

# Concentração de Medida

Thiago Rodrigo Ramos

26 de dezembro de 2024

## Sumário

<b>1</b>	<b>Introdução às desigualdades de concentração</b>	<b>2</b>
1.1	Markov e seus amigos . . . . .	2
1.2	Variáveis sub-gaussianas . . . . .	4
1.3	Desigualdade de Hoeffding . . . . .	8
<b>2</b>	<b>Estimando a variância</b>	<b>10</b>
2.1	Efron-Stein . . . . .	10
2.2	Diferenças limitadas . . . . .	12
<b>A</b>	<b>Desigualdades básicas</b>	<b>13</b>
A.1	Fatorial . . . . .	13
<b>B</b>	<b>Convexidade</b>	<b>13</b>
B.1	Jensen . . . . .	13

## 1 Introdução às desigualdades de concentração

Baseado em [Boucheron et al., 2013].

### 1.1 Markov e seus amigos

**Teorema 1** (Desigualdade de Markov). *Seja  $X$  uma variável aleatória não-negativa. Então*

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

*Demonstração.* Usando o fato que  $X \geq 0$ , temos que:

$$\begin{aligned} \mathbb{E}[X] &= \mathbb{E}[X \cdot \mathbb{I}[X \geq t]] + \mathbb{E}[X \cdot \mathbb{I}[X < t]] \\ &\geq t \cdot \mathbb{P}(X \geq t) + 0. \end{aligned}$$

□

Note que dado qualquer função não-decrescente e positiva  $\phi$ , então pela desigualdade de Markov,

$$\mathbb{P}(X \geq t) \leq \mathbb{P}(\phi(X) \geq \phi(t)) \leq \frac{\mathbb{E}[\phi(X)]}{\phi(t)}.$$

De fato, considerando  $\phi : x \mapsto x^2$ , conseguimos o seguinte resultado.

**Corolário 1** (Desigualdade de Chebyshev). *Dado uma variável aleatória qualquer  $X$ , temos que*

$$\mathbb{P}(|X - \mathbb{E}[X]| \geq t) \leq \frac{\text{Var}[X]}{t^2}.$$

**Exemplo 1.** Sejam  $\{X_i\}_{i=1}^n$  v.a. i.i.d. com média  $\mu$  e variância  $\sigma^2$ . Então para qualquer  $t > 0$ , por Chebyshev:

$$\mathbb{P} \left( \frac{1}{n} \left| \sum_{i=1}^n X_i - \mu \right| \geq t \right) \leq \frac{n\sigma^2}{n^2 t^2} \rightarrow_n 0,$$

provando a uma versão da lei fraca dos grandes números.

Uma outra escolha natural para aplicarmos Markov é a função  $\phi : x \mapsto e^{\lambda x}$  para algum  $\lambda > 0$ .

**Corolário 2** (Desigualdade de Chernoff). Dado uma variável aleatória qualquer  $X$ , temos que

$$\mathbb{P}(X \geq t) \leq e^{-\lambda t} \mathbb{E}[e^{\lambda X}]. \quad (1)$$

**TODO:** Falar que o ótimo não é atingido por Chernoff, mas sim por algum momento (ex. 2.5).

Note que o que aparece do lado direito na desigualdade de Chernoff é algo que depende da função geradora de momento de  $X$ , dada por  $\mathbb{E}[e^{\lambda X}]$ . Por questões que ficarão mais claras no futuro, vamos reescrever a expressão 1 da seguinte forma:

$$\mathbb{P}(X \geq t) \leq e^{-\lambda t} \mathbb{E}[e^{\lambda X}] = e^{-(\lambda t - \psi_X(\lambda))},$$

onde  $\psi_X(\lambda) = \log(\mathbb{E}[e^{\lambda X}])$ .

Perceba que a conclusão da desigualdade de Chernoff vale para qualquer  $\lambda > 0$ , portanto podemos minimizar o lado direito em  $\lambda$  e assim temos o seguinte

$$\mathbb{P}(X \geq t) \leq e^{-\psi_X^*(t)}, \quad (2)$$

onde

$$\psi_X^*(t) = \sup_{\lambda \geq 0} (\lambda t - \psi_X(\lambda)).$$

Além disso, como  $\psi_X(0) = 0$  e em  $\psi_X^*(t)$  tomamos o supremo sobre os  $\lambda \geq 0$ , é claro que  $\psi_X^*(t) \geq 0$ . A função  $\psi_X^*$  é conhecida como transformada de Cramér de  $X$  e mais geralmente se o supremo é sobre todos os valores de  $\lambda$ , é conhecida como Função Dual Fenchel-Legendre de  $\psi_X$ .

