Chernoff-Bounds

Wolfgang Mulzer

1 The General Bound

Let \( P = (p_1, \ldots, p_m) \) and \( Q = (q_1, \ldots, q_m) \) be two distributions on \( m \) elements, i.e., \( p_i, q_i \geq 0 \), for \( i = 1, \ldots, m \), and \( \sum_{i=1}^{m} p_i = \sum_{i=1}^{m} q_i = 1 \). The Kullback-Leibler divergence or relative entropy of \( P \) and \( Q \) is defined as

\[
D_{KL}(P \parallel Q) := \sum_{i=1}^{m} p_i \ln \frac{p_i}{q_i}.
\]

If \( m = 2 \), i.e., \( P = (p, 1-p) \) and \( Q = (q, 1-q) \), we also write \( D_{KL}(p \parallel q) \). The Kullback-Leibler divergence provides a measure of distance between the distributions \( P \) and \( Q \): it represents the expected loss of efficiency we incur if we encode an \( m \)-letter alphabet with distribution \( P \) with a code that is optimal for distribution \( Q \). We can now state the general form of the Chernoff-Bound:

**Theorem 1.1.** Let \( X_1, \ldots, X_n \) be independent random variables with \( X_i \in \{0, 1\} \) and \( \Pr[X_i = 1] = p \), for \( i = 1, \ldots, n \). Set \( X := \sum_{i=1}^{n} X_i \). Then, for any \( t \in [0, 1-p] \), we have

\[
\Pr[X \geq (p + t)n] \leq e^{-D_{KL}(p \parallel (p + t))n}.
\]

2 Three Proofs

2.1 The Moment Method

The usual proof of Theorem 1.1 uses the exponential function \( \exp \) and Markov’s inequality. It is called moment method because \( \exp \) simultaneously encodes all moments of \( X \), i.e., \( X, X^2, X^3, \) etc. The proof technique is very general and can be used to obtain several variants of Theorem 1.1. Let \( \lambda > 0 \) a parameter to be determined later. We have

\[
\Pr[X \geq (p + t)n] = \Pr[\lambda X \geq \lambda(p + t)n] = \Pr[e^{\lambda X} \geq e^{\lambda(p + t)n}].
\]

From Markov’s inequality, we obtain

\[
\Pr[e^{\lambda X} \geq e^{\lambda(p + t)n}] \leq \frac{\mathbb{E}[e^{\lambda X}]}{e^{\lambda(p + t)n}}.
\]

Now, the independence of the \( X_i \) yields

\[
\mathbb{E}[e^{\lambda X}] = \mathbb{E}\left[e^{\lambda \sum_{i=1}^{n} X_i}\right] = \mathbb{E}\left[\prod_{i=1}^{n} e^{\lambda X_i}\right] = \prod_{i=1}^{n} \mathbb{E}[e^{\lambda X_i}] = (pe^{\lambda} + 1 - p)^n.
\]

Thus,

\[
\Pr[X > (p + t)n] \leq \left(\frac{pe^{\lambda} + 1 - p}{e^{\lambda(p + t)}}\right)^n, \tag{1}
\]
for every $\lambda > 0$. Optimizing for $\lambda$ using calculus, we get that the right hand side is minimized if

$$e^\lambda = \frac{(1 - p)(p + t)}{p(1 - p - t)}.$$

Plugging this into (1), we get

$$\Pr[X > (p + t)n] \leq \left[ \left( \frac{p}{p + t} \right)^{p + t} \left( \frac{1 - p}{1 - p - t} \right)^{1 - p - t} \right]^n = e^{-D_{KL}(p+t\|p)n},$$

as desired.

