0368-3248-01-Algorithms in Data Mining **Fall 2013**

Lecture 2: Probabilistic Inequalities

Lecturer: Edo Liberty

Warning: *This note may contain typos and other inaccuracies which are usually discussed during class. Please do not cite this note as a reliable source. If you find mistakes, please inform me.*

Fact 0.1 (Markov's inequality)**.** *For any positive random variable X:*

$$\Pr(X > t) \le \frac{E[X]}{t} \tag{1}$$

Fact 0.2 (Chebyshev's inequality)**.** *For any random variable X*

$$\Pr[|X - E[X]| > t] \le \frac{\sigma^2(X)}{t^2} \tag{2}$$

Theorem 0.1 (Chernoff's bound)**.** *Let X_i be a set of* **independent** *random variables such that $\mathbb{E}[X_i] = 0$ and $|X_i| \le 1$ almost surely. Also define $\sigma_i^2 = \mathbb{E}[X_i^2]$ and $\sigma^2 = \sum_i \sigma_i^2$. Then:*

$$\Pr[\sum_i X_i \ge t] \le max(e^{-t^2/4\sigma^2}, e^{-t/2})$$

Proof.

$$\begin{aligned}
\Pr[\sum_i X_i \ge t] &= \Pr[\lambda \sum_i X_i \ge \lambda t] \quad (\text{for } \lambda \ge 0) && (3)\\
&= \Pr[e^{\lambda \sum_i X_i} \ge e^{\lambda t}] \quad (\text{because } e^x \text{ is monotone}) && (4)\\
&\le \mathbb{E}[e^{\lambda \sum_i X_i}]/e^{\lambda t} \quad (\text{by Markov}) && (5)\\
&= \Pi_i \mathbb{E}[e^{\lambda X_i}]/e^{\lambda t} && (6)
\end{aligned}$$

Now, for $x \in [0, 1]$ we have that $e^x \le 1 + x + x^2$ so $\mathbb{E}[e^{\lambda X_i}] \le 1 + \mathbb{E}[\lambda X_i] + \lambda^2 \mathbb{E}[X_i^2] \le 1 + \lambda^2 \sigma_i^2$. Now, since $1 + x \le e^x$ we have that $1 + \lambda^2\sigma_i^2 \le e^{\lambda^2\sigma_i^2}$. Combining the above we have that $\mathbb{E}[e^{\lambda X_i}] \le e^{\lambda^2\sigma_i^2}$

$$\begin{aligned}
\Pi_i \mathbb{E}[e^{\lambda X_i}]/e^{\lambda t} &\le \Pi_i \mathbb{E}[e^{\lambda^2\sigma_i^2}]/e^{\lambda t} && (7)\\
&= e^{\lambda^2\sigma^2 - \lambda t} && (8)
\end{aligned}$$

Now, optimizing over $\lambda \in [0, 1]$ we get that $\lambda = min(1, t/2\sigma^2)$ which completes the proof.

0.1 Other Useful forms

Chernoff's inequality: Let $X_1, \ldots, X_n$ be independent $\{0, 1\}$ valued random variables. Each X_i takes the value 1 with probability p_i and 0 else. Let $X = \sum_{i=1}^n X_i$ and let $\mu = E[X] = \sum_{i=1}^n p_i$. Then:

$$\begin{aligned}
\Pr[X > (1+\varepsilon)\mu] &\le e^{-\mu\varepsilon^2/4}\\
\Pr[X < (1-\varepsilon)\mu] &\le e^{-\mu\varepsilon^2/2}
\end{aligned}$$

Or, using the union bound:

$$\Pr[|X - \mu| > \varepsilon\mu] \le 2e^{-\mu\varepsilon^2/4}$$

$\square$