ESTIMADORES INESPERADOS PARA O VALOR ESPERADO

Roberto Imbuzeiro Oliveira

IMPA

- http://w3.impa.br/~rimfo

— Baseado em artigos —

de outros

Christian Brownlees
Olivier Catoni
Shahar Mendelson

com outros

Luc Devroye
Emilien Joly
Matthieu Lerasle
Gábor Lugosi

PARTE 1

Introdução

VALORES ESPERADOS

- Qual a altura média de um brasileiro?
- Qual é a sobrevida média de um paciente de câncer depois de um tratamento experimental?
- Qual o consumo médio de energia elétrica em seu bairro nas noites de janeiro?

PROBLEMA BÁSICO

Queremos estimar médias sobre populações (= valores esperados) de modo que a chance de erros grandes seja mínima.

Nossa abstração A população é dada por uma distribuição de probabilidade sobre $\mathbb{R}$.

DEFINIÇÕES MATEMÁTICAS

P uma distribuição de probabilidade sobre $\mathbb{R}$

- cada $A \subset \mathbb{R}$ mensurável tem uma probabilidade

$P(A) \in [0,1]$, com $P(\mathbb{R}) = 1$, $P(\emptyset) = 0$

& $P(A_1 \cup A_2 \cup \dots \cup A_n \cup \dots) = \sum_{n=1}^{\infty} P(A_n)$

para $A_1, \dots, A_n, \dots$ disjuntos

Exemplos

$P(A) := \int_A p(x)\,dx$ onde p é densidade

$P(A) = \sum_{n \in \mathbb{N} \cap A} p_n$ onde $p_n \geq 0$, $\sum_{n \in \mathbb{N}} p_n = 1$.

MAIS DEFINIÇÕES

μ_P := valor esperado de P

$$= \int_{\mathbb{R}} x \, P(dx) \quad \left(\text{ou } \mathbb{E}_P(X) \right)$$

$$= \int_{\mathbb{R}} x \, p(x) \, dx \quad (\text{se } P \text{ tem densidade } p)$$

$$= \sum_{n \in \mathbb{N}} n \, p_n \quad (\text{se } P \text{ é dada por pesos sobre } \mathbb{N})$$

OS DADOS

- Em geral não temos acesso a toda uma população (ou seja, a P).

- Suporemos que temos uma amostra de n variáveis aleatórias independentes e identicamente distribuídas $X_1^n := (X_1, \dots, X_n) =_d P^n$

$$\mathbb{P}\left(\bigcap_{i=1}^{n} \{X_i \in A_i\}\right) = \prod_{i=1}^{n} P(A_i)$$

OBJETIVO (V. 2.0)

- Queremos estimar μ_P a partir de $X_1^n =_d P^n$, com P desconhecida

[Hipótese: $P \in \mathcal{P}$, uma família conhecida.]

Estimador $\hat{E}_n : \mathbb{R}^n \to \mathbb{R}$

▷ queremos $|\hat{E}_n(X_1^n) - \mu_P| << 1$ com a maior probabilidade possível.

MEDINDO A QUALIDADE

Catoni : na melhor das hipóteses, não podemos esperar nada melhor que

$$|\hat{E}_n(X_1^n) - \mu_P| \sim \frac{\sigma_P}{\sqrt{n}}.$$

σ_P = desvio padrão (suporemos finito)

$$\sigma_P^2 := \mathbb{E}_P[(X-\mu_P)^2] \quad \text{(variância)}$$

$$= \int_{\mathbb{R}} (x-\mu_P)^2 \, P(dx).$$

MEDINDO A QUALIDADE (II)

$\mathcal{P}$ família de distribuições sobre $\mathbb{R}$

$$P \in \mathcal{P} \mapsto \begin{cases} \mu_P := \mathbb{E}_P(X) \\ \sigma_P^2 := \mathbb{E}_P((X-\mu)^2) \end{cases}$$

$\hat{E}_n : \mathbb{R}^n \to \mathbb{R}$ dado

Erro médio quadrático

$$\mathbb{E}_P\left[\left(\hat{E}_n(X_1^n) - \mu_P\right)^2\right] = \frac{\sigma_P^2}{n}$$

$\forall P \in \mathcal{P}$

Flutuações

$$\mathbb{P}_P\left(\left|\hat{E}_n(X_1^n) - \mu_P\right| > \frac{\sigma_P\, r(\delta)}{\sqrt{n}}\right) \leq \delta$$

com $\delta \in (0,1)$, $\forall P \in \mathcal{P}$

& $r(\delta)$ menor possível.

CRITÉRIO DAS FLUTUAÇÕES

Dada uma família $\mathcal{P}$ e um tamanho de amostra n, queremos $\hat{E}_n : \mathbb{R}^n \to \mathbb{R}$ tal que:

- dada qualquer $P \in \mathcal{P}$
- dado qualquer $\delta \in (\delta_{min}, 1/2)$

$$\mathbb{P}_P\left(|\hat{E}_n(X_1^n) - \mu_P| > \frac{\sigma_P}{\sqrt{n}}\, r(\delta) \right) \leq \delta$$

com $r(\delta)$ o menor possível para a família $\mathcal{P}$.

($\hat{E}_n$ escolhido p/ minimizar $r(\delta)$ e δ_{min}!)