### 2.2 Chvátal’s Method

Let $B(n, p)$ the random variable that gives the number of heads in $n$ independent Bernoulli trials with success probability $p$. It is well known that

$$\Pr[B(n, p) = l] = \binom{n}{l} p^l (1 - p)^{n-l},$$

for $l = 0, \ldots, n$. Thus, for any $\tau \geq 1$ and $k \geq pn$, we get

$$\Pr[B(n, p) \geq k] = \sum_{i=k}^{n} \binom{n}{i} p^i (1 - p)^{n-i} \leq \sum_{i=k}^{n} \binom{n}{i} p^i (1 - p)^{n-i} \tau^{-k} \leq \sum_{i=0}^{k-1} \binom{n}{i} p^i (1 - p)^{n-i} \tau^{-k} = \sum_{i=0}^{n} \binom{n}{i} p^i (1 - p)^{n-i} \tau^{-k}.$$

Thus, using the Binomial theorem, we obtain

$$\Pr[B(n, p) \geq k] \leq \sum_{i=0}^{n} \binom{n}{i} p^i (1 - p)^{n-i} \tau^{-k} = \tau^{-k} \sum_{i=0}^{n} \binom{n}{i} (p\tau)^i (1 - p)^{n-i} = \frac{(p\tau + 1 - p)^n}{\tau^k}.$$

If we write $k = (p + t)n$ and $\tau = e^\lambda$, we can conclude

$$\Pr[B(n, p) \geq (p + t)n] \leq \left( \frac{pe^\lambda + 1 - p}{e^{\lambda(p+t)}} \right)^n.$$

This is the same as (1), so we can complete the proof of Theorem 1.1 as in Section 2.1.

### 2.3 The Impagliazzo-Kabanets Method

Let $\lambda \in [0, 1]$ be a parameter to be chosen later. Let $I \subseteq \{1, \ldots, n\}$ be a random index set obtained by including each element $i \in \{1, \ldots, n\}$ with probability $\lambda$. We estimate the probability $\Pr[\prod_{i \in I} X_i = 1]$ in two different ways, where the probability is over the random choice of $X_1, \ldots, X_n$ and $I$.

On the one hand, using the union bound and independence, we have

$$\Pr[\prod_{i \in I} X_i = 1] \leq \sum_{S \subseteq \{1, \ldots, n\}} \Pr[I = S \land \prod_{i \in S} X_i = 1] = \sum_{S \subseteq \{1, \ldots, n\}} \Pr[I = S] \cdot \prod_{i \in S} \Pr[X_i = 1] = \sum_{S \subseteq \{1, \ldots, n\}} \lambda^{|S|} (1 - \lambda)^{|S| - |S|} \cdot p^{|S|} = \sum_{s=0}^{n} \binom{n}{s} (\lambda p)^s (1 - \lambda)^{n-s} = (\lambda p + 1 - \lambda)^n, \quad (2)$$
by the Binomial theorem. On the other hand, by the law of total probability,

$$\Pr\left[ \prod_{i \in I} X_i = 1 \right] = \Pr\left[ \prod_{i \in I} X_i = 1 \mid X \geq (p + t)n \right] \Pr[X \geq (p + t)n].$$

Now, fix $X_1, \ldots, X_n$ with $X \geq (p + t)n$. For the fixed choice of $X_1 = x_1, \ldots, X_n = x_n$, the probability $\Pr[\prod_{i \in I} x_i = 1]$ is exactly the probability that $I$ avoids all the $n - X$ indices $i$ where $x_i = 0$. Thus,

$$\Pr\left[ \prod_{i \in I} x_i = 1 \right] = (1 - \lambda)^{n - X} \geq (1 - \lambda)^{(1 - p - t)n}.$$

Since the bound holds uniformly for every choice of $x_1, \ldots, x_n$ with $X \geq (p + t)n$, we get

$$\Pr\left[ \prod_{i \in I} X_i = 1 \mid X \geq (p + t)n \right] \geq (1 - \lambda)^{(1 - p - t)n},$$

so

$$\Pr\left[ \prod_{i \in I} X_i = 1 \right] \geq (1 - \lambda)^{(1 - p - t)n} \Pr[X \geq (p + t)n].$$

Combining with (2),

$$\Pr[X \geq (p + t)n] \leq \left( \frac{\lambda p + 1 - \lambda}{(1 - \lambda)(1 - p - t)} \right)^n.$$ (3)

Using calculus, we get that the right hand side is minimized for $\lambda = t/(1 - p)(p + t)$ (note that $\lambda \leq 1$ for $t \leq 1 - p$). Plugging this into (3),

$$\Pr[X > (p + t)n] \leq \left[ \left( \frac{p}{p + t} \right)^{p + t} \left( \frac{1 - p}{1 - p - t} \right)^{1 - p - t} \right]^n = e^{-D_{KL}(p + t\|p)n},$$

as desired.

