**Chebyshev's association inequality**

Let \( f, g : \mathbb{R} \rightarrow \mathbb{R} \) be increasing functions. If \( X \) is a random variable, then

\[
E[f(X)g(X)] \geq E[f(X)] \cdot E[g(X)]
\]

**Proof**

Let \( Y \) be distributed as \( X \) but independent of \( X \). Then

\[
E[(X - E[X])(g(X) - E[g(X)])] \geq 0
\]

Take expected values.

**Harris' inequality**

Let \( f, g : \mathbb{R} \rightarrow \mathbb{R} \) be increasing functions (in all variables). Let \( X = (X_1, \ldots, X_n) \) where \( X_1, \ldots, X_n \) are independent. Then

\[
E[f(X)g(X)] \geq E[f(X)] \cdot E[g(X)]
\]

**Proof**

By induction. \( n = 1 \) is Chebyshev's association inequality. Suppose the theorem is true for functions of \( n - 1 \) variables. Then

\[
E[f(X)g(X)] = E[E[f(X)g(X)|X_1, \ldots, X_n]]
\]

\[
\geq E[E[f(X)|X_1, \ldots, X_{n-1}]] \cdot E[g(X)|X_1, \ldots, X_{n-1}]
\]

By Chebyshev's inequality.

Because \( f, g \) are increasing in the \( n \)-th variable.

Now define \( f(X_1, \ldots, X_{n-1}) = E[f(X)|X_1, \ldots, X_{n-1}] \) and \( g \) similarly.

By independence \( f, g \) are increasing and the result follows by the induction hypothesis.
Concentration inequalities

We start with Markov's inequality: if $X \geq 0$, then for all $t > 0$,
\[ \mathbb{P}(X \geq t) \leq \frac{EX}{t} \]

This implies Chebyshev's inequality: for any random variable $X$ and for $t > 0$,
\[ \mathbb{P}(|X - \mathbb{E}X| > t) = \mathbb{P}((X - \mathbb{E}X)^2 > t^2) \leq \frac{\mathbb{E}(X - \mathbb{E}X)^2}{t^2} = \frac{\text{var}(X)}{t^2} \]

**Exercise:** Prove the Chebyshev-Cantelli inequality:
\[ \mathbb{P}(X - \mathbb{E}X > t) \leq \frac{\text{var}(X)}{\text{var}(X) + t^2} \]

(always $\leq 1$!)

In particular, if $X_1, \ldots, X_n$ are i.i.d. with $\text{var}(X) = \sigma^2$, then
\[ \mathbb{P}(\sqrt{n}(|\bar{X} - \mathbb{E}X| > t) \leq \frac{\text{var}(\bar{X})}{t^2} = \frac{n\sigma^2}{t^2} \]

Under additional conditions, much better bounds are available. Recall that by the Central Limit Theorem,
\[ \mathbb{P}\left( \frac{1}{\sqrt{n}} \sum_{i=1}^{n}(X_i - \mathbb{E}X) > t \right) \to \mathbb{P}(N > t) \leq e^{-\frac{t^2}{2\sigma^2}} \]

so one expects sub-Gaussian tail bounds. Such bounds are themselves bounded:
\[ \mathbb{P}(X > t) = \mathbb{P}(e^{sX} > e^{st}) \leq \frac{e^{sX}}{e^{st}} \quad \text{this bound can be minimized in } s! \]
\[ X = \frac{1}{n} \sum_{i=1}^{n} X_i, \quad E e^{sX} = E \prod_{i=1}^{n} e^{sX_i} = (E e^{sX})^n \]

So Chernoff bounds are especially useful for sums of independent random variables.

For example, if \( X \in [a, b] \), then for all \( s \in \mathbb{R} \),
\[ E e^{sX} \leq e^{\frac{b^2}{a} s^2} \leq \text{Exercise!} \]

\( \Rightarrow \) Hoeffding's lemma.

Which implies Hoeffding's inequality: if \( X_1, \ldots, X_n \) are independent \( \in \mathbb{R} \), then for all \( t > 0 \),
\[ \Pr(\sum_{i=1}^{n}(X_i - E X_i) \geq t) \leq e^{-\frac{2t^2}{\sum_{i=1}^{n} Var(X_i)}} \]

And another version is Bernstein's inequality: if \( X_1, \ldots, X_n \) are i.i.d. such that \( X_i \leq 1 \) and \( E X_i = 0 \), then
\[ \Pr(\sum_{i=1}^{n} X_i > t) \leq e^{-\frac{t^2}{2\sum_{i=1}^{n} Var(X_i)}} \]

Johnson–Lindenstrauss lemma. A beautiful application of the probabilistic method: let \( a_1, \ldots, a_n \in \mathbb{R}^D \) and \( \varepsilon > 0 \).

For what values of \( D < C \) does there exist a function \( X : \mathbb{R}^D \rightarrow \mathbb{R}^n \) such that \( \forall i, j = 1, \ldots, n \),
\[ (1 - \varepsilon) \|a_i - a_j\|^2 \leq \|X(a_i) - X(a_j)\|^2 \leq (1 + \varepsilon) \|a_i - a_j\|^2 \]

Such a function is an \( \varepsilon \)-isometry. We show that such an embedding exists whenever \( D > \frac{100}{\varepsilon^2} \log n \), independent of \( n \).
We prove that in fact there exist linear embeddings with the desired property. Let

$$W : \mathbb{R}^d \to \mathbb{R}^d$$

be given by its matrix \( W = (W_{ij})_{d \times d} \) such that the \( W_{ij} \) are \( \mathcal{N}(0, \frac{1}{n}) \) random variables.

Then for any \( a \in \mathbb{R}^d \),

$$E \| W(a) \|^2 = \sum_{i=1}^{d} E \left( \sum_{j=1}^{d} a_j W_{ij} \right)^2 = \| a \|^2$$

Thus, we need to show that, with positive probability,

$$\max_{i,j=1,\ldots,n} \left\| \frac{W_i(a_j) - W_i(a_j)^*}{\| a_i - a_j \|^2} \right\|^2 - 1 < \epsilon$$

\[ \| W(a) \|^2 - 1 = \| W(a) \|^2 - 1 \]

is a sum of independent zero-mean random variables.

Chernoff bound and the union bound give the desired result. Exercise: work out the details.