**TODO:** Explicar como generalizar para  $\lambda$  qualquer e falar da diferenciabilidade e convexidade.

**Exemplo 2** (Distribuição Normal). Se  $X \sim N(0, \sigma^2)$ , é fácil mostrar que  $M_X(\lambda) = e^{\lambda^2 \sigma^2 / 2}$ , logo  $\psi_X(\lambda) = \frac{\lambda^2 \sigma^2}{2}$  e

$$\psi_X^*(t) = \frac{t^2}{2\sigma^2}.$$

Logo, por Eq. 2, temos que

$$\mathbb{P}(X \geq t) \leq e^{-t^2/(2\sigma^2)}.$$

De fato, esse limitante é o melhor possível a menos de uma constante. **TODO: mostrar isso.**

**Exemplo 3** (Soma de i.i.d.). Sejam  $\{X_i\}_{i=1}^n$  v.a. i.i.d. e defina  $S_n = \sum_{i=1}^n X_i$ . Então,  $\mathbb{E}[e^{\lambda S_n}] = \prod_{i=1}^n \mathbb{E}[e^{\lambda X_i}] = M_X(\lambda)^n$  e portanto  $\psi_{S_n}(\lambda) = n\psi_X(\lambda)$  e dessa forma

$$\psi_{S_n}^*(t) = \sup_{\lambda > 0} (\lambda t - n\psi_X(\lambda)) = n \sup_{\lambda > 0} (\lambda \frac{t}{n} - \psi_X(\lambda)) = n\psi_X^*(t/n).$$

Por exemplo, se  $X_i \sim N(0, 1)$  então  $\psi_{S_n}^* = n\psi_X^*(t/n) = n \frac{t^2}{2n^2}$  e então

$$\mathbb{P}(S_n \geq t) \leq e^{-t^2/(2n)}.$$

Podemos também considerar a probabilidade da soma ser maior que  $nt$  e concluir que

$$\mathbb{P}\left(\frac{S_n}{n} \geq t\right) \leq e^{-nt^2/2}.$$

## 1.2 Variáveis sub-gaussianas

Dizemos que uma v.a.  $X$  com  $\mathbb{E}[X] = 0$  é sub-gaussiana se sua função geradora de momento é limitada pela função geradora de momento de uma v.a. normal com variância  $v^2$ , ou seja,

$$\psi_X(\lambda) \leq \frac{\lambda^2 v^2}{2}, \quad \lambda \in \mathbb{R}. \quad (3)$$

**TODO: falar que precisa ser centrada**

Note que por Chernoff, se  $X$  é centrada e subgaussiana, então

$$\mathbb{P}(X \geq t) \leq e^{-x^2/(2v^2)},$$

já que essa se comporta basicamente como uma normal com variância  $v^2$  (Exemplo 1.1). A partir desse fato, conseguimos caracterizar as variáveis sub-gaussianas de outras formas.

**Teorema 2.** Seja  $X$  uma v.a. com  $\mathbb{E}[X] = 0$  e suponha que para algum  $v^2$ ,

$$\max\{\mathbb{P}(X > t), \mathbb{P}(-X > t)\} \leq e^{-x^2/(2v^2)},$$

então para qualquer  $q \geq 1$ ,

$$\mathbb{E}[X^{2q}] \leq 2q!(2v^2)^q \leq q!(4v^2)^q.$$

Reciprocamente, se existir constante  $C$  tal que

$$\mathbb{E} [X^{2q}] \leq q!C^q,$$

então  $X$  é sub-gaussiana com  $v^2 = 4C$ .

**Observação 1** (Momentos de uma Normal e fatorial duplo). Seja  $X$  uma v.a. com distribuição Normal de média 0 e variância  $\sigma^2$ . Então sabemos que

$$\mathbb{E} [X^{2q}] = \sigma^{2q}(2q-1)!!,$$

onde  $n!!$  é o fatorial duplo, isto é,  $n!!$  é o produto de todos os números menores que  $n$  que possuem a mesma paridade.