3 Useful Consequences

3.1 The Lower Tail

Corollary 3.1. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in \{0, 1\}$ and $\Pr[X_i = 1] = p$, for $i = 1, \ldots, n$. Set $X := \sum_{i=1}^n X_i$. Then, for any $t \in [0, p]$, we have

$$\Pr[X \leq (p - t)n] \leq e^{-D_{KL}(p-t\|p)n}.$$

Proof.

$$\Pr[X \leq (p - t)n] = \Pr[n - X \geq n - (p - t)n] = \Pr[X' \geq (1 - p + t)n],$$

where $X' = \sum_{i=1}^n X'_i$ with $X'_i \in \{0, 1\}$ independent random variables such that $\Pr[X'_i = 1] = 1 - p$. The result follows from $D_{KL}(1 - p + t\|1 - p) = D_{KL}(p - t\|p)$. □
3.2 Motwani-Raghavan version

Corollary 3.2. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in \{0, 1\}$ and $\Pr[X_i = 1] = p$, for $i = 1, \ldots, n$. Set $X := \sum_{i=1}^n X_i$ and $\mu = pn$. Then, for any $\delta \geq 0$, we have

$$\Pr[X \geq (1 + \delta)\mu] \leq \left(\frac{e^\delta}{(1 + \delta)^{1+\delta}}\right)^\mu,$$

and

$$\Pr[X \leq (1 - \delta)\mu] \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}}\right)^\mu.$$

Proof. Setting $t = \delta \mu/n$ in Theorem 1.1 yields

$$\Pr[X \geq (1 + \delta)\mu] \leq \exp\left(-n \left[p(1 + \delta)\ln(1 + \delta) + p\left(\frac{1 - p}{p} - \delta\right)\ln\left(1 - \delta \frac{p}{1 - p}\right)\right]\right) = \left(\frac{(1 - \delta p/(1 - p))^{\delta - (1 - p)/p}}{(1 + \delta)^{1+\delta}}\right)^\mu \leq \left(\frac{e^{-\delta^2 p/(1 - p)}}{(1 + \delta)^{1+\delta}}\right)^\mu \leq \left(\frac{e^{-\delta}}{(1 + \delta)^{1+\delta}}\right)^\mu.$$

Setting $t = \delta \mu/n$ in Corollary 3.1 yields

$$\Pr[X \leq (1 - \delta)\mu] \leq \exp\left(-n \left[-p(1 - \delta)\ln(1 - \delta) + p\left(\frac{1 - p}{p} + \delta\right)\ln\left(1 + \delta \frac{p}{1 - p}\right)\right]\right) = \left(\frac{(1 + \delta p/(1 - p))^\delta - (1 - p)/p}{(1 - \delta)^{1-\delta}}\right)^\mu \leq \left(\frac{e^{-\delta^2 p/(1 - p) - \delta}}{(1 - \delta)^{1-\delta}}\right)^\mu \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}}\right)^\mu.$$

3.3 Handy Versions

Corollary 3.3. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in \{0, 1\}$ and $\Pr[X_i = 1] = p$, for $i = 1, \ldots, n$. Set $X := \sum_{i=1}^n X_i$ and $\mu = pn$. Then, for any $\delta \in (0, 1)$, we have

$$\Pr[X \leq (1 - \delta)\mu] \leq e^{-\delta^2 \mu/2}.$$

Proof. By Corollary 3.2

$$\Pr[X \leq (1 - \delta)\mu] \leq \left(\frac{e^{-\delta}}{(1 - \delta)^{1-\delta}}\right)^\mu.$$

Using the power series expansion of $\ln(1 - \delta)$, we get

$$(1 - \delta)\ln(1 - \delta) = -(1 - \delta)\sum_{i=1}^\infty \frac{\delta^i}{i} = -\delta + \sum_{i=2}^\infty \frac{\delta^i}{(i - 1)i} \geq -\delta + \delta^2/2.$$

