Concentração de Medida

Thiago Rodrigo Ramos

27 de dezembro de 2024

Sumário

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*

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

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

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

□

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

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

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*

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

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

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

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*

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

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 $\mathbb{E}\left[e^{\lambda X}\right]$. Por questões que ficarão mais claras no futuro, vamos reescrever a expressão 1 da seguinte forma:

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

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

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

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

onde

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

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 ψ_X^* é conhecida como transformada de Crammér de X e mais geralmente se o supremo é sobre todos os valores de λ, é conhecida como Função Dual Fenchel-Legendre de ψ_X.

TODO: Explicar como generalizar para λ 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

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

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 $S_n = \sum_{1=i}^n X_i$. Então, $\mathbb{E}\left[e^{\lambda S_n}\right] = \prod_{i=1}^n \mathbb{E}\left[e^{\lambda X_i}\right] = M_X(\lambda)^n$ 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

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

Podemos também considerar a probabilide 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,

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

TODO: falar que precisa ser centrada

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

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

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,*

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

então para qualquer $q \geq 1$,

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

Reciprocamente, se existir constante C tal que

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

então X ś sub-gaussina 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* σ^2. *Então sabemos que*

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

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

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

Agora note que,

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

E portanto,

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

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

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

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

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

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

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

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

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

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

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

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

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

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*

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

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

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

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

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

Dessa forma, temos o seguinte

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

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

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

□

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 métade 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 $\mathrm{Var}[X] = \mathrm{Var}\left[X - \frac{b+a}{2}\right] \leq \frac{(b-a)^2}{4}$.

Agora, seja dP a distribuição de X e consdiere 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_λ é de fato uma distribuição.

Agora note que

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

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

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

onde $\theta \in [0, \lambda]$. $\square$

Na prova acima, perceba que se $\mathbb{E}\left[X\right] = 0$ 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 $f : \mathcal{X} \to \mathbb{R}$.

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

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

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

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

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 Δ_j já teve todas as variáveis depois de j integradas e portanto só sobrou aleatoriedade antes de j. Portanto,

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

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

Concluímos então que

$$\operatorname{Var}[Z] = \sum_{i=1}^{n} \mathbb{E}\left[\Delta_i^2\right]. \tag{5}$$

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

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

Por Jensen, temos então que

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

concluindo então que

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

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

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

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 $f(X_1, \ldots, X_n)$, é automaticamente estendida para a variância de Z. Por exemplo, pela caracterização da esperança como o minimizador do erro quadrático, também temos que

$$\mathrm{Var}\left[Z\right] \leq \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}_{i:i}\left[Z - \tilde{Z}_i\right]^2\right] = \sum_{i=1}^{n} \mathbb{E}\left[(Z - \tilde{Z}_i)^2\right],$$

para qualquer $\tilde{Z}_i$ mensurável no produto ignorando i, sendo uma igualdade com a variância que encontramos anteriormente quando $\tilde{Z}_i = \mathbb{E}_{i:i}\left[Z\right]$ Outro exemplo: sabemos que se X, X' são cópias i.i.d., então

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

Logo, temos que

$$\mathbb{E}\left[\mathrm{Var}_{i:i}\left[Z\right]\right] = \frac{1}{2}\mathbb{E}\left[\mathbb{E}_{i:i}\left[Z - Z_i'\right]^2\right]$$

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

$$\begin{aligned}\mathrm{Var}\left[Z\right] &\leq \frac{1}{2}\sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}_{i:i}\left[Z - Z_i'\right]^2\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}_{i:i}\left[Z - Z_i'\right]_+^2\right] \\ &= \sum_{i=1}^{n} \mathbb{E}\left[\mathbb{E}_{i:i}\left[Z - Z_i'\right]_-^2\right],\end{aligned}$$

onde usamos o fato de que $(Z - Z')_+ \sim (Z' - Z)_+ = (Z - Z')_-$ e $Z_+ Z_- = 0$.

Exemplo 4 (Diferenças limitadas). *Vamos estudar o que acontece quando a função f satisfaz uma propriedade muito comum. Dizemos que $f : \mathcal{X} \to \mathbb{R}$ satisfazer a propriedade das diferenças limitadas se existem $c_1, \dots, c_n$ positivos tal que:*

$$\sup_{x_1, x_1', \dots, x_n, x_n'} |f(x_1, \dots, x_i, \dots, x_n) - f(x_1, \dots, x_i', \dots, x_n)| \le c_i,$$

isto é, se fixado todas as coordenadas, exceto a i-esima, 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,

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

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

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

Exemplo 5 (Maior autovalor de matrizes aleatórias). *Seja A uma matriz simétrica com coordenadas $X_{i,j}$ i.i.d. e $|X_{i,j}| \le 1$, estamos interessados em estudar $Z = f(A) = \lambda_1$ onde λ_1 é seu maior autovalor. Sabemos que*

$$\lambda_1 = \sup_{\|u\|=1} u^t A u.$$

Além disso, se v é autovetor correspondente à λ_1, então

$$v^t A v = \lambda_1 v^t v = \lambda_1$$

pela ortonormalidade da decomposição de matrizes simétricas.