Note que

$$n!! = \frac{(n+1)!}{(n+1)!!},$$

e que

$$(2q)!! = \prod_{i=0}^{q-1} (2q-2i) = 2^q \prod_{i=0}^{q-1} (q-i) = 2^q q!.$$

Portanto,

$$(2q-1)!! = \frac{(2q)!}{(2q)!!} = \frac{(2q)!}{2^q q!}.$$

Dessa forma, temos o seguinte

$$\begin{aligned} \mathbb{E} [X^{2q}] &= \sigma^{2q}(2q-1)!! \\ &= \sigma^{2q} \frac{(2q)!}{2^q q!}. \end{aligned}$$

Agora note que,

$$\frac{(2q)!}{q!} = \prod_{i=1}^q (q+i) \geq \prod_{i=1}^q (i+i) = 2^q q!.$$

E portanto,

$$\begin{aligned} \mathbb{E} [X^{2q}] &= \sigma^{2q}(2q-1)!! \\ &= \sigma^{2q} \frac{(2q)!}{2^q q!} \\ &\geq \sigma^{2q} \frac{1}{2^q} 2^q q! \\ &= (\sigma^2)^q q!. \end{aligned}$$

Logo, o Teorema 2 nos diz é que podemos caracterizar uma v.a. sub-gaussiana comparando o seu limitate do tipo Chernoff com o de uma Normal, ou comparando os seus momentos pares com o de uma normal.

*Prova do Teorema 2.* O argumento que vamos utilizar é bem comum nesse tipo de prova. Vamos assumir por simplicidade que  $v = 1$ . Note que

$$\mathbb{E} [X^{2q}] = 2q \int_0^\infty x^{2q-1} \mathbb{P}(|X| > x) dx,$$

fazendo a mudança de variável  $u^{2q} = x$ . Dessa forma

$$\mathbb{E} [X^{2q}] \leq 4q \int_0^\infty x^{2q-1} e^{-x^2/2} dx,$$

onde um fator de 2 aparece pelo union bound de  $\{|X| > t\} = \{X > t\} \cup \{-X > t\}$ .

Fazendo a mudança de variável  $2u = x^2$ , temos que

$$\begin{aligned} \mathbb{E} [X^{2q}] &\leq 4q \int_0^\infty x^{2q-1} e^{-x^2/2} dx \\ &= 4q \int_0^\infty (2x)^{q-1} e^{-x} dx \\ &= q2^{q+1} \int_0^\infty x^{q-1} e^{-x} dx \\ &= q2^{q+1} \Gamma(q-1+1) \\ &= 2q!(2v^2)^q. \end{aligned}$$

Reciprocamente, suponha que  $\mathbb{E} [X^{2q}] \leq q!C^q$ . Tome  $X'$  uma cópia i.i.d. de  $X$ . Então,

$$\mathbb{E} [e^{\lambda X}] \mathbb{E} [e^{-\lambda X}] = \mathbb{E} [e^{\lambda(X-X')}] = \sum_{i=0}^\infty \frac{\lambda^{2q} \mathbb{E} [(X-X')^{2q}]}{(2q)!},$$

já que se  $Z = X - X'$ , note que  $Z \sim -Z$  e como  $x \mapsto x^{2q-1}$  são funções ímpares, temos que  $\mathbb{E} [Z^{2q-1}] = \mathbb{E} [(-Z)^{2q-1}] = -\mathbb{E} [Z^{2q-1}]$  e portanto  $\mathbb{E} [Z^{2q-1}] = 0$ .

Como  $g(x) = x^{2q}$  é convexa, temos que  $g(x/2 - y/2) \leq g(x)/2 + g(-y)/2$ , isto é,

$$\frac{1}{2^{2q}}(x-y)^{2q} \leq \frac{1}{2}(x^{2q} + y^{2q}),$$

dessa forma, temos que

$$\begin{aligned}
\mathbb{E} \left[ e^{\lambda X} \right] \mathbb{E} \left[ e^{-\lambda X} \right] &= \sum_{i=0}^{\infty} \frac{\lambda^{2q} \mathbb{E} \left[ (X - X')^{2q} \right]}{(2q)!} \\
&\leq \sum_{i=0}^{\infty} \frac{\lambda^{2q} 2^{2q} \mathbb{E} \left[ X^{2q} \right]}{(2q)!} \\
&\leq \sum_{i=0}^{\infty} \lambda^{2q} 2^{2q} C^q \frac{q!}{(2q)!} \\
&\leq \sum_{i=0}^{\infty} \lambda^{2q} 2^{2q} C^q \frac{1}{2^q q!} \\
&\leq \sum_{i=0}^{\infty} \lambda^{2q} 2^q C^q \frac{1}{q!} \\
&\leq \sum_{i=0}^{\infty} \frac{(2C\lambda^2)^q}{q!}.
\end{aligned}$$