COMENTÁRIOS SOBRE CRITÉRIO

$$\left[\forall \delta \in (\delta_{min}, 1/2) : \sup_{P \in \mathcal{P}} \mathbb{P}_P \left(|\hat{E}_n(X_1^n) - \mu_P| > \frac{\sigma_P}{\sqrt{n}} r(\delta) \right) \leq \delta \right]$$

▷ Nossos alvos serão:

1) $\mathcal{P}$ muito grande ("não paramétrica")

2) $\delta_{min} \approx \exp(-c_{\mathcal{P}}\, n)$
(prob. exponencialmente pequena de falha)

3) $r(\delta) \approx \sqrt{\log(1/\delta)}$ subgaussiana

Nada disso pode ser muito melhorado...

EXEMPLOS DE FAMÍLIAS $\mathcal{P}$

$$\left[\forall \delta \in (\delta_{min}, 1/2) : \sup_{P \in \mathcal{P}} \mathbb{P}_P \left(\left| \hat{E}_n(X_1^n) - \mu_P \right| > \frac{\sigma_P}{\sqrt{n}} r(\delta) \right) \leq \delta \right]$$

- $\mathcal{P}_2 := \{ \text{todas as } P \text{ com } \sigma_P^2 < +\infty \}$
- $\mathcal{P}_{2,\sigma_1^2,\sigma_2^2} = \{ P \in \mathcal{P}_2 : \sigma_1^2 \leq \sigma_P^2 \leq \sigma_2^2 \}$
- $\mathcal{P}_{\alpha,K} = \{ P \in \mathcal{P}_2 : \mathbb{E}_P\left[(X-\mu_P)^\alpha \right] \leq K \sigma_P^\alpha \}$ $(\alpha > 2)$
- $\mathcal{P}_{sym} = \{ P \in \mathcal{P}_2 : P \text{ simétrica c/ relação a } \mu_P \}$

POR QUE, AFINAL?

- Se você quer probabilidade muito baixa de erro...
- Se você quer estimar muitas médias de 1 vez:

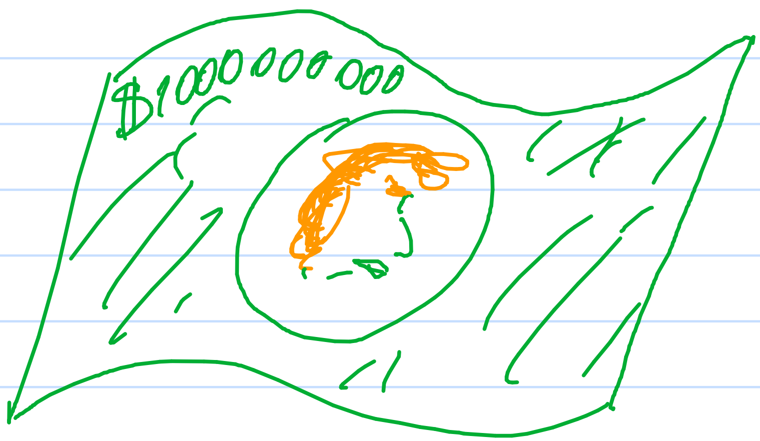
$$\left[\mu_{P_1}, \mu_{P_2}, \mu_{P_3}, \dots, \mu_{P_k}\right] \qquad k >> 1$$

$$\mathbb{P}_{P_i}\left(\left|\hat{E}_n(X_1^n(i)) - \mu_{P_i}\right| > \frac{\sigma_{P_i}}{\sqrt{n}}\, r\left(\cancel{\delta}\, \tfrac{\delta}{k}\right)\right) \leq \cancel{\delta}\, \tfrac{\delta}{k}$$

"Correção de Bonferroni"

$$\Rightarrow \mathbb{P}\left(\bigcup_{i=1}^{k}\left\{\left|\hat{E}_n(X_1^n(i)) - \mu_{P_i}\right| > \frac{\sigma_{P_i}}{\sqrt{n}}\, r\left(\tfrac{\delta}{k}\right)\right\}\right) \leq \delta$$

BIG DATA



RESUMO DOS RESULTADOS

- O estimador óbvio seria a <u>média amostral</u>

$$\hat{M}_n(X_1^n) := \frac{1}{n}\sum_{i=1}^{n} X_i ,$$

mas ela <u>não atinge nossos objetivos</u>

- <u>Outros estimadores</u>, bem diferentes, dão <u>resultados quase-ótimos.</u>

- Extensões a vetores e a problemas de regressão...

O QUE VEREMOS

- Resultados positivos, dados por estimadores explícitos

- Resultados negativos: dada $\mathcal{P}$, encontramos $P_1, P_2 \in \mathcal{P}$ com médias diferentes, mas tais que $X_1^n =_d P_1^n$ e $Y_1^n =_d P_2^n$ são parecidas.

- Versões simples de métodos desenvolvidos em Estatística Não-Paramétrica (cotas minimax, Lepskii)

ANÁLISE DA MÉDIA AMOSTRAL

ou: Chebyshev contra Gauss

(com vitória para o russo)

BREVE INTERLÚDIO

- Quem é menor?

- $\dfrac{1}{n^{10}}$ ou $e^{-\frac{n}{100}}$ $(n \to +\infty)$?

- $\sqrt{\ln 1/\delta}$ ou $\dfrac{1}{\sqrt{\delta}}$ $(\delta \approx 0)$?

