Bound, Equalities and Inequalities

Hadar Tal

hadar.tal@mail.huji.ac.il

This paper is a summary of the educational materials and lectures from

- **Wikipedia**
- **3Blue1Brown** YouTube channel

Winter 2024

Contents

Chapter 1

Bounds

Chapter 2

Equalities

2.1 Properties of Binomial Coefficients

2.1.1 Symmetry Rule for Binomial Coefficients

Theorem 2.1.1. *For all $n, k \in \mathbb{N}$, the following holds*

$$\binom{n}{k} = \binom{n}{n-k} \tag{2.1}$$

Proof. The proof is by definition of the binomial coefficient

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} \tag{2.2}$$

and the symmetry of the factorial function

$$n! = n \cdot (n-1) \cdot \ldots \cdot 2 \cdot 1 = 1 \cdot 2 \cdot \ldots \cdot (n-1) \cdot n = n! \tag{2.3}$$

which implies that

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{n!}{(n-k)!k!} = \binom{n}{n-k} \tag{2.4}$$

□

Example. *The symmetry rule for binomial coefficients states that the number of ways to choose k elements out of n is the same as the number of ways to choose $n-k$ elements out of n.*

2.1.2 Pascal's Rule for Binomial Coefficients

Theorem 2.1.2. *For all $n, k \in \mathbb{N}$, the following holds*

$$\binom{n}{k} = \binom{n-1}{k} + \binom{n-1}{k-1} \tag{2.5}$$

Proof.

$$\binom{n}{k} = \frac{n!}{k!(n-k)!} = \frac{(n-1)!}{k!(n-1-k)!} + \frac{(n-1)!}{(k-1)!(n-1-(k-1))!} = \binom{n-1}{k} + \binom{n-1}{k-1} \tag{2.6}$$

□

Example. *choosing a team of k players from n candidates: you can either include a specific player in your team and choose the rest k-1 players from the remaining n-1 candidates, or not include that specific player, thus choosing all k players from the remaining n-1 candidates.*

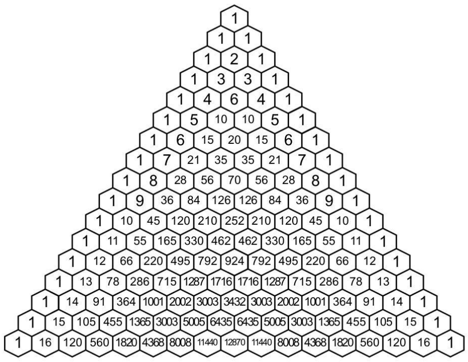

Figure 2.1: Pascal's Triangle

2.1.3 Sum of Binomial Coefficients over Lower Index

Theorem 2.1.3. *For all $n \in \mathbb{N}$, the following holds*

$$\sum_{i=0}^{n} \binom{n}{i} = 2^n \tag{2.7}$$

Proof. The proof is by induction on n. For $n = 0$, the base case is

$$\sum_{i=0}^{0} \binom{0}{i} = \binom{0}{0} = 1 = 2^0 \tag{2.8}$$

For the induction step, assume that the theorem holds for $n = k$. Then

$$\sum_{i=0}^{k+1} \binom{k+1}{i} = \sum_{i=0}^{k+1} \left(\binom{k}{i} + \binom{k}{i-1} \right) = \sum_{i=0}^{k} \binom{k}{i} + \sum_{i=0}^{k} \binom{k}{i-1} = 2^k + 2^k = 2^{k+1} \tag{2.9}$$

□

Example. *The sum of binomial coefficients over the lower index is the same as counting all the subsets of a set of size n which is 2^n.*

2.1.4 Factors of Binomial Coefficient

Theorem 2.1.4. *For all $r \in \mathbb{R}, k \in \mathbb{Z}$, the following holds:*

$$k\binom{r}{k} = r\binom{r-1}{k-1} \tag{2.10}$$

Proof. By definition, the binomial coefficient $\binom{r}{k}$ is given by

$$\binom{r}{k} = \frac{r!}{k!(r-k)!}. \tag{2.11}$$

Multiplying both sides by k, we have

$$k\binom{r}{k} = k\frac{r!}{k!(r-k)!} = \frac{r!}{(k-1)!(r-k)!} = r\frac{(r-1)!}{(k-1)!(r-k)!} = r\binom{r-1}{k-1}. \tag{2.12}$$

□

Example. *Forming a committee of k members from a group of r individuals, with one member to be chosen as chairperson.*

2.1.5 Increasing Sum of Binomial Coefficients

Theorem 2.1.5. *For all $n \in \mathbb{N}$, the following holds*

$$\sum_{k=0}^{n} k\binom{n}{k} = n2^{n-1} \tag{2.13}$$

Proof.