Ou seja,

$$\mathbb{E} \left[ e^{\lambda X} \right] \leq e^{2C\lambda^2} \frac{1}{\mathbb{E} \left[ e^{-\lambda X} \right]} \leq e^{2C\lambda^2},$$

já que  $\mathbb{E} [X] = 0$  e portanto por Jensen temos que  $\mathbb{E} \left[ e^{-\lambda X} \right] \geq e^{-\lambda \mathbb{E}[X]} = 1$ .  $\square$

A hipótese de sub-gaussianidade é interessante para generalizar resultados que em geral valem para v.a. normais. A seguir vamos provar a desigualdade maximal.

**Teorema 3** (Desigualdade Maximal). *Sejam  $\{X_i\}_{i=1}^n$  v.a. i.i.d., centradas com sub-gaussianas com constante  $v^2$ . Então*

$$\mathbb{E} \left[ \max_{i=1, \dots, n} X_i \right] \leq \sqrt{v^2 n \log(n)}$$

*Demonstração.* Por Jensen, e o fato de que  $x \mapsto e^{\lambda x}$  é crescente para  $\lambda > 0$  temos

$$\begin{aligned}
e^{\lambda \mathbb{E}[\max_{i=1, \dots, n} X_i]} &\leq \mathbb{E} \left[ e^{\lambda \max_{i=1, \dots, n} X_i} \right] \\
&= \mathbb{E} \left[ \max_{i=1, \dots, n} e^{\lambda X_i} \right] \\
&\leq \sum_{i=1}^n \mathbb{E} \left[ e^{\lambda X_i} \right].
\end{aligned}$$

Como as v.a. são sub-gaussianas, temos

$$\begin{aligned} e^{\lambda \mathbb{E}[\max_{i=1,\dots,n} X_i]} &\leq \sum_{i=1}^n \mathbb{E} \left[ e^{\lambda X_i} \right] \\ &\leq n e^{\lambda^2 v^2 / 2}. \end{aligned}$$

Dessa forma, temos o seguinte

$$\mathbb{E} \left[ \max_{i=1,\dots,n} X_i \right] \leq \frac{1}{\lambda} \left( \log n + \frac{\lambda^2 v^2}{2} \right),$$

que é minimizado quando  $\lambda = \sqrt{2(\log n)/v^2}$  o que nos dá

$$\mathbb{E} \left[ \max_{i=1,\dots,n} X_i \right] \leq \sqrt{2v^2 \log n}.$$

□

### 1.3 Desigualdade de Hoeffding

Na seção anterior aprendemos sobre v.a. sub-gaussianas, agora vamos mostrar que toda v.a. limitada é subgaussiana. Alguns elementos da prova do teorema serão úteis no futuro.

**Teorema 4** (Lema de Hoeffding). *Seja  $X$  uma v.a. com  $\mathbb{E}[X] = 0$  e tal que  $X \in [a, b]$ . Então temos que  $\psi_X(\lambda) \leq \frac{\lambda^2(b-a)^2}{4}$ .*

*Demonstração.* Note que a distância de  $X$  metade do intervalo  $(a, b)$  é sempre menor que a metade do intervalo  $(a, b)$ , isto é,

$$\left| X - \frac{b+a}{2} \right| \leq \frac{b-a}{2}.$$

Dessa forma temos que  $\text{Var}[X] = \text{Var} \left[ X - \frac{b+a}{2} \right] \leq \frac{(b-a)^2}{4}$ .

Agora, seja  $dP$  a distribuição de  $X$  e considere a distribuição  $dP_\lambda(x) = e^{-(\psi_X(\lambda) - \lambda x)} dP(x)$ . Note que

$$\int dP_\lambda(x) = \int e^{-(\psi_X(\lambda) - \lambda x)} dP(x) = \mathbb{E}_P \left[ e^{\lambda X} \right] / \mathbb{E}_P \left[ e^{\lambda X} \right] = 1$$

logo,  $dP_\lambda$  é de fato uma distribuição.