A MÉDIA AMOSTRAL

$$\hat{M}_n(X_1^n) := \frac{1}{n} \sum_{i=1}^{n} X_i$$

Erro médio quadrático

$$\mathbb{E}_P\left[\left(\hat{M}_n(X_1^n) - \mu_P\right)^2\right] = \frac{\sigma_P^2}{n} \quad \forall P \text{ com } \sigma_P < +\infty.$$

DESIGUALDADE DE CHEBYSHEV

$$\mathbb{P}_P\left(\left|\hat{M}_n(X_1^n) - \mu_P\right| > \frac{\sigma_P}{\sqrt{\delta n}} \right) \leq \delta$$

$$\left[r(\delta) = \frac{1}{\sqrt{\delta}} \gg \ln(1/\delta) \right]$$

Prova

$$\mathbb{P}_P\left(\left|\hat{M}_n(X_1^n) - \mu_P\right| > x \right) \leq \frac{\mathbb{E}_P\left[(M_n(X_1^n) - \mu_P)^2 \right]}{x^2}$$

& tome $x := \dfrac{\sigma_P}{\sqrt{\delta n}}$.

MUITO MELHOR NO LIMITE !

Escreva $\overline{\Phi}(\lambda) := \int_{\lambda}^{+\infty} \frac{e^{-t^2/2}}{\sqrt{2\pi}}\,dt \quad \left(\underset{\lambda >> 1}{\sim} e^{-\lambda^2/2}, \right)$

$\overline{\Phi}^{-1}\left(\frac{\delta}{2}\right) = \lambda$ com $\overline{\Phi}(\lambda) = \delta/2$

$\sim \sqrt{2 \ln(2/\delta)}$ para $\delta << 1$

Teorema Central do Limite

$$\mathbb{P}_P\left(\left| \hat{M}_n(X_1^n) - M_P \right| > \frac{\sigma_P}{\sqrt{n}} \overline{\Phi}^{-1}(\delta/2) \right) \xrightarrow[n \to +\infty]{} \delta .$$

Assintoticamente, $r(\delta) \sim \sqrt{\ln 2/\delta}$ tá OK.

CHEBYSHEV OU GAUSS ?

$$\mathbb{P}_P\left(|\hat{M}_n(X_1^n) - m_P| > \frac{\sigma_P}{\sqrt{\delta n}}\right) \leq \delta$$

OU

$$\mathbb{P}_P\left(|\hat{M}_n(X_1^n) - m_P| > \frac{\sigma_P}{\sqrt{n}}\sqrt{2\ln\left(\frac{2}{\delta}\right)}\right) \leq \delta + \underbrace{o(1)}_{\substack{\text{vai a } 0 \text{ qdo} \\ n \to +\infty}}$$

?

CHEBYSHEV VENCE...

Teorema (Catoni) Existe um $c>0$ tal que para quaisquer n e δ, $\exists P$ com $\sigma_P < +\infty$ e

$$\mathbb{P}_P\left(|\hat{M}_n(X_1^n) - \mu_P| > \frac{\sigma_P}{\sqrt{\delta n}}\right) \geq c\,\delta.$$

- Se P é fixo e $n \to +\infty$, $r(\delta) \sim \sqrt{\ln 1/\delta}$
- Se para cada n escolho P, $r(\delta) \sim \sqrt{\frac{1}{\delta}}$

... MESMO SE TENTAMOS FAVORECER GAUSS!

$$\mathcal{P}_{4,K} := \{ P : \mathbb{E}_P[(X-\mu_P)^4] \leq K\sigma_P^2 \}$$

Barry Esséen

$$\left| \mathbb{P}_P\left(\left|\hat{M}_n(X_1^n) - \mu_P\right| \geq \frac{\sigma_P}{\sqrt{n}} \Phi^{-1}(\delta/2) \right) - \delta \right| = O\left(\frac{1}{n^c}\right)$$

▷ Bom quando $\delta > \delta_{min} \sim \frac{1}{n^c}$, mas nós queremos $\delta_{min} \sim e^{-cn}$.

Gauss perde mesmo para famílias $\mathcal{P}$ menores.

Ver Catoni...

VALE À PENA SER OTIMISTA?

"Ene vai a infinito! Levará algum tempo, mas eu acredito!"

Augusto Cury

Não faz sentido!

A convergência para Gaussianas é lenta!

SERIA MELHOR RESIGNAR-SE ?

"A vida é curta,
as amostras são pequenas
e Gauss está morto."

Emil Cioran

Não faz sentido!
Há estimadores
melhores!

ANTES DE PROSSEGUIR

Catoni se $\mathcal{P}$ contém todas as Gaussianas

$\left(\text{densidade } p(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right)\right)$

então

$$\left[r(\delta) \geq \Phi^{-1}(\delta) \sim \sqrt{2 \ln 1/\delta} \right]$$

▶ $\hat{E}_n$ é L-subgaussiano se $r(\delta) \leq L\sqrt{\ln(1/\delta)+1}$

PARTE 2

Um primeiro estimador inesperado

A MEDIANA DAS MÉDIAS

- $\mathcal{P}_2 := \{ P : \sigma_P^2 < +\infty \}$
- $\delta_{min} := \exp(1 - n/8)$

Teorema $\forall \delta \in (\delta_{min}, 1/2) \; \exists \hat{E}_{n,\delta} : \mathbb{R}^n \to \mathbb{R}$

$$\forall P \in \mathcal{P}_2 : \mathbb{P}_P\left(\left| \hat{E}_{n,\delta}(X_1^n) - \mu_P \right| > \frac{2\sqrt{e}\,\sigma_P}{\sqrt{n}} \sqrt{1 + \ln 1/\delta} \right) \leq \delta .$$

Estimador que depende do δ desejado!