$$\sum_{k=0}^{n} k\binom{n}{k} = \sum_{k=1}^{n} k\binom{n}{k} = \sum_{k=1}^{n} n\binom{n-1}{k-1} = n\sum_{k=1}^{n}\binom{n-1}{k-1} = n\sum_{k=0}^{n-1}\binom{n-1}{k} = n\cdot 2^{n-1} \tag{2.14}$$

□

Chapter 3

Inequalities

3.1 Vector Norms

Theorem 3.1.1. ***Cauchy-Schwarz Inequality***
Let u, v be vectors of an inner product space. Then

$$|\langle u, v\rangle|^2 \leq \|u\| \, \|v\|$$

Theorem 3.1.2. ***Triangle Inequality***
Let u, v be vectors of an inner product space. Then

$$\|u + v\| \leq \|u\| + \|v\|$$

3.2 Probability

3.2.1 Markov's Inequality

Theorem 3.2.1. ***Markov's Inequality***
Let X be a non-negative random variable and $a > 0$. Then

$$P(X \geq a) \leq \frac{E[X]}{a}$$

Proof.

$$E[X] = \int_0^\infty x f(x) dx \geq \int_a^\infty x f(x) dx \geq \int_a^\infty a f(x) dx = a \int_a^\infty f(x) dx = aP(X \geq a) \tag{3.1}$$

□

3.2.2 Chebyshev's Inequality

Theorem 3.2.2. ***Chebyshev's Inequality***
Let X be a random variable with mean μ and variance σ^2. Then

$$P(|X - \mu| \geq a) \leq \frac{\sigma^2}{a^2}$$

Proof.

$$P(|X - \mu| \geq a) = P((X - \mu)^2 \geq a^2) \leq \frac{E[(X - \mu)^2]}{a^2} = \frac{\sigma^2}{a^2} \tag{3.2}$$

□

3.2.3 Chernoff Bound for Sum of Independent Bernoulli Variables

Theorem 3.2.3. ***Chernoff Bound*** *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables with $X_i \sim$ Bernoulli(p_i) and $\Pr(X_i = 1) = p_i$. Define $\mu = E\left[\sum_{i=1}^n X_i\right] = \sum_{i=1}^n p_i$. Then for any $\delta > 0$, the following bounds hold:*

$$\Pr\left(\sum_{i=1}^n X_i \geq (1+\delta)\mu\right) \leq \left(\frac{e^{\delta}}{(1+\delta)^{1+\delta}}\right)^{\mu}, \quad \text{if } \delta > 0$$

$$\Pr\left(\sum_{i=1}^n X_i \leq (1-\delta)\mu\right) \leq \left(\frac{e^{-\delta}}{(1-\delta)^{1-\delta}}\right)^{\mu}, \quad \text{if } 0 < \delta < 1$$

Proof. Define $t = \log(1+\delta)$ for the upper tail bound and $t = -\log(1-\delta)$ for the lower tail bound. By using the moment-generating function approach, we have:

$$E[e^{tX_i}] = p_i e^t + (1-p_i) = 1 + p_i(e^t - 1) \leq e^{p_i(e^t-1)} \tag{3.3}$$

Thus, for the upper bound:

$$\begin{aligned} E\left[e^{t\sum_{i=1}^n X_i}\right] &= \prod_{i=1}^n E[e^{tX_i}] = \prod_{i=1}^n \left(1 + p_i(e^t-1)\right) \\ &\leq \prod_{i=1}^n e^{p_i(e^t-1)} = e^{\sum_{i=1}^n p_i(e^t-1)} = e^{\mu(e^t-1)} \end{aligned} \tag{3.4}$$

$$\Pr\left(\sum_{i=1}^n X_i \geq (1+\delta)\mu\right) \leq \frac{e^{\mu(e^t-1)}}{e^{t(1+\delta)\mu}} = e^{-\mu t\delta}\left(\frac{e^t}{1+\delta}\right)^{\mu} \tag{3.5}$$

Simplifying further using Taylor series approximations and properties of logarithms, we obtain the upper bound as stated in the theorem.
The proof for the lower bound follows similarly, using $t = -\log(1-\delta)$ and applying analogous steps. □

Remarks: This proof illustrates the power of the Chernoff bound for providing exponentially decreasing bounds on the tail probabilities of the sum of independent Bernoulli trials, particularly useful in scenarios involving large numbers of independent events or trials.

