INEQUALITIES

Chebyshev's association inequality

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

$$\mathbb{E}[f(X)g(X)] \geq \mathbb{E}[f(X)] \cdot \mathbb{E}[g(X)]$$

Proof Let X' be distributed as X but independent of it. Then $(f(X) - f(X')) \cdot (g(X) - g(X')) \geq 0$.

Take expected values. $\boxed{\text{m}}$

Harris' inequality Let $f, g : \mathbb{R}^n \to \mathbb{R}$ be increasing functions (in all variables). Let $X = (X_1, \dots, X_n)$ where $X_1, \dots, X_n$ are independent. Then

$$\mathbb{E}[f(X) \cdot g(X)] \geq \mathbb{E}[f(X)] \cdot \mathbb{E}[g(X)].$$

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

$$\mathbb{E}[f(X) \cdot g(X)] = \mathbb{E}\,\mathbb{E}[f(X) g(X) \mid X_1, \dots, X_{n-1}]$$

$$\geq \mathbb{E}\left[\mathbb{E}[f(X) \mid X_1, \dots, X_{n-1}] \cdot \mathbb{E}[g(X) \mid X_1, \dots, X_{n-1}]\right]$$

↑ By Chebyshev's inequality because f, g are increasing in the n-th variable.

Now define $f'(X_1, \dots, X_{n-1}) = \mathbb{E}[f(X) \mid X_1, \dots, X_{n-1}]$ and g' similarly. By independence, f', g' are increasing and the result follows by the induction hypothesis. $\boxed{\text{m}}$

Concentration inequalities

We start with Markov's inequality: if $X \geq 0$, then for all $t > 0$,

$$\mathbb{P}(X \geq t) \leq \frac{\mathbb{E}X}{t}$$

This implies Chebyshev's inequality: for any random variable X and for $t > 0$,

$$\mathbb{P}(|X - \mathbb{E}X| \geq t) = \mathbb{P}\left((X-\mathbb{E}X)^2 \geq t^2\right) \leq \frac{\mathbb{E}(X-\mathbb{E}X)^2}{t^2} = \frac{\mathrm{var}(X)}{t^2}$$

Exercise: Prove the Chebyshev–Cantelli inequality:

$$\mathbb{P}(X - \mathbb{E}X \geq t) \leq \frac{\mathrm{var}(X)}{\mathrm{var}(X) + t^2}$$

(always ≤ 1!)

In particular, if $X_1, \ldots, X_n$ are i.i.d. with $\mathrm{var}(X_i) = \sigma^2$, then

$$\mathbb{P}\left(\left|\sum_{i=1}^n X_i - n\mathbb{E}X\right| > t\right) \leq \frac{\mathrm{var}\left(\sum_i X_i\right)}{t^2} = \frac{n\sigma^2}{t^2}$$

Under additional conditions, much better bounds are available. Recall that by the Central Limit Theorem,

$$\mathbb{P}\left(\frac{1}{\sqrt{n}} \sum_{i=1}^n (X_i - \mathbb{E}X) > t\right) \xrightarrow{n\to\infty} \mathbb{P}(N > t) \leq e^{-\frac{t^2}{2\sigma^2}}$$

standard normal — for $t > 0$

so one expects sub-Gaussian tail bounds. Such bounds are "Chernoff bounds":

$$\mathbb{P}(X > t) \underset{\text{if } s>0}{=} \mathbb{P}\left(e^{sX} > e^{st}\right) \leq \frac{\mathbb{E}\left[e^{sX}\right]}{e^{st}}$$

this bound can be minimized in s!

If $X = \sum_{i=1}^n X_i$, $\mathbb{E} e^{s \sum_i X_i} \underset{\text{i.i.d.}}{=} \mathbb{E} \prod_{i=1}^n e^{sX_i} = \left(\mathbb{E} e^{sX_1}\right)^n$

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

For example, if $X \in [a,b]$, then for all $s \in \mathbb{R}$,

$$\mathbb{E}\left[e^{sX}\right] \leq e^{s^2/8} \Leftarrow \text{Exercise!}$$

Hoeffding's lemma.

Which implies Hoeffding's inequality: if $X_1, \dots, X_n$ are independent $\in [0,1]$, then for all $t > 0$,

$$\mathbb{P}\left(\sum_{i=1}^n (X_i - \mathbb{E}X_i) > t\right) \leq e^{-2t^2/n}$$

Another version is Bernstein's inequality: if $X_1, \dots, X_n$ are i.i.d. such that $X_i \leq 1$ and $\mathbb{E}X_i = 0$, then

$$\mathbb{P}\left(\sum_{i=1}^n X_i > t\right) \leq e^{-t^2/(2n\sigma^2 + 2t/3)}$$

Johnson–Lindenstrauss lemma. A beautiful application of the probabilistic method: let $a_1, \dots, a_n \in \mathbb{R}^D$ and $\varepsilon > 0$. For what values of $d < D$ does there exist a function $f: \mathbb{R}^D \to \mathbb{R}^d$ such that $\forall\, i,j = 1, \dots, n$,

$$(1-\varepsilon)\|a_i - a_j\|^2 \leq \|f(a_i) - f(a_j)\|^2 \leq (1+\varepsilon)\|a_i - a_j\|^2 \ ?$$

Such a function is an ε-isometry. We show that such an embedding exists whenever $d \geq \frac{100}{\varepsilon^2} \log n$

independent of D !!!

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 $W = (W_{i,j})_{d \times D}$ such that the $W_{i,j}$ are $N(0, \frac{1}{d})$ random variables.

Then for any $a \in \mathbb{R}^D$,

$$\mathbb{E}\|W(a)\|^2 = \sum_{i=1}^{d} \underbrace{\mathbb{E}\left(\sum_{j=1}^{D} a_j W_{i,j}\right)^2}_{= \sum_{j=1}^{D} \frac{1}{d} a_j^2} = \|a\|^2$$

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

$$\max_{i,j = 1,\dots,n} \underbrace{\left| \left\| \frac{W(a_i) - W(a_j)}{\|a_i - a_j\|} \right\|^2 - 1 \right|}_{\left\| W\left(\frac{a_i - a_j}{\|a_i - a_j\|}\right) \right\|^2 - 1 = \|W(b_{i,j})\|^2 - 1} < \varepsilon$$

unit vector

$$\|W(b)\|^2 - 1 = \sum_{i=1}^{d} \Big(\underbrace{\sum_{j=1}^{D} W_{i,j} b_j}_{N(0, \frac{1}{d})}\Big)^2 - 1$$ is a sum of independent zero-mean random variables.

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