Mas já é um começo.

DOIS FATOS USADOS

- $$\underbrace{\mathbb{P}\left(\mathrm{Bin}\left(b, \frac{1}{2e}\right) > \frac{n}{2}\right)}_{} \leq e^{-b}$$

  b lançamentos independentes de uma moeda com $\mathbb{P}(\text{CARA}) = \frac{1}{2e}$

- $$\mathbb{P}_p\left(\left| \frac{1}{k} \sum_{i=1}^{k} X_i - \mu_p \right| > \sqrt{\frac{2e}{k}}\, \sigma_p \right) \leq \frac{1}{2e}$$

  Chebyshev num caso particular

POR QUE MEDIANA DAS MÉDIAS?

Tome $b := \log(1/\delta)$, $k = \frac{n}{b}$ (inteiros hoje!).

$$\left(\begin{array}{c} b \\ \text{blocos} \end{array}\right) \quad \underbrace{X_1 \; X_2 \cdots X_k}\,\Big|\,\underbrace{X_{k+1}\; X_{k+2} \cdots X_{2k}}\,\Big|\,\underbrace{X_{2k+1} \cdots X_{3k}}\,\Big|\cdots X_{n-1}\; X_n \Big|$$

$$\left(\begin{array}{c} \text{médias} \\ \text{dos blocos} \end{array}\right) \quad Y_1 = \frac{1}{k}\sum_{i=1}^{k} X_i \qquad Y_2 = \frac{1}{k}\sum_{i=k+1}^{2k} X_i \qquad Y_3 = \frac{1}{k}\sum_{i=2k+1}^{3k} X_i \quad \cdots \quad Y_b$$

$$\left[\hat{E}_n(X_1^n) = \text{mediana}\,(Y_1, Y_2, \dots, Y_b)\right]$$

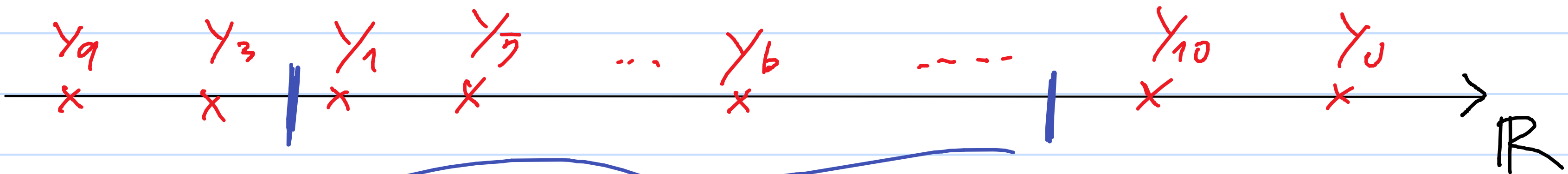
ANÁLISE

$X_1^n =_d P^n$; média μ_p e desvio σ_p

▶ $Y_1 = \frac{1}{k}\sum_{i=1}^{k} X_i$, $\mathbb{P}(Y_1 \notin I) \leq \frac{1}{2e}$.

Mesmo vale para outras Y_i , $i = 2, \ldots, b$.

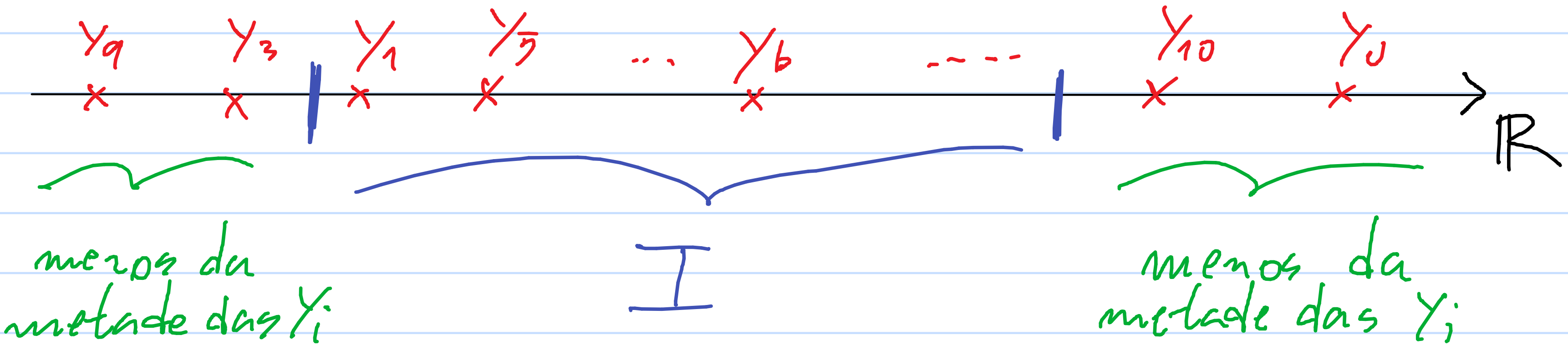
Além disso, são independentes (blocos disjuntos).