Theorem 3.2.4. *Let $X_1, X_2, \ldots, X_n$ be independent Bernoulli random variables where each X_i takes values in $\{0, 1\}$. Define $Z_n = \sum_{i=1}^n X_i$ and $\mu = E[Z_n]$. Then for any $\delta > 0$, it holds that:*

$$\Pr(|Z_n - \mu| \geq \delta\mu) \leq 2e^{-\frac{\delta^2\mu}{2}}. \tag{3.6}$$

3.2.4 Hoeffding's Inequality

Hoeffding's Lemma

Theorem 3.2.5. ***Hoeffding's Lemma*** *Let X be a random variable such that $a \leq X \leq b$. Then for any $\lambda \in \mathbb{R}$,*

$$E\left[e^{\lambda(X - E[X])}\right] \leq \exp\left(\frac{\lambda^2(b-a)^2}{8}\right).$$

Proof. Define the function $\psi(t) = E[e^{t(X-E[X])}]$. We will show that $\psi(t) \leq \exp\left(\frac{t^2(b-a)^2}{8}\right)$ for any $t \in \mathbb{R}$.
First, observe that $\psi(0) = 1$. To bound $\psi(t)$, we use the convexity of the exponential function. For any $t \in \mathbb{R}$,

$$e^{t(X-E[X])} \leq \frac{b-X}{b-a} e^{t(a-E[X])} + \frac{X-a}{b-a} e^{t(b-E[X])}.$$

Taking the expectation on both sides, we get

$$\psi(t) = E[e^{t(X-E[X])}] \leq \frac{b-E[X]}{b-a} e^{t(a-E[X])} + \frac{E[X]-a}{b-a} e^{t(b-E[X])}.$$

Next, we apply Jensen's inequality to the convex function $f(x) = e^{tx}$:

$$\psi(t) \leq e^{tE[X-E[X]]} = e^0 = 1.$$

However, this is not enough to prove the lemma. We need a sharper bound. Using the fact that $e^x \leq 1 + x + \frac{x^2}{2} e^{|x|}$ for any $x \in \mathbb{R}$, we get

$$\psi(t) \leq 1 + tE[X - E[X]] + \frac{t^2}{2} E[(X - E[X])^2] e^{|t|(b-a)}.$$

Since $E[X - E[X]] = 0$ and $E[(X - E[X])^2] \leq (b-a)^2/4$, we obtain

$$\psi(t) \leq 1 + \frac{t^2(b-a)^2}{8} e^{|t|(b-a)}.$$

For small t, the exponential term $e^{|t|(b-a)} \approx 1$, so

$$\psi(t) \leq \exp\left(\frac{t^2(b-a)^2}{8}\right).$$

This completes the proof. □

Remarks: Hoeffding's Lemma is a key result that provides an exponential bound on the moment-generating function of a bounded random variable. This lemma is fundamental in deriving Hoeffding's inequality and other concentration inequalities, offering a way to control the tail behavior of sums of bounded independent random variables.

Theorem 3.2.6. ***Hoeffding's Inequality***
Let $X_1, X_2, \ldots, X_n$ be independent random variables such that X_i takes values in the interval $[a_i, b_i]$. Define the sum of these variables as $S_n = \sum_{i=1}^n X_i$ and let $\mu = E[S_n]$. Then for any $\epsilon > 0$,

$$\Pr(|S_n - \mu| \geq \epsilon) \leq \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^n (b_i - a_i)^2}\right),$$

Proof. Consider the moment-generating function of X_i. For any $\lambda > 0$, we have:

$$E\left[e^{\lambda(X_i - E[X_i])}\right] \leq \exp\left(\frac{\lambda^2(b_i - a_i)^2}{8}\right), \tag{3.7}$$

which follows from Hoeffding's lemma. Therefore, for the sum $S_n = \sum_{i=1}^n X_i$, we can write:

$$E\left[e^{\lambda(S_n-\mu)}\right] = \prod_{i=1}^n E\left[e^{\lambda(X_i - E[X_i])}\right] \leq \prod_{i=1}^n \exp\left(\frac{\lambda^2(b_i - a_i)^2}{8}\right) = \exp\left(\frac{\lambda^2}{8} \sum_{i=1}^n (b_i - a_i)^2\right). \tag{3.8}$$

Using Markov's inequality, we get:

$$\Pr(S_n - \mu \geq \epsilon) = \Pr\left(e^{\lambda(S_n-\mu)} \geq e^{\lambda\epsilon}\right) \leq \frac{E\left[e^{\lambda(S_n-\mu)}\right]}{e^{\lambda\epsilon}} \leq \exp\left(\frac{\lambda^2}{8}\sum_{i=1}^{n}(b_i - a_i)^2 - \lambda\epsilon\right). \quad (3.9)$$

By optimizing the bound over λ, we set $\lambda = \frac{4\epsilon}{\sum_{i=1}^{n}(b_i-a_i)^2}$ to minimize the exponent:

$$\Pr(S_n - \mu \geq \epsilon) \leq \exp\left(-\frac{2\epsilon^2}{\sum_{i=1}^{n}(b_i - a_i)^2}\right). \quad (3.10)$$

The bound for $\Pr(S_n - \mu \leq -\epsilon)$ follows by applying the same argument to $-X_i$.

□

Remarks: Hoeffding's inequality provides a powerful tool for bounding the probability that the sum of bounded independent random variables deviates significantly from its expected value. This inequality is particularly useful in scenarios where we need uniform bounds that hold regardless of the distribution of the individual variables, as long as they are within specified bounds.