Thus,

$$\Pr[X \leq (1 - \delta)\mu] \leq e^{[1 - \delta + \delta^2/2]\mu} = e^{-\delta^2 \mu/2},$$

as claimed.
Corollary 3.4. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in \{0, 1\}$ and $\Pr[X_i = 1] = p$, for $i = 1, \ldots, n$. Set $X := \sum_{i=1}^n X_i$ and $\mu = pn$. Then, for any $\delta \geq 0$, we have
\[
\Pr[X \geq (1 + \delta)\mu] \leq e^{-\min\{\delta^2, \delta\} \mu / 4}.
\]

**Proof.** We may assume that $(1 + \delta)p \leq 1$. Then Theorem 1.1 gives
\[
\Pr[X \geq (1 + \delta)pn] \leq e^{-D_{\text{KL}}((1 + \delta)p \parallel p)n}.
\]
Define $f(\delta) := D_{\text{KL}}((1 + \delta)p \parallel p)$. Then
\[
f'(\delta) = p \ln(1 + \delta) - p \ln(1 - \delta p / (1 - p))
\]
and
\[
f''(\delta) = \frac{p}{(1 + \delta)(1 - p - \delta p)} \geq \frac{p}{1 + \delta}.
\]
By Taylor’s theorem, we have
\[
f(\delta) = f(0) + \delta f'(0) + \frac{\delta^2}{2} f''(\xi),
\]
for some $\xi \in [0, \delta]$. Since $f(0) = f'(0) = 0$, it follows that
\[
f(\delta) = \frac{\delta^2}{2} f''(\xi) \geq \frac{\delta^2 p}{2(1 + \xi)} \geq \frac{\delta^2 p}{2(1 + \delta)}.
\]
For $\delta \geq 1$, we have $\delta / (1 + \delta) \geq 1/2$, for $\delta < 1$, we have $1 / (\delta + 1) \geq 1/2$. This gives for all $\delta \geq 0$
\[
f(\delta) \geq \min\{\delta^2, \delta\} p / 4,
\]
and the claim follows. \hfill \square

Corollary 3.5. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in \{0, 1\}$ and $\Pr[X_i = 1] = p$, for $i = 1, \ldots, n$. Set $X := \sum_{i=1}^n X_i$ and $\mu = pn$. Then, for any $\delta > 0$, we have
\[
\Pr[|X - \mu| \geq \delta \mu] \leq 2e^{-\min\{\delta^2, \delta\} \mu / 4}.
\]

**Proof.** Combine Corollaries 3.3 and 3.4. \hfill \square

Corollary 3.6. Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in \{0, 1\}$ and $\Pr[X_i = 1] = p$, for $i = 1, \ldots, n$. Set $X := \sum_{i=1}^n X_i$ and $\mu = pn$. For $t \geq 2e\mu$, we have
\[
\Pr[X \geq t] \leq 2^{-t}.
\]

**Proof.** By Corollary 3.2
\[
\Pr[X \geq (1 + \delta)\mu] \leq \left( \frac{e^\delta}{(1 + \delta)^{1+\delta}} \right)^\mu \leq \left( \frac{e}{1 + \delta} \right)^{(1+\delta)\mu}.
\]
For $\delta \geq 2e - 1$, the denominator in the right hand side is at least $2e$, and the claim follows. \hfill \square

4 Generalizations

We mention a few generalizations of the three proof techniques for Section 2. Since the consequences from Section 3 are based on simple algebraic manipulation of the bounds, they same consequences also hold for the generalized settings.
4.1 Hoeffding-Extension

**Theorem 4.1.** Let $X_1, \ldots, X_n$ be independent random variables with $X_i \in [0,1]$ and $E[X_i] = p_i$. Set $X := \sum_{i=1}^n X_i$ and $p := (1/n) \sum_{i=1}^n p_i$. Then, for any $t \in [0, 1 - p]$, we have

$$\Pr[X \geq (p + t)n] \leq e^{-D_{KL}(p+t\parallel p)n}.$$ 

**Proof.** The proof generalizes the moment method. Let $\lambda > 0$ a parameter to be determined later. As before, Markov’s inequality yields

$$\Pr[e^{\lambda X} \geq e^{\lambda(p+t)n}] \leq \frac{E[e^{\lambda X}]}{e^{\lambda(p+t)n}}.$$ 

