Applying Hoeffding's ineq. for $-X_i$ (which are also independent and symmetric Bernoulli), we obtain

$$ P \left\{ \frac{\sum_{i=1}^{n} a_i X_i}{\sqrt{n}} > t \right\} \leq e^{-t^2} \hspace{1cm} \text{(1)} $$

Combining both inequalities, we get

$$ P \left\{ \left| \frac{\sum_{i=1}^{n} a_i X_i}{\sqrt{n}} \right| > t \right\} \leq 2e^{-t^2/2} \hspace{1cm} \text{(2)} $$

2. Assumption $\sum a_i^2 = 1$ can be removed by rescaling $\left( a_i \rightarrow \frac{a_i}{\sum a_i^2} \right)$

$$ \forall a_i \hspace{1cm} P \left\{ \left| \frac{\sum_{i=1}^{n} a_i X_i}{\sqrt{n}} \right| > t \right\} \leq 2 \exp \left( - \frac{t^2}{2 \sum a_i^2} \right) \hspace{1cm} \text{(3)} $$

3. For non-symmetric Bernoulli ($P(X_i = 1) = p_i$, $P(X_i = -1) = 1 - p_i$), a sharp result is Chernoff's inequality (with a similar argument).

4. The method of proving Hoeffding's inequality (due to S. Bernstein) can be applied for more general distributions. In particular, one can prove:

**General Hoeffding's inequality**

Let $X_i$ be independent r.v's such that $E X_i = 0$, $-c_i \leq X_i \leq c_i$ a.s. for all $i$.

Then

$$ P \left\{ \left| \frac{\sum_{i=1}^{n} X_i}{\sqrt{\sum_{i=1}^{n} c_i^2}} \right| > t \right\} \leq 2 \exp \left( - \frac{t^2}{\sum_{i=1}^{n} c_i^2} \right), \hspace{1cm} t > 0 \hspace{1cm} \text{(4)} $$

Ineq. (4) is a special case of this theorem for $a_i X_i$ (since $-a_i \leq a_i X_i \leq a_i$).

Other deviation inequalities in the same spirit:

- Bernstein's inequality - see later
- Bennett's inequality

References: [V. Petrov, Sums of independent random variables]
What is the largest class of r.v.'s for which Hoeffding-type inequalities hold, i.e. the sums have "sub-gaussian" tails:
\[ \Pr \left\{ \sum_{i=1}^{n} X_i > t \right\} \leq e^{-t^2/n} \]  
(\text{**})

A necessary condition is that \( X_i \) themselves have "sub-gaussian" tails:
\[ \Pr \left\{ X_i > t \right\} \leq e^{-ct^2}, \quad t > 0 \]  
(\text{**})

(One can see this by setting \( n = 1 \) in (\text{**})).

(\text{**}) is also a sufficient condition.

One can show this by the same Bernstein's argument.

This deserves a special attention.

We will study "sub-gaussian" r.v.'s (\text{**}) in some detail now.

1.3. Sub-gaussian random variables

- Reference for this section: [R. Vershynin, Introduction to non-asymptotic...]

Start with Gaussian.

Three basic properties of \( g \sim \mathcal{N}(0,1) \), which describe how "large" \( g \) is:

1. Tails: \( \Pr \{ |g| > t \} \leq 2e^{-t^2/2}, \quad t > 1 \)

2. Moments: \( (\mathbb{E} (g^{2p})^{1/2})^{1/p} \leq \frac{\Gamma((p+1)/2)}{\Gamma(1/2)} \) \( = O(\sqrt{p}) \), \( p \geq 2 \)

Recall basic facts about moments of r.v. \( X \):

- \( \|X\|_p := (\mathbb{E} |X|^p)^{1/p} \) is a norm; it defines \( L_p (\Omega, \mathbb{P}) \) space.
- \( \|X\|_p \leq C \|X\|_q \) when \( p \leq q \) (i.e. \( \|X\|_p \) increases in \( p \)).
- \( \|X\|_\infty := \text{ess. sup} |X| \), i.e. \( M = \|X\|_\infty \) iff \( |X| \leq M \) a.s.

