Desigualdade probabilística de Hoeffding ilustrada com um caso específico simples[1]

No artigo (de Machine Learning) de Sébastien Bubeck e Mark Sellke *A Universal Law of Robustnesss via Isoperimetry*[2] é utilizada a desigualdade de Hoeffding (1963) [1] na demonstração do teorema principal. Esta desigualdade probabilística apresenta o seguinte enunciado, em tradução do original:

Desigualdade de Hoeffding [Hoeffding (1963) Theorem 2]: *Se X_1, X_2, ..., X_n forem variáveis aleatórias independentes e $a_i \leq X_i \leq b_i$, $i = 1, 2, \ldots, n$, então, para $t > 0$:*

$$\mathbb{P}\left[\overline{X} - \mu \geq t\right] \leq \exp\left(-2n^2t^2 / \sum\nolimits_{i=1}^{n} (b_i - a_i)^2\right).$$

Notação: $\overline{X} = \frac{S}{n}$ e $\mu = \mathbb{E}\left[\overline{X}\right] = \frac{\mathbb{E}[S]}{n}$, em que a soma $S = \sum_{i=1}^{n} X_i$ e $\mathbb{E}\left[\overline{X}\right]$ é a esperança matemática de $\overline{X}$.

Embora a demonstração deste teorema não faça referência explícita à desigualdade de Markov, no caso particular do exercício que apresento a seguir, vou usá-la para facilitar a sua demonstração, à semelhança do que é feito neste vídeo de MIT RES.6-012 *Introduction to Probability* [2]. No exercício, à excepção da utilização da desigualdade de Markov, sigo, para mais fácil generalização, os passos da demonstração do teorema original adaptada ao caso apresentado.

Desigualdade de Markov[3]: *se Z for uma variável aleatória positiva ou nula cuja esperança matemática se designa por $\mathbb{E}[Z]$ e c uma constante positiva, então*

$$\mathbb{P}(Z \geq c) \leq \frac{\mathbb{E}[Z]}{c}.$$

Exercício (caso particular da desigualdade de Hoeffding): *Sejam X_1, X_2, $\cdots$, X_n variáveis aleatórias independentes que tomam, com igual probabilidade, os valores -1 e 1*

$$\mathbb{P}(X_i = -1) = \mathbb{P}(X_i = 1) = 1/2, \qquad i = 1, \ldots, n.$$

[1] Américo Tavares, `https://problemasteoremas.wordpress.com/2022/04/04/desigualdade-de-hoeffding-ilustrada-com-um-caso-especifico-simples/`

[2] `https://arxiv.org/abs/2105.12806`

[3] `https://en.wikipedia.org/wiki/Markov%27s_inequality`

Designando a média das variáveis aleatórias por $\overline{X} = S/n$, em que $S = \sum_{i=1}^{n} X_i$ é a sua soma, e fazendo uso da desigualdade de Markov, determine o seguinte majorante da probabilidade da média ser pelo menos t

$$\mathbb{P}\left[\overline{X} \geq t\right] \leq e^{-nt^2/2}, \qquad t > 0. \tag{1}$$

Resolução: Se substituirmos $\overline{X}$ em $\mathbb{P}\left[\overline{X} \geq t\right]$, obtemos a probabilidade equivalente

$$\mathbb{P}\left[\overline{X} \geq t\right] = \mathbb{P}\left[\frac{S}{n} \geq t\right] = \mathbb{P}\left[S \geq nt\right].$$

Seja, agora, h uma constante positiva arbitrária. Como a condição $S \geq nt$ é equivalente a $e^{hS} \geq e^{hnt}$, podemos escrever

$$\mathbb{P}\left[\overline{X} \geq t\right] = \mathbb{P}\left[S \geq nt\right] = \mathbb{P}\left[e^{hS} \geq e^{hnt}\right].$$

Para majorar esta última probabilidade usamos a desigualdade de Markov, que aplicada a este caso se traduz em

$$\mathbb{P}\left[e^{hS} \geq e^{hnt}\right] \leq \frac{\mathbb{E}\left[e^{hS}\right]}{e^{hnt}}. \tag{2}$$

Substituindo o valor de S, tem-se

$$\frac{\mathbb{E}\left[e^{hS}\right]}{e^{hnt}} = \frac{\mathbb{E}\left[e^{h\left(\sum_{i=1}^{n} X_i\right)}\right]}{e^{hnt}} = \frac{\prod_{i=1}^{n} \mathbb{E}\left[e^{hX_i}\right]}{e^{hnt}}. \tag{3}$$