Using independence, we get

$$E[e^{\lambda X}] = E[e^{\lambda \sum_{i=1}^n X_i}] = \prod_{i=1}^n E[e^{\lambda X_i}]. \quad (4)$$

Now we need to estimate $E[e^{\lambda X_i}]$. The function $z \mapsto e^{\lambda z}$ is convex, so $e^{\lambda z} \leq (1 - z)e^{\lambda 0} + ze^{\lambda 1}$ for $z \in [0,1]$. Hence,

$$E[e^{\lambda X_i}] \leq E[1 - X_i + X_i e^{\lambda}] = 1 - p_i + p_i e^{\lambda}.$$ 

Going back to (4),

$$E[e^{\lambda X}] \leq \prod_{i=1}^n (1 - p_i + p_i e^{\lambda}).$$

Using the arithmetic-geometric mean inequality $\prod_{i=1}^n x_i \leq \left((1/n) \sum_{i=1}^n x_i\right)^n$, for $x_i \geq 0$, this is

$$E[e^{\lambda X}] \leq (1 - p + pe^{\lambda})^n.$$ 

From here we continue as in Section 2.1.$\square$

4.2 Hypergeometric Distribution

Chvátal’s proof generalizes to the hypergeometric distribution.

**Theorem 4.2.** Suppose we have an urn with $N$ balls, $P$ of which are red. We randomly draw $n$ balls from the urn without replacement. Let $H(N, P, n)$ denote the number of red balls in the sample. Set $p := P/N$. Then, for any $t \in [0, 1 - p]$, we have

$$\Pr[H(N, P, n) \geq (p + t)n] \leq e^{-D_{KL}(p+t\parallel p)n}.$$ 

**Proof.** It is well known that

$$\Pr[H(N, P, n) = l] = \binom{P}{l} \binom{N - P}{n - l} \binom{N}{l}^{-1},$$

for $l = 0, \ldots, n$.

**Claim 4.3.** For every $j \in \{0, \ldots, n\}$, we have

$$\binom{N}{n}^{-1} \sum_{i=j}^n \binom{P}{i} \binom{N - P}{n - i} \binom{i}{j} \leq \binom{n}{j} p^j.$$
Proof. Consider the following random experiment: take a random permutation of the \( N \) balls in the urn. Let \( S \) be the sequence of the first \( n \) elements in the permutation. Let \( X \) be the number of \( j \)-subsets of \( S \) that contain only red balls. We compute \( \mathbb{E}[X] \) in two different ways. On the one hand,

\[
\mathbb{E}[X] = \sum_{i=j}^{n} \Pr[S \text{ contains } i \text{ red balls}] \binom{i}{j} = \sum_{i=j}^{n} \binom{N}{i}^{-1} \binom{N-P}{n-i} \binom{i}{j}.
\]  

(5)

On the other hand, let \( I \subseteq \{1, \ldots, n\} \) with \( |I| = j \). Then the probability that all the balls in the positions indexed by \( I \) are red is

\[
\frac{P}{N} \cdot \frac{P-1}{N-1} \cdots \frac{P-j+1}{N-j+1} \leq \left( \frac{P}{N} \right)^j = p^j.
\]

Thus, by linearity of expectation \( \mathbb{E}[X] \leq \binom{n}{j} p^j \). Together with (5), the claim follows.  

\[\square\]

Claim 4.4. For every \( \tau \geq 1 \), we have

\[
\binom{N}{n}^{-1} \sum_{i=0}^{n} \binom{P}{i} \binom{N-P}{n-i} \tau^i \leq (1 + (\tau - 1)p)^n.
\]