I = intervalo favorável = $\left[\mu_p - \sigma_p\sqrt{\frac{2e}{k}}, \mu_p + \sigma_p\sqrt{\frac{2e}{k}}\right]$

ANÁLISE (II)

Observação Se + da metade das Y_i cai em I, então $\hat{E}_n(X_1^n) \in I$.



$$\Rightarrow \mathbb{P}(\hat{E}_n(X_1^n) \notin I) \leq \mathbb{P}\left(\#\{i=1,\dots,b \,;\, Y_i \notin I\} \geq \frac{b}{2}\right)$$

ANÁLISE (III)

$$\mathbb{P}\left(\hat{E}_n(X_1^n) \notin I\right) \leq \mathbb{P}\left(\#\{i=1,\dots,b \,;\, Y_i \notin I\} \geq \frac{b}{2}\right)$$

Eventos $\{Y_i \notin I\}$ independentes com prob. $\leq \frac{1}{2e}$

$$\Rightarrow \quad \mathbb{P}\left(\hat{E}_n(X_1^n) \notin I\right) \leq \mathbb{P}\left(\mathrm{Bin}\left(b, \frac{1}{2e}\right) \geq \frac{b}{2}\right) \leq e^{-b}.$$

Agora lembre que $\boxed{b = \log 1/\delta}$ e

$$I = \left[\mu_P - \underbrace{\sigma_P\sqrt{\frac{2e}{k}}}_{\sqrt{\frac{2e \log 1/\delta}{n}}}\ ,\ \mu_P + \underbrace{\sigma_P\sqrt{\frac{2e}{k}}}_{\sqrt{\frac{2e \log 1/\delta}{n}}}\right] \quad \text{com} \quad \boxed{k = \frac{n}{b} = \frac{n}{\log 1/\delta}}$$

COMENTÁRIOS SOBRE O ESTIMADOR

- Reinventado várias vezes:
  Alon/Matthias/Szégedy, Nemirovskii/Yudin,
  Jerrum/Sinclair ...

- Ressuscitado por Daniel Hsu em um blog
  a partir de trabalhos semelhantes de Catoni.

- Duas limitações: depende de δ e pede $\delta \geq \delta_{min} = e^{1-\frac{n}{8}}$

O QUE TEMOS × O QUE QUEREMOS

$\mathcal{P}$ dada, $\mathcal{P} \subset \mathcal{P}_2 = \{P : \sigma_P < +\infty\}$

(Temos)

$$\forall \delta \geq \delta_{min} = e^{1-n/8} \quad \exists\, \hat{E}_{n,\delta} : \mathbb{R}^n \to \mathbb{R}$$

$$\sup_{P \in \mathcal{P}} \mathbb{P}\left(\left|\hat{E}_n(X_1^n) - \mu_P\right| \geq L\sigma_P \sqrt{\frac{\ln(1/\delta)+1}{n}} \right) \leq \delta$$

(Queremos)

$$\exists\, \hat{E}_n : \mathbb{R}^n \to \mathbb{R} \quad \forall \delta \geq \delta_{min} \; (= \;?)$$

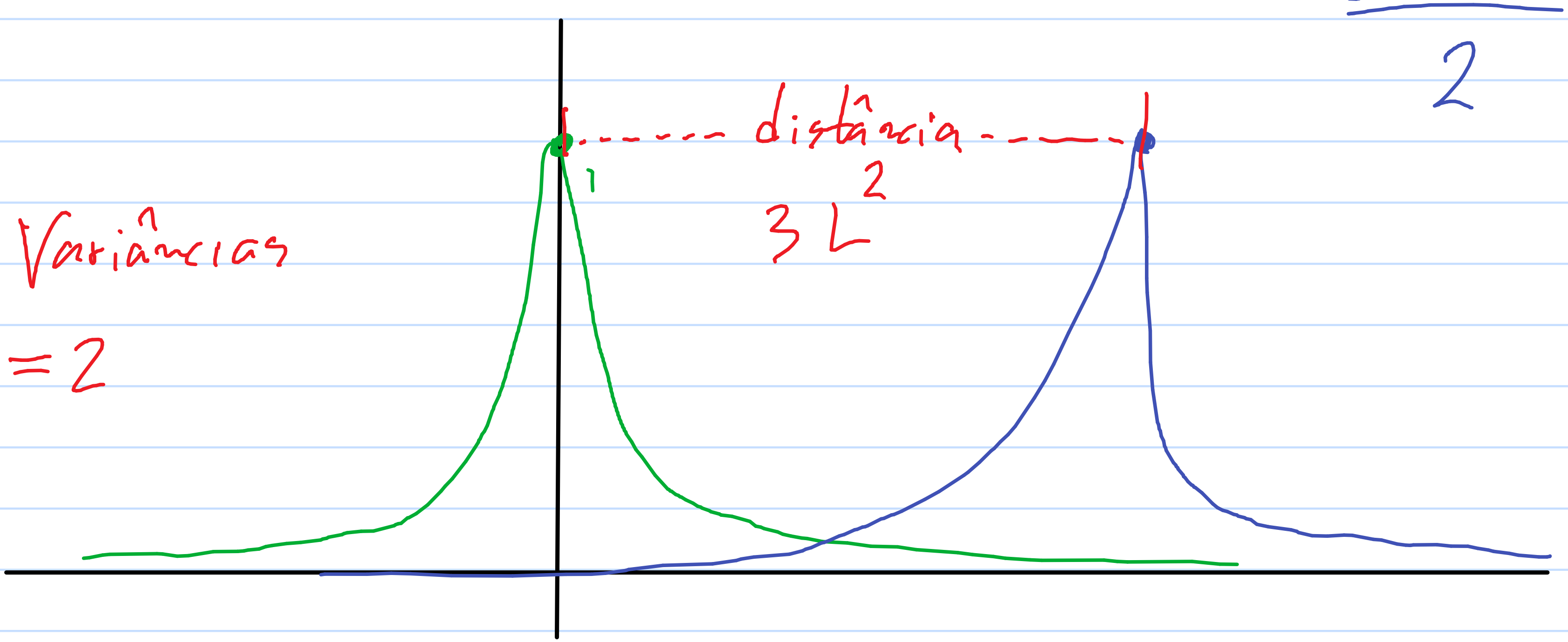
$$\sup_{P \in \mathcal{P}} \mathbb{P}\left(\left|\hat{E}_n(X_1^n) - \mu_P\right| \geq L\sigma_P \sqrt{\frac{\ln(1/\delta)+1}{n}} \right) \leq \delta$$

