $X_i \in [a_i, b_i]$ for all i. Then

$$\Pr[X \geq \mathbf{E}X + t] \leq \exp\left(-\frac{2t^2}{R}\right)$$

for any $t \geq 0$, where $R = \sum_i (b_i - a_i)^2$.

- Rephrased: $X \leq \mathbf{E}X + \sqrt{\frac{1}{2} R \ln \delta^{-1}}$ with probability $1 - \delta$.
- The inequality $\Pr[X \leq \mathbf{E}X - t] \leq \exp\left(-\frac{2t^2}{R}\right)$ is obtained by applying Theorem 3 to $-X$.
- The inequality $\Pr[|X - \mathbf{E}X| \geq t] \leq 2\exp\left(-\frac{2t^2}{R}\right)$ is obtained by union bounding Theorem 3 and the previous inequality together.
- Exponential relationship between t and δ.
- Disadvantage: The X_i must be fully independent.

Chernoff

Theorem 4 (Chernoff bound).

Let $X = \sum_i X_i$ where $X_1, \ldots, X_k$ are *independent* random variables such that $X_i \in [0, 1]$ for all i. Then

$$\Pr[X \geq \mathbf{E}X \cdot t] \leq e^{-\mathbf{E}X \cdot h(t)} \qquad \text{for } t \geq 1$$

$$\Pr[X \leq \mathbf{E}X \cdot s] \leq e^{-\mathbf{E}X \cdot h(s)} \qquad \text{for } s \in (0, 1)$$

where $h(t) = t \ln t - t + 1$.

- If the exact value of $\mathbf{E}X$ is unknown but we have the bounds $L \le \mathbf{E}X \le U$, then the inequalities

$$\begin{aligned} \Pr[X \ge U \cdot t] &\le \exp(-U \cdot h(t)) && \text{for } t \ge 1 \\ \Pr[X \le L \cdot s] &\le \exp(-L \cdot h(s)) && \text{for } s \in (0,1) \end{aligned}$$

  still hold.[1]

- Exponential relationship between $\mathbf{E}X \cdot t$ and δ.

- Disadvantage: The X_i must be fully independent.

To apply Theorem 4, we often want to lower-bound $h(t)$. The following looser bounds are obtained by this method. I recommend graphing $h(t)$ to get a sense of which lower bound is appropriate for the particular regime of t appearing in your application.

Corollary 5 (Chernoff bounds (looser versions))**.**

Let $X = \sum_i X_i$ where $X_1, \ldots, X_k$ are *independent* random variables such that $X_i \in [0,1]$ for all i. Then:

$$\begin{aligned} \Pr[X \ge \mathbf{E}X \cdot t] &\le \exp\left(-\mathbf{E}X \cdot \frac{(t-1)^2}{t+1}\right) && \text{for } t \ge 1 \\ \Pr[X \ge \mathbf{E}X \cdot t] &\le \exp(-\mathbf{E}X \cdot t) && \text{for } t \ge 6.4 \\ \Pr[X \ge \mathbf{E}X \cdot t] &\le \exp(-\mathbf{E}X \cdot 2t) && \text{for } t \ge 19.1 \\ \Pr[X \le \mathbf{E}X \cdot s] &\le \exp\left(-\mathbf{E}X \cdot \frac{(t-1)^2}{2}\right) && \text{for } s \in (0,1) \end{aligned}$$

[1] The second one is easy but the first is trickier; it can be shown by applying Theorem 4 to $Y = X + U - \mathbf{E}X$, considering $U - \mathbf{E}X$ as a sum of $\lceil U - \mathbf{E}X \rceil$ trivial random variables.