Proof. Using Claim 4.3 and the Binomial theorem (twice),

\[
\binom{N}{n}^{-1} \sum_{i=0}^{n} \binom{P}{i} \binom{N-P}{n-i} \tau^i = \binom{N}{n}^{-1} \sum_{i=0}^{n} \binom{P}{i} \binom{N-P}{n-i} (1 - (\tau - 1))^i \\
= \binom{N}{n}^{-1} \sum_{i=0}^{n} \binom{P}{i} \binom{N-P}{n-i} \sum_{j=0}^{i} \binom{i}{j} (\tau - 1)^j \\
= \binom{N}{n}^{-1} \sum_{j=0}^{n} (\tau - 1)^j \sum_{i=j}^{n} \binom{P}{i} \binom{N-P}{n-i} \binom{i}{j} \\
\leq \sum_{j=0}^{n} \binom{n}{j} (\tau - 1)^j p^j = (1 + (\tau - 1)p)^n,
\]

as claimed.  

\[\square\]

Thus, for any \( \tau \geq 1 \) and \( k \geq pn \), we get as before

\[
\Pr[H(N, P, n) \geq k] = \binom{N}{n}^{-1} \sum_{i=k}^{n} \binom{P}{i} \binom{N-P}{n-i} \leq \binom{N}{n}^{-1} \sum_{i=0}^{n} \binom{P}{i} \binom{N-P}{n-i} \tau^{i-k} \leq \frac{(\tau r + 1 - p)^n}{\tau^k},
\]

by Claim 4.4. For here the proof proceeds as in Section 2.2.  

\[\square\]

4.3 General Impagliazzo-Kabanets

Theorem 4.5. Let \( X_1, \ldots, X_n \) be random variables with \( X_i \in [0, 1] \). Suppose there exist \( p_i \in [0, 1] \), \( i = 1, \ldots, n \), such that for every index set \( I \subseteq \{1, \ldots, n\} \), we have \( \Pr[\prod_{i \in I} X_i = 1] \leq \prod_{i \in I} p_i \). Set \( X := \sum_{i=1}^{n} X_i \) and \( p := (1/n) \sum_{i=1}^{n} p_i \). Then, for any \( t \in [0, 1 - p] \), we have

\[
\Pr[X \geq (p + t)n] \leq e^{-D_{KL}(p+t||p)n}.
\]
Proof. Let $\lambda \in [0,1]$ be a parameter to be chosen later. Let $I \subseteq \{1,\ldots,n\}$ be a random index set obtained by including each element $i \in \{1,\ldots,n\}$ with probability $\lambda$. As before, we estimate the probability $\Pr[\prod_{i \in I} X_i = 1]$ in two different ways, where the probability is over the random choice of $X_1,\ldots,X_n$ and $I$. Similarly to before,

\[
\Pr[\prod_{i \in I} X_i = 1] = \Pr[\prod_{i \in I} X_i = 1] \leq \sum_{S \subseteq \{1,\ldots,n\}} \Pr[I = S \land \prod_{i \in S} X_i = 1] 
\leq \sum_{S \subseteq \{1,\ldots,n\}} \Pr[I = S] \cdot \Pr[\prod_{i \in S} X_i = 1] \leq \sum_{S \subseteq \{1,\ldots,n\}} \lambda^{|S|}(1 - \lambda)^{n - |S|} \cdot \prod_{i \in S} p_i. \tag{6}
\]

We define $n$ independent random variables $Z_1,\ldots,Z_n$ as follows: for $i = 1,\ldots,n$, with probability $1 - \lambda$, we set $Z_i = 1$, and with probability $\lambda$, we set $Z_i = p_i$. By (6), and using independence and the arithmetic-geometric mean inequality,

\[
\Pr[\prod_{i \in I} X_i = 1] = \mathbb{E}[\prod_{i = 1}^n Z_i] = \prod_{i = 1}^n \mathbb{E}[Z_i] = \prod_{i = 1}^n (1 - \lambda + p_i \lambda) \leq (1 - \lambda + p\lambda)^n. \tag{7}
\]

The proof of the lower bound remains unchanged and yields

\[
\Pr[\prod_{i \in I} X_i = 1] \geq (1 - \lambda)^{(1-p-t)n} \Pr[X \geq (p + t)n],
\]

as before. Combining with (7) and optimizing for $\lambda$ finishes the proof, see Section 2.3. \qed