POR QUE É NECESSÁRIO QUE $\delta_{min} \geq e^{-cn}$.

P_0 = Laplace com média $\mu_0 = 0$ densidade $\frac{e^{-|x|}}{2}$

P_1 = Laplace com média $\mu_1 = 3L^2$ densidade $\frac{e^{-|x-\mu_1|}}{2}$

Variâncias $= 2$



O QUE OCORRE

$$p_0(x_1^n) = \text{densidade de } X_1^n = P_0^n = \frac{1}{2^n} e^{-\sum_{i=1}^n |x_i|}$$

$$p_1(x_1^n) = \text{densidade de } X_1^n = P_1^n = \frac{1}{2^n} e^{-\sum_{i=1}^n |x_i - 3\epsilon_n^2|}$$

- razão entre densidades $\in \left[\exp -(3\epsilon_n^2), \exp(3\epsilon_n^2)\right]$

- $\dfrac{\mathbb{P}_{P_0}(X_1^n \in E)}{\mathbb{P}_{P_1}(X_1 \in E)} \in \left[\exp(-3\epsilon_n^2), \exp(3\epsilon_n^2)\right]$

CONCLUSÃO : $\delta_{min} \geq e^{1-4L^2 n}$

Um estimador L-subgaussiano dependendo de δ para $\delta = \delta_{min} = \exp(1-5L^2 n)$ teria:

$$\mathbb{P}_{P_0}\left(\hat{E}_n(X_1^n) - \overset{0}{\mu_0} > \sqrt{2}L\underbrace{\sqrt{\frac{1+\ln\frac{1}{\delta}}{n}}}_{\text{menor que } 5L^2}\right) \leq \delta = e^{1-5L^2 n}$$

$$\mathbb{P}_{P_1}\left(\hat{E}_n(X_1^n) - \overset{10L^2}{\mu_1} > \sqrt{2}L\overbrace{\sqrt{\frac{1+\ln\frac{1}{\delta}}{n}}}\right) \geq 1-\delta$$

CONCLUINDO A CONCLUSÃO

... mas, se

$$\mathcal{E} = \{ \hat{E}_n(X_1^n) > 5L^2 \},$$

$$\mathbb{P}_{P_1}(\mathcal{E}) \geq 1 - \delta \Rightarrow \mathbb{P}_{P_0}(\mathcal{E}) \geq e^{-5L^2 n}(1-\delta)$$

$$\gg \delta$$

$$\Rightarrow \mathbb{P}_{P_0}\left(\hat{E}_n^1(X_1^n) > \mu_0 - \sqrt{2}\, L \sqrt{\frac{\ln 1/\delta + 1}{n}} \right) \gg \delta \;!$$

RESUMINDO

- Existe um estimador δ-dependente para o valor esperado que funciona para todas as P com $\sigma_P < +\infty$

- Provamos que os δ's permitidos devem ser $\geq \exp(1 - 5L^2 n)$ para qualquer família $\mathcal{P} \supseteq \{P_0, P_1\}$

- Pergunta: um mesmo estimador pode funcionar para todos os $\delta \in (\delta_{min}, 1/2)$?

PARTE 3

Estimadores para múltiplos δ's

A PARTIR DAQUI

[L. Devroye, M. Lerasle, G. Lugosi, R.I.O.

" Sub-Gaussian Mean Estimators"

Annals of Statistics, 2016]

PERGUNTA BÁSICA

- $\mathcal{P}$ = família de distribuições de probabilidade sobre $\mathbb{R}$ com $\sigma_P < +\infty$
- $n \in \mathbb{N} \setminus \{0\}$, $L > 0$, $\delta_{min} \in (0, 1/2)$ dados;

▶ Existe um $\hat{E}_n : \mathbb{R}^n \to \mathbb{R}$ tal que $\forall \delta \in (\delta_{min}, 1/2)$

$$\sup_{P \in \mathcal{P}} \mathbb{P}_P\left(|\hat{E}_n(X_1^n) - \mu_P| > L\sqrt{\frac{\ln(1/\delta) + 1}{n}} \right) \leq \delta$$

?