Dado que a função exponencial $f(x) = e^{hx}$ é convexa, o seu gráfico é limitado superiormente, no intervalo $[-1, 1]$, pela recta que une os pontos $(-1, f(-1)) = \left(-1, e^{-h}\right)$ e $(1, f(1)) = \left(1, e^{h}\right)$, cuja equação é dada por

$$y = \frac{1-x}{2}e^{-h} + \frac{x+1}{2}e^{h}.$$

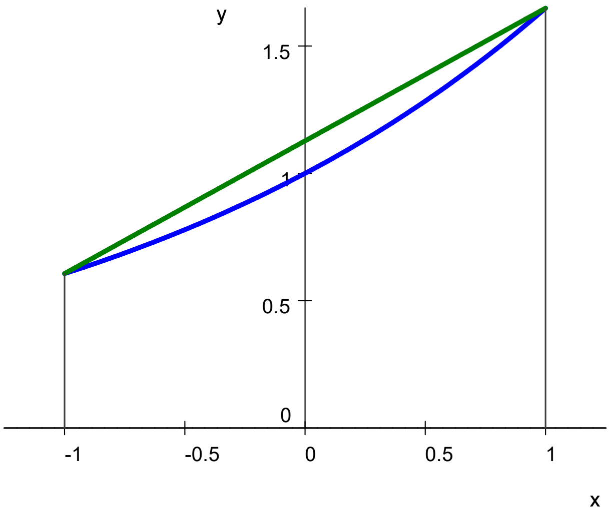
Assim,

$$f(x) = e^{hx} \leq \frac{1-x}{2}e^{-h} + \frac{x+1}{2}e^{h}, \qquad x \in [-1, 1], \tag{4}$$

pelo que $\mathbb{E}\left[e^{hX_i}\right] = \mathbb{E}\left[f(X_i)\right]$ satisfaz a condição

$$\mathbb{E}\left[e^{hX_i}\right] \leq \frac{1 - \mathbb{E}\left[X_i\right]}{2}e^{-h} + \frac{\mathbb{E}\left[X_i\right] + 1}{2}e^{h} = \frac{1}{2}\left(e^{-h} + e^{h}\right), \tag{5}$$

atendendo a que $\mathbb{E}\left[X_i\right]=0$, pois a distribuição de cada X_i é simétrica.



Gráficos de e^{hx} (azul) e $\dfrac{1-x}{2}e^{-hx}+\dfrac{x+1}{2}e^{hx}$ (verde), em $[-1,1]$, $h=1/2$

De (2), (3) e (5) resulta então

$$\mathbb{P}\left[\overline{X}\geq t\right]\leq\frac{\prod_{i=1}^{n}\mathbb{E}\left[e^{hX_i}\right]}{e^{hnt}}\leq\left(\frac{e^{-h}+e^{h}}{2e^{ht}}\right)^{n}. \tag{6}$$

Para facilitar o resto do cálculo, vamos agora reescrever $\left(e^{-h}+e^{h}\right)/2$:

$$\frac{e^{-h}+e^{h}}{2}=e^{-h-\ln 2+\ln\left(1+e^{2h}\right)}:=e^{L(h)}, \tag{7}$$

em que o expoente $L\left(h\right)=-h-\ln 2+\ln\left(1+e^{2h}\right)$. As duas primeiras derivadas de $L\left(h\right)$ são

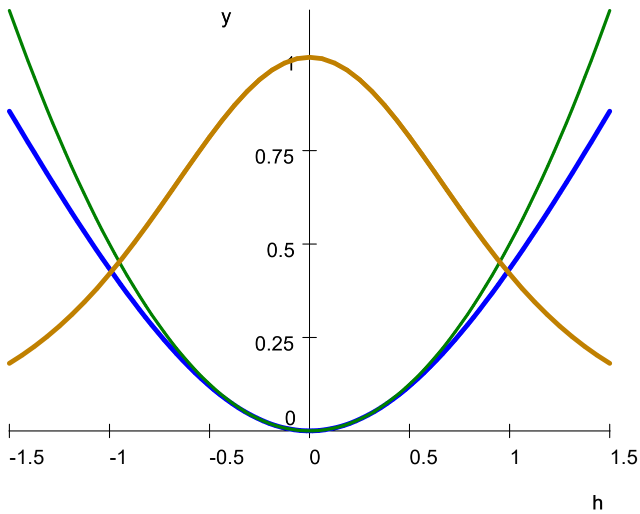
$$L'\left(h\right)=-1+\frac{2}{1+e^{-2h}}\qquad\text{e}\qquad L''\left(h\right)=\frac{4e^{-2h}}{\left(1+e^{-2h}\right)^{2}}.$$

A segunda derivada admite a seguinte factorização

$$L''\left(h\right)=4\times\frac{1}{e^{-2h}+1}\times\frac{e^{-2h}}{e^{-2h}+1}=4\times\frac{1}{e^{-2h}+1}\left(1-\frac{1}{e^{-2h}+1}\right), \tag{8}$$

isto é, $L''\left(h\right)=4u(1-u)$ em que $u=1/(e^{-2h}+1)\in\,]0,1[$, uma vez que $0<1/(e^{-2h}+1)<1$. Ora, o máximo de $4u(1-u)$ ocorre quando $u=1/2$, donde $L''\left(h\right)\leq 4\times(1/2)(1-1/2)=1$. Pelo desenvolvimento em série de Taylor, dado que $L\left(0\right)=L'\left(0\right)=0$, vem

$$L\left(h\right)=L(0)+L'\left(0\right)h+\frac{L''\left(0\right)}{2}h^{2}+\cdots\leq\frac{1}{2}h^{2}. \tag{9}$$