Vamos tentar estudar a variância de Z, para isso note que se $A_{i,j}'$ é uma cópia de A com a entrada $X_{i,j}'$ sendo uma cópia i.i.d. de $X_{i,j}$, então para um v qualquer, temos que

$$\begin{aligned} f(A) - f(A_{ij}') &= v^t A v - \sup_{\|u\|=1} u^t A_{ij}' u \\ &\le v^t A v - v^t A_{ij}' v \\ &= v^t (A - A_{ij}') v \\ &= v^t ((e_i e_j^t + e_j e_i^t)(X_{ij} - X_{ij}')) v \\ &= v^t ((e_i e_j^t + e_j e_i^t)) v (X_{ij} - X_{ij}') \\ &= 2 v_i v_j (X_{ij} - X_{ij}') \\ &\le 4 v_i v_j. \end{aligned}$$

Dessa forma, por Efron-Stein e o fato que $\|v\| = 1$,

$$\operatorname{Var}\left[f(A)\right] \leq \sum_{i \leq j} \mathbb{E}\left[(f(A) - f(A'_{ij}))_+^2\right] \leq \sum_{i \leq j} 16 v_i^2 v_j^2 = 16.$$

TODO: *Para tomarmos o quadrado dos dois lados, precisamos garantir que as coisas sao positivas, por isso é importante truncar positivamente o lado esquerdo para garantir que é positivo!!! Isso corrige o bound*

Exemplo 6 (Desigualdade de Poincaré)**.** *Suponha que temos uma função contínua* $f : [0,1]^n \to \mathbb{R}$ *que é convexa coordenada a coordenada e tem derivadas parciais.*

Por convexidade temos $f'(x)(x - y) > f(x) - f(y)$, *portanto, contanto que* $f(x) - f(y) > 0$, *teremos que* $f'(x)^2(x - y)^2 > (f(x) - f(y))^2$.

Considere então $Z_i = \min_{x'_i} f(X_1, \ldots, x_i, \ldots, X_n)$ *e* X_i^* *o valor que atinge esse mínimo, que existe por compacidade e continuidade. Então,*

$$\begin{aligned}
(Z - Z_i)^2 &= (f(X_1, \ldots, X_i, \ldots, X_n) - f(X_1, \ldots, X'_i, \ldots, X_n))^2 \\
&\leq \partial_i f(X)^2 (X_i - X'_i)^2 \\
&\leq \partial_i f(X)^2.
\end{aligned}$$

Dessa forma

$$\operatorname{Var}\left[Z\right] \leq \sum_{i=1}^{n} \mathbb{E}\left[(Z - Z_i)^2\right] \leq \sum_{i=1}^{n} \partial_i f(X)^2 = \|\nabla f(X)\|^2.$$

2.2 Cotas exponenciais via Efron-Stein

Vamos apresentar agora um truque para conseguirmos cotas exponenciais utilizando as cotas da variância que temos via Efron-Stein. Para isso, primeiro vamos definir os quantis

$$Q_\alpha = \inf\{z : \mathbb{P}\left\{Z \leq z\right\}) \geq \alpha\}.$$

Considere a seguinte função

$$g_{a,b}(x) = \begin{cases} b & \text{se } f(x) \geq b \\ a & \text{se } f(x) \leq a \\ f(x) & \text{se } a < f(x) < b \end{cases}$$

Note que, se $a \geq Mf(X)$, onde $M = Q_{1/2}$ é a mediana, temos

$$
\begin{aligned}
\mathbb{E}\left[g_{a,b}(X)\right] &= b\mathbb{E}\left[\mathbb{I}\left\{f(X) \geq b\right\}\right] + a\mathbb{P}\left\{f(X) \leq a\right\}) + \mathbb{E}\left[f(X)\mathbb{I}\left\{a < f(X) < b\right\}\right] \\
&\leq b\mathbb{E}\left[\mathbb{I}\left\{f(X) \geq b\right\}\right] + a\mathbb{P}\left\{f(X) \leq a\right\}) + b\mathbb{E}\left[\mathbb{I}\left\{a < f(X) < b\right\}\right] \\
&= a\mathbb{P}\left\{f(X) \leq a\right\}) + b\mathbb{E}\left[\mathbb{I}\left\{a < f(X)\right\}\right] \\
&= a\mathbb{P}\left\{f(X) \leq a\right\}) + b\mathbb{P}\left\{a < f(X)\right\}) \\
&= a\mathbb{P}\left\{f(X) \leq a\right\}) + b(1 - \mathbb{P}\left\{f(X) \leq a\right\})) \\
&= a\mathbb{P}\left\{f(X) \leq a\right\}) + b - b\mathbb{P}\left\{f(X) \leq a\right\}) \\
&= b + (b-a)(-\mathbb{P}\left\{f(X) \leq a\right\})) \\
&\leq b + (b-a)(-\mathbb{P}\left\{f(X) \leq MZ\right\})) \\
&\leq b - (b-a)/2 \\
&= (b+a)/2,
\end{aligned}
$$

já que $(b-a) > 0$ e $\mathbb{P}\left\{f(X) \leq a\right\}) \geq \mathbb{P}\left\{f(X) \leq MZ\right\}) \geq 1/2$.

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