CSC 310: Information Theory **Winter 2026**

Tutorial 1: Concentration of Random Bits and Entropy Computation

Instructor: Swastik Kopparty *TA: Ziyang Jin*

Date: 7 Jan 2026

1.1 Markov's Inequality and Chebyshev's Inequality

Theorem 1.1 (Markov's Inequality) *Let X be a non-negative random variable. Then for any $a > 0$,*

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}.$$

Proof: Note that

$$\begin{aligned}
\mathbb{E}[X] &= \Pr[X < a] \cdot \mathbb{E}[X|X < a] + \Pr[X \geq a] \cdot \mathbb{E}[X|X \geq a] \\
&\geq \Pr[X \geq a] \cdot \mathbb{E}[X|X \geq a] && \text{(since } X \text{ is non-negative)} \\
&\geq \Pr[X \geq a] \cdot a && \text{(since } \mathbb{E}[X|X \geq a] \geq a\text{)}.
\end{aligned}$$

Reordering the terms, we obtain

$$\Pr[X \geq a] \leq \frac{\mathbb{E}[X]}{a}$$

as wanted. ■

Theorem 1.2 (Chebyshev's Inequality) *Let X be a random variable. Let $\mu = \mathbb{E}[X]$ and let $\sigma^2 = \mathrm{Var}(X)$ (the variance). Then for any $k > 0$,*

$$\Pr\left[|X - \mu| \geq k\right] \leq \frac{\sigma^2}{k^2}.$$

Proof: Note that

$$\begin{aligned}
& \Pr[|X - \mu| \geq k] \\
= \ & \Pr[(X - \mu)^2 \geq k^2] && \text{(squaring on both sides)} \\
\leq \ & \frac{\mathbb{E}[(X - \mu)^2]}{k^2} && \text{(apply Markov's inequality on r.v. } (X - \mu)^2\text{)} \\
= \ & \frac{\sigma^2}{k^2} && \text{(since } \mathrm{Var}(X) = \mathbb{E}[(X - \mu)^2]\text{)}.
\end{aligned}$$

■

Alternatively, when $\sigma^2 > 0$ (the non-trivial case), we can also write Chebyshev's inequality as follows:

$$\Pr\left[|X - \mu| \geq k\sigma\right] \leq \frac{1}{k^2}.$$

Remark 1.3 *Markov's inequality, Chebyshev's inequality, and Chernoff bounds are also known as concentration bounds. The rule of thumb for concentration bounds is that the more information you have, the better bound you can obtain. If you only know the expectation and non-negativity of the random variable, then you can use Markov's inequality. If you also know the variance of the random variable, you can use Chebyshev's inequality. In the future, we will also look at Chernoff bounds, which applies specifically to the sum of n independent coin tosses.*

1.2 Concentration of n Independent Random Bits

Let $X_1, ..., X_n$ be n independent random variables, and for $i \in [n]$ (where $[n]$ denotes $\{1, 2, \ldots, n\}$),

$$X_i = \begin{cases} 1 & \text{with probability } p, \\ 0 & \text{with probability } 1-p. \end{cases}$$

We can see that each X_i is an independent random bit, and we have a string of n independent random bits. Let random variable $X = \sum_{i=1}^{n} X_i$. Thus, X can be interpreted as the total number of 1's among these n independent random bits. You can also view X as a binomial random variable with n trials and with probability p. By knowledge of binomial random variable, we have

$$\mathbb{E}[X] = np,$$
$$\mathrm{Var}(X) = np(1-p).$$

By Chebyshev's inequality (assuming $0 < p < 1$), we have

$$\Pr\left[|X - np| \geq k\sqrt{np(1-p)}\right] \leq \frac{1}{k^2}. \tag{1.1}$$

Plugging $k = \frac{\epsilon\sqrt{n}}{\sqrt{p(1-p)}}$ where $\epsilon > 0$ is a constant into (1.1), we obtain

$$\Pr\left[|X - np| \geq \epsilon n\right] \leq \frac{1}{\frac{\epsilon^2}{p(1-p)}n}. \tag{1.2}$$

Note that $\frac{\epsilon^2}{p(1-p)}$ is just some constant. Thus, we have

$$\Pr\left[|X - np| \geq \epsilon n\right] \leq O\left(\frac{1}{n}\right) = o(1) \tag{1.3}$$

as $n \to \infty$. This says that $\Pr[|X - np| \geq \epsilon n]$ can be arbitrarily small for sufficiently large n. For example, if we take $\epsilon = 0.001, p = 0.5, n = 2.5 \times 10^7$, then $\Pr[|X - np| \geq 0.001n] \leq 0.01$, which means 99% of the outcomes lie within the interval $[np - 0.001n, np + 0.001n]$. Equation (1.3) is the conclusion we will use in future classes.

1.3 Entropy Computation and Compression

Consider a discrete random variable Y with alphabet $\{a, b, c\}$ defined as follows:

$$Y = \begin{cases} a & \text{w.prob. } \frac{1}{2}, \\ b & \text{w.prob. } \frac{1}{4}, \\ c & \text{w.prob. } \frac{1}{4}. \end{cases}$$

We compute the entropy of Y as follows:

$$H(Y) = -\frac{1}{2}\log_2\frac{1}{2} - \frac{1}{4}\log_2\frac{1}{4} - \frac{1}{4}\log_2\frac{1}{4} = \frac{1}{2}\log_2\frac{1}{\frac{1}{2}} + \frac{1}{4}\log_2\frac{1}{\frac{1}{4}} + \frac{1}{4}\log_2\frac{1}{\frac{1}{4}} = 1.5.$$

Consider a random string s of length n, where each character is generated independently by Y. To naïvely represent s in binary, since there are 3 outcomes a, b, c, we can use $\lceil \log_2 3 \rceil = 2$ bits to represent each character. Since there are n characters, the naïve encoding takes a total of $2n$ bits.

To compress it, we can use a single bit "0" to represent a, and use two bits "10" to represent b, and use two bits "11" to represent c. Intuitively, since a appears more frequently, we should use fewer bits to represent a, and since b and c appear less frequently, so we are okay with using more bits to represent them. Under this representation, a is expected to appear $\frac{n}{2}$ times in s, and b and c are expected to appear $\frac{n}{4}$ each. Therefore, the expected length of such encoding is $1 \times \frac{n}{2} + 2 \times \frac{n}{4} + 2 \times \frac{n}{4} = 1.5n$.

Claim 1.4 *The encoding scheme above takes* $(1.5 + \epsilon)n$ *bits with probability* > 0.999 *for sufficiently large* n.

Exercise: Use Chebyshev's inequality to convince yourself that the claim above is true.