UM RESULTADO NEGATIVO

Teorema Fixe $L>0$ e tome $\delta_{min} = \delta_{min,n} \to 0$.

Tome ainda $\sigma_{1,n}^2$, $\sigma_{2,n}^2$ com $R_n = \frac{\sigma_{2,n}^2}{\sigma_{1,n}^2} \to +\infty$

Então não existe $\hat{E}_n$ com as propriedades

desejadas para a classe

$$\mathcal{P}_{2,\sigma_{1,n}^2,\sigma_{2,n}^2} = \{P : \sigma_{1,n}^2 \leq \sigma_P^2 \leq \sigma_{2,n}^2\}$$

se n é grande o suficiente.

Em particular, não há estimador de múltiplos δ para $\mathcal{P}_2$.

ELEMENTOS DA PROVA

- Reescalonando, podemos imaginar que

$$\left[\, \sigma_{1,n}^2 \ll \sigma_{2,n}^2 \ll 1 \,\right]$$

Considere

$$\begin{cases} P_0 = \text{Poisson com média } \sigma_{2,n} \\ P_1 = \text{Poisson com média } \sigma_{1,n} \end{cases}$$

(Variância = média p/ Poissons)

ESTATÍSTICA SUFICIENTE

▶ $S := \sum_{i=1}^{n} X_i$ é estatística suficiente p/ Poisson.

$$\left[\, \mathbb{P}_{P_i}\left(\hat{E}^1_n(X_1^n) \in I \mid S = k \right) \text{ independe de } i = 0 \text{ ou } 1 \,\right]$$

Por outro lado, como $P_0 = \text{Poisson}(\sigma^2_{2,n})$

& $P_1 = \text{Poisson}(\sigma^2_{1,n})$, $\dfrac{\sigma^2_{2,n}}{\sigma^2_{1,n}} = R_n >> 1$

$$\left[\, \mathbb{P}_{P_1}\left(S = \lceil \sigma^2_{2,n}\, n \rceil \right) \geq \exp(-c R_n \ln R_n)\, \mathbb{P}_{P_0}\left(S = \lceil \sigma^2_{2,n}\, n \rceil\right) \,\right]$$

CONCLUSÃO DO ESBOÇO DA PROVA

- Se existisse $\hat{E}_n : \mathbb{R}^n \to \mathbb{R}$ p/ múltiplos δ's.

$$\mathbb{P}_{P_0}\left(\hat{E}_n(X_1^n) \geq \frac{\sigma_{2,n}^2}{2}\right) = 1 - o(1)$$

$$\Rightarrow \mathbb{P}_{P_0}\left(\hat{E}_n(X_1^n) \geq \frac{\sigma_{2,n}^2}{2} \;\middle|\; S = \underbrace{\lceil \sigma_{2,n}^2\, n \rceil}_{\text{valor + provável de } S \text{ sob } P_0}\right) \geq \frac{1}{2}$$

$$\Rightarrow \mathbb{P}_{P_1}\left(\hat{E}_n(X_1^n) \geq \frac{\sigma_{2,n}^2}{2}\right) \geq \frac{1}{2}\, \mathbb{P}_{P_1}\left(S = \lceil \sigma_{2,n}^2 \cdot n \rceil\right)$$

$$\gg \sigma_{1,n}^2 = \sigma_{1,n} L \sqrt{\frac{\ln(1/\delta) + 1}{n}} \qquad \gtrsim \exp\left(-c\, R_n \ln R_n\right)$$

RESIGNAÇÃO E DEPRESSÃO ?

"O sonho acabou."

John Lennon

Não! Há resultados positivos a serem provados!

UM RESULTADO POSITIVO

$\mathcal{F}_{2,\sigma_1^2,\sigma_2^2} = \{ P : \sigma_1^2 \leq \sigma^2 \leq \sigma_2^2 \}$, $R = \dfrac{\sigma_2^2}{\sigma_1^2}$

Teorema $\exists\, L = L(R)$, $\delta_{min,n} = 4e^{1-\frac{n}{8}}$

$\forall n \in \mathbb{N}$, $n \geq 3$ $\exists\, \hat{E}_n : \mathbb{R}^n \to \mathbb{R}$ $\forall \delta \in \left(\delta_{min,n}, \frac{1}{2}\right)$

$$\left[\sup_{P \in \mathcal{F}} \mathbb{P}_P \left(|\hat{E}_n(X_1^n) - \mu_P| > L\sigma_P \sqrt{\frac{\ln(1/\delta) + 1}{n}} \right) \leq \delta \right]$$

QUAL É A DIFERENÇA?

Neste caso, podemos construir intervalos de confiança subgaussianos para cada $\delta \geq e^{1-n/8}$:

$$\hat{a}_{\pm}(X_1^n, \delta) := \hat{E}_{n,\delta}(X_1^n) \pm 2\sqrt{e}\,\sigma_2 \sqrt{\frac{1+\ln\frac{1}{\delta}}{n}}$$

$$\left[\hat{a}_{+}(X_1^n, \delta) - \hat{a}_{-}(X_1^n, \delta) \leq L_0(R)\,\sigma_P \sqrt{\frac{1+\ln\frac{1}{\delta}}{n}} \right.$$

$$\& \quad \mathbb{P}_P\left(\mu_P \notin [\hat{a}_{-}(X_1^n, \delta), \hat{a}_{+}(X_1^n, \delta)]\right) \leq \delta$$

$$\left. \forall P \in \mathcal{P}_{2,\sigma_1,\sigma_2}. \right]$$

O QUE FAREMOS COM INTERVALOS?