Agora note que

$$\begin{aligned}
\psi''(\lambda) &= \left( \log(\mathbb{E}[e^{\lambda X}]) \right)'' \\
&= \left( \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} \right)' \\
&= \frac{\mathbb{E}[X^2 e^{\lambda X}] \mathbb{E}[e^{\lambda X}] - (\mathbb{E}[Xe^{\lambda X}])^2}{\mathbb{E}[e^{\lambda X}]^2} \\
&= \frac{\mathbb{E}[X^2 e^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} - \left( \frac{\mathbb{E}[Xe^{\lambda X}]}{\mathbb{E}[e^{\lambda X}]} \right)^2 \\
&= \mathbb{E}_{P_\lambda}[X^2] - \mathbb{E}_{P_\lambda}[X]^2 \\
&= \text{Var}_{P_\lambda}[X] \\
&\leq \frac{(b-a)^2}{4}.
\end{aligned}$$

Além disso,  $\psi(0) = 0$  e  $\psi'(0) = \mathbb{E}[X] = 0$ . Logo

$$\psi(\lambda) = \psi(0) + \psi'(0)\lambda + \frac{\lambda^2}{2}\psi''(\theta) \leq (b-a)^2/4,$$

onde  $\theta \in [0, \lambda]$ . □

Na prova acima, perceba que se  $\mathbb{E}[X] = 0$  então limitar  $\psi''(\lambda)$  é suficiente para limitar  $\psi(\lambda)$ , isso nos sera útil no futuro.

## 2 Estimando a variância

### 2.1 Efron-Stein

Nessa capítulo vamos estudar v.a. da forma  $Z = f(X_1, \dots, X_n)$  onde  $X_i$  são v.a. independentes e  $f : \mathcal{X} \rightarrow \mathbb{R}$ .

Sabemos que no caso de soma de v.a.s independentes,  $S_n = X_1 + \dots + X_n$ , temos que

$$\text{Var} [S_n] = \sum_{i=1}^n \text{Var} [X_i],$$

isto é, a variância da função soma é aditiva. Nessa seção o nosso objetivo é encontrar algo parecido para funções quaisquer  $f$ .

Antes de continuarmos, vamos introduzir a seguinte notação para a esperança condicional

$$\mathbb{E}_{a:b} [Z] = \mathbb{E} [Z | X_1, \dots, X_{a-1}, X_{b+1}, \dots, X_n], \quad (4)$$

isto é, o operador  $\mathbb{E}_{a:b} [\cdot]$  integra as variáveis entre  $a$  e  $b$  e o resto é fixo.

Dessa forma temos que

$$Z - \mathbb{E} [Z] = \sum_{i=1}^n \Delta_i,$$

onde  $\Delta_i = \mathbb{E}_{i+1:n} [Z] - \mathbb{E}_{i:n} [Z]$ . Perceba que  $\Delta_j$  já teve todas as variáveis depois de  $j$  integradas e portanto só sobrou aleatoriedade antes de  $j$ . Portanto,

$$\begin{aligned} \text{Var} [Z] &= \mathbb{E} \left[ \left( \sum_{i=1}^n \Delta_i \right)^2 \right] \\ &= \mathbb{E} \left[ \sum_{i,j=1}^n \Delta_i \Delta_j \right] \\ &= \mathbb{E} \left[ \sum_i \Delta_i^2 + 2 \sum_{i < j} \Delta_i \Delta_j \right] \\ &= \sum_i \mathbb{E} [\Delta_i^2] + 2 \sum_{i < j} \mathbb{E} [\Delta_i \Delta_j] \\ &= \sum_i \mathbb{E} [\Delta_i^2] + 2 \sum_{i < j} \mathbb{E} [\mathbb{E}_{i+1:n} [\Delta_i \Delta_j]]. \end{aligned}$$

Note que  $\mathbb{E}_{i+1:n} [\Delta_i \Delta_j] = \Delta_i \mathbb{E}_{i+1:n} [\Delta_j] = \Delta_i \cdot 0$ .