3. M.G.F.: \( M_g(t) = \mathbb{E} e^{tg} = e^{t^2/2}, \quad t \in \mathbb{R} \)
Now let \( X \) be a general r.v.; all three properties are equivalent:

- **Sub-Gaussian property**

Let \( X \) be a r.v. Then the following properties are equivalent:

1. **Tails:** \( P\{|X| > t\} \leq \exp\left(1 - t^2/K_i^2\right) \) for all \( t > 0 \)

2. **Moments:** \( (E|X|^p)^{1/p} \leq K_i \sqrt{p} \) for all \( p > 1 \)

3. **M.G.F.:** \( \exp(\lambda X) \leq \exp(K_i^2 \lambda^2) \) for all \( \lambda \in \mathbb{R} \)

Moreover, if \( EX = 0 \) then (1) and (2) are equivalent to:

4. **M.G.F.:** \( \exp(\lambda^2) \leq \exp(\lambda^2 K_i^2) \) for all \( \lambda > 0 \)

The precise meaning of the equivalence: \( \exists \) absolute const \( C \) such that property (i) implies property (j) with \( K_j \leq CK_i \), for any two \( i, j \in \{1, 2, 3\} \)

**Proof.** (1) \( \iff \) (2). Assume (1) holds. W.l.o.g. \( K_1 = 1 \) (by rescaling \( X \rightarrow X/K_i \)).

Recall the identity for expected value for non-negative r.v.'s \( Z \):

\[
E Z = \int_0^\infty P\{Z > u\} \, du
\]

(proved by integration by parts)

Apply the identity for \( Z = |X|^p \).

\[
E |X|^p = \int_0^\infty P\{|X|^p > u\} \, du.
\]

\[
= \int_0^\infty P\{|X| > t^{1/p}\} \, dt^{p-1} \, dt \quad \text{(change of var. } u = t^p)\]

\[
\leq \int_0^\infty e^{-t^2} \, dt^{p-1} \quad \text{by prop.(i)}
\]

\[
\leq \frac{\Gamma\left(\frac{p}{2}\right)}{\sqrt{\pi}} \quad \text{(change of var. } t^2 = s \quad \text{and def of Gamma function)}
\]

\[
\leq \frac{\Gamma\left(\frac{p}{2}\right)^{\frac{1}{2}}}{\sqrt{\pi}} \quad \text{by Stirling's formula)}
\]

Take \( p \)-th root \( \Rightarrow \) Property (2).

-8-
(2) ⇒ (3). Assume (2) holds; W.L.O.G. \( K_2 = 1 \).

Taylor series:

\[
E \exp \left( c X^2 \right) = E \left[ 1 + \sum_{p=1}^{\infty} \frac{\left( c X^2 \right)^p}{p!} \right] = 1 + \sum_{p=1}^{\infty} \frac{c^p E \left( X^{2p} \right)}{p!}
\]

\[\leq 1 + \sum_{p=1}^{\infty} \frac{(2cp)^p}{(p/e)^p} \quad \text{by prop. (2)}\]

\[= 1 + \sum_{p=1}^{\infty} (2ce)^p.\]

Choose \( c = \frac{1}{4e} \) ⇒ \( E \exp \left( c X^2 \right) \leq 1 + \sum_{p=1}^{\infty} \left( \frac{1}{2} \right)^p \leq 2.\)

⇒ property (3)

(3) ⇒ (1). Assume (3) holds; W.L.O.G. \( K_2 = 1 \). "Exponential Chebyshev".

\[
P \{ |X| \geq t \} = P \left\{ \exp \left( X^2 \right) \geq \exp \left( t^2 \right) \right\}
\]

\[\leq \frac{E \exp \left( X^2 \right)}{\exp(t^2)} \quad \text{by prop. (3)}\]

\[\leq \exp \left( 1 - t^2 \right).\]

⇒ Property (1)

\[\text{QED.}\]

HW: Prove the second part of Lemma, for Property (4).

A proof is actually given in [R. Vershynin, Introduction to the non-asymptotic...]

but is a bit lengthy.

Can you find a better proof?: (Shorter, more elegant)?