- Dados $\hat{I}_1, \hat{I}_2, \dots, \hat{I}_K$ intervalos com $\mathbb{P}(\mu_p \notin \hat{I}_k) \leq 2^{-k}$ $(1 \leq k \leq K)$, tomamos:

$$\begin{cases} \hat{K} = \min \{ k \leq K : \hat{I}_k \cap \hat{I}_{k+1} \cap \dots \hat{I}_K \neq \emptyset \} \\ \hat{E} = \text{ponto médio de } \hat{I}_{\hat{K}} \cap \hat{I}_{\hat{K}+1} \cap \dots \cap \hat{I}_K . \end{cases}$$

Afirmação: $\forall k \leq K : \mathbb{P}_P(|\hat{E} - \mu_p| > |\hat{I}_k|) \leq 2^{1-k}$

PROVA DA AFIRMAÇÃO

Evento bom $\quad \forall j \in \{h, h+1, \dots, K\} \quad \mu_p \in I_j$

$$\Rightarrow \quad I_h \cap I_{h+1} \cap \dots \cap I_K \ni \mu_p$$

$$\Rightarrow \quad \hat{K} \leq h \quad e \quad \mu_p, \hat{E} \in I_h$$

$$\Rightarrow \quad |\hat{E} - \mu_p| \leq |I_h| \quad ☆$$

$$☆ \quad \mathbb{P}_p(|\hat{E} - \mu_p| > |I_h|) = \mathbb{P}(\text{evento ruim})$$

$$= \mathbb{P}(\exists j \geq h : \mu_p \notin I_j) \leq \sum_{j \geq h} 2^{-j} = 2^{1-h}$$

NO NOSSO CASO

$$\hat{I}_k = \left[a_-(X_1^n, 2^{-k}),\ a_+(X_1^n, 2^{-k})\right]$$

com $$|\hat{I}_k| < L_0(R)\,\sigma_P \sqrt{\frac{1+\ln 2^k}{n}}$$

$$\Rightarrow \mathbb{P}_P\left(|\hat{E}_n(X_1^n) - \mu_P| > L_0'(R)\,\sigma_P \sqrt{\frac{1+k}{n}}\right) \leq 2^{1-k}$$

& escolhemos $k = 1 + \lceil \log_2 \frac{1}{\delta} \rceil$ p/ obter o resultado desejado.

OUTRA APLICAÇÃO DO MÉTODO

- P é dita **k-regular** se

$$\forall j \geq k: \quad \mathbb{P}_P\left(\frac{1}{j}\sum_{i=1}^{j} X_i \leq m_P\right) \geq \frac{1}{3}$$

$$\& \quad \mathbb{P}_P\left(\frac{1}{j}\sum_{i=1}^{j} X_i \geq m_P\right) \geq \frac{1}{3}$$

$$\left[\ \mathcal{P}_{k\text{-reg}} := \{ P : \sigma_P < +\infty,\ P\ k\text{-regular} \}\ \right]$$

Obs: $\bigcup_k \mathcal{P}_{k\text{-reg}} = \mathcal{P}_2$ pelo Teo Central do Limite!

EXEMPLOS

- $\mathcal{P}_{\alpha,K} := \{ P : \sigma_P < +\infty \ \& \ \mathbb{E}[(X-\mu_P)^\alpha] \leq K\sigma_P^\alpha \}$

  Contida em $\mathcal{P}_{k\text{-reg}}$ p/ $k \sim K^{\frac{2\alpha}{\alpha-2}}$

- $\mathcal{P}_{sym} := \{ P : \sigma_P < +\infty \ \& \ \forall x \in \mathbb{R} \ \ \mathbb{P}_P(X-\mu_P \leq x) = \mathbb{P}_P(\mu_P - X \leq x) \}$

  Contida em $\mathcal{P}_{1\text{-reg}}$.

RESULTADOS

$$\begin{cases} \delta_{min,n} := 4\exp\left(1 - \dfrac{n}{128k}\right) \\ \\ k \leq 326\, n \end{cases}$$

Teorema $\exists\, L > 0$ independente de k e $\exists\, \hat{E}_n : \mathbb{R}^n \longrightarrow \mathbb{R}$ $\forall\, \delta \in (\delta_{min,n}, 1/2)$

$$\left[\sup_{P \in \mathcal{P}_{h\text{-}reg}} \mathbb{P}_P\left(\left| \hat{E}_n(X_1^n) - \mu_P \right| > \frac{L\sigma_P}{\sqrt{n}} \sqrt{\ln\frac{1}{\delta} + 1} \right) \leq \delta \right]$$

ELEMENTOS DA PROVA

$$\left[b \sim \ln\left(\frac{1}{\delta}\right) \right]$$

$\left(b \text{ blocos} \right)$ $\quad \underbrace{X_1 \; X_2 \; \cdots \; X_k} \;\Big|\; \underbrace{X_{k+1} \; X_{k+2} \; \cdots \; X_{2k}} \;\Big|\; \underbrace{X_{2k+1} \; \cdots \; X_{3k}