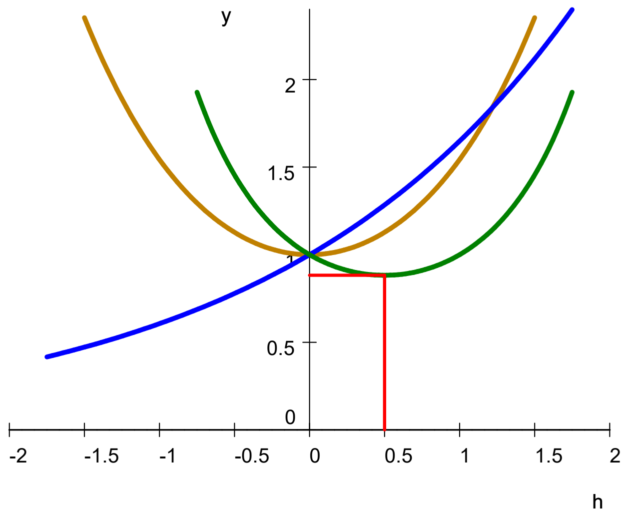
Gráficos de $L(h)$ (azul claro), $h^2/2$ (verde) e $L''(h)$ (castanho claro)

No gráfico anterior mostram-se os andamentos de $L(h)$, $h^2/2$ e $L''(h)$. De (7) e (9), resulta que $\left(e^{-h}+e^{h}\right)/2=e^{L(h)}\leq e^{h^2/2}$, e de (6),

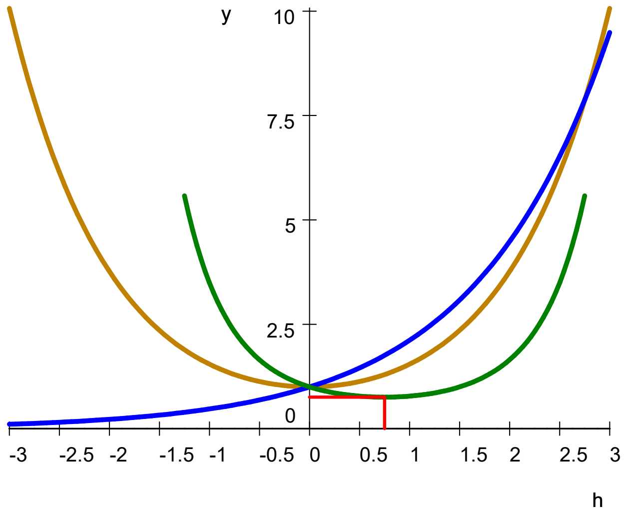
$$\mathbb{P}\left[\overline{X}\geq t\right]\leq\left(\frac{e^{-h}+e^{h}}{2e^{ht}}\right)^{n}\leq e^{-hnt+nh^{2}/2}. \tag{10}$$

Designe-se o expoente de (10) por $g(h)=-hnt+nh^2/2$. Visto que $\min_h\exp\left[g(h)\right]=\exp\left[\min_h g(h)\right]$ e $g'(h)=nh-nt=0$, o segundo membro de (10) tem o mínimo em $h_{\min}=t$. Finalmente, inserindo este valor em (10), obtemos o majorante $e^{-hnt+nh^2/2}=e^{-nt^2/2}$ de $\mathbb{P}\left[\overline{X}\geq t\right]$ indicado em (1), o que demonstra a desigualdade de Hoeffding, neste caso específico.

Apresento a seguir dois exemplos gráficos para os casos $t=1/2$ e $t=3/4$.



Gráficos de $\left(e^h + e^{-h}\right)/2$ (castanho claro), e^{ht} (azul) e $e^{-ht+h^2/2}$ (verde), para $t = 1/2$. O mínimo da curva verde ocorre em $h_{\text{min}} = t = 1/2$



Gráficos de $\left(e^h + e^{-h}\right)/2$ (castanho claro), e^{ht} (azul) e $e^{-ht+h^2/2}$ (verde), para $t = 3/4$. O mínimo da curva verde ocorre em $h_{\text{min}} = t = 3/4$

Referências

[1] Wassily Hoeffding, *Probability inequalities for sums of bounded random variables* (PDF). Journal of the American Statistical Association. 58 (301): 13–30. Acessível via Wikipedia https://en.wikipedia.org/wiki/Hoeffding%27s_inequality

[2] John Tsitsiklis, MIT RES.6-012 Introduction to Probability, Spring 2018, S18.3 *Hoeffding's Inequality* https://youtu.be/MWcO8ZTOQQQ