Concluimos então que

$$\text{Var} [Z] = \sum_{i=1}^n \mathbb{E} [\Delta_i^2]. \quad (5)$$

Agora note que  $\mathbb{E}_{i+1:n} [\mathbb{E}_{i:i} [Z]] = \mathbb{E}_{i:n} [Z]$  e portanto

$$\Delta_i^2 = (\mathbb{E}_{i+1:n} [Z - \mathbb{E}_{i:i} [Z]])^2.$$

Por Jensen, temos então que

$$\Delta_i^2 = (\mathbb{E}_{i+1:n} [Z - \mathbb{E}_{i:i} [Z]])^2 \leq \mathbb{E}_{i+1:n} [(Z - \mathbb{E}_{i:i} [Z])^2],$$

concluindo então que

$$\begin{aligned} \mathbb{E} [\Delta_i^2] &\leq \mathbb{E} [(Z - \mathbb{E}_{i:i} [Z])^2] = \mathbb{E} [\mathbb{E}_{i:i} [(Z - \mathbb{E}_{i:i} [Z])^2]] \\ &= \mathbb{E} [\text{Var}_{i:i} [Z]]. \end{aligned}$$

Juntando isso com Eq. 5, temos o que é conhecido com desigualdade de Efron-Stein:

$$\text{Var} [Z] \leq \sum_{i=1}^n \mathbb{E} [(Z - \mathbb{E}_{i:i} [Z])^2] = \sum_{i=1}^n \mathbb{E} [\text{Var}_{i:i} [Z]].$$

Essas identidades nos dizem que para estimar a variância de  $Z$ , basta estudarmos como  $f$  se comporta elemento a elemento.

Além disso, qualquer propriedade que conseguimos derivar para uma coordenada de  $f(X_1, \dots, X_n)$ , é automaticamente estendida para a variância de  $Z$ . Por exemplo, pela caracterização da esperança como o minimizador do erro quadrático, também temos que

$$\text{Var} [Z] \leq \inf_{\tilde{Z}_i} \sum_{i=1}^n \mathbb{E} [\mathbb{E}_{i:i} [Z - \tilde{Z}_i]^2].$$

Outro exemplo: sabemos que se  $X, X'$  são cópias i.i.d., então

$$\mathbb{E} [X - X']^2 = \mathbb{E} [X - \mathbb{E} [X] - (X' - \mathbb{E} [X'])]^2 = 2\text{Var} [X].$$

Logo, temos que

$$\mathbb{E} [\text{Var}_{i:i} [Z]] = \frac{1}{2} \mathbb{E} [\mathbb{E}_{i:i} [Z - Z'_i]^2]$$

onde  $Z'_i$  é uma cópia de  $Z$  mas com coordenada  $X_i$  trocada por uma cópia i.i.d.  $X'_i$ . E portanto, também temos que

$$\text{Var} [Z] \leq \frac{1}{2} \sum_{i=1}^n \mathbb{E} [\mathbb{E}_{i:i} [Z - Z'_i]^2].$$

## 2.2 Diferenças limitadas

Vamos estudar o que acontece quando a função  $f$  satisfaz uma propriedade muito comum. Dizemos que  $f : \mathcal{X} \rightarrow \mathbb{R}$  satisfazer a propriedade das diferenças limitadas se existem  $c_1, \dots, c_n$  positivos tal que:

$$\sup_{x_1, x'_1, \dots, x_n, x'_n} |f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, x'_i, \dots, x_n)| \leq c_i,$$

isto é, se fixado todas as coordenadas, exceto a  $i$ -ésima, então  $f$  não muda mais que  $c_i$ .

Usando o mesmo argumento que a distância de  $Z$  até o ponto médio dos extremos do intervalo que está contigo é no máximo metade do tamanho do intervalo, isto é,

$$|Z - (b + a)/2| \leq (b - a)/2$$

onde  $b = \sup_{x'_i} f(X_1, \dots, x'_i, \dots, X_n)$  e  $a = \inf_{x'_i} f(X_1, \dots, x'_i, \dots, X_n)$ , temos que

$$\text{Var}[Z] \leq \sum_{i=1}^n \mathbb{E}[\text{Var}_{i:i}[Z]] \leq \sum_{i=1}^n (b - a)^2/4 \leq \sum_{i=1}^n c^2/4.$$

## A Desigualdades básicas

### A.1 Fatorial

**TODO:** Falar de  $2p!$  e  $2p!/p!$

## B Convexidade

### B.1 Jensen

## Referências

[Boucheron et al., 2013] Boucheron, S., Lugosi, G., and Massart, P. (2013). *Concentration Inequalities: A Nonasymptotic Theory of Independence*. Oxford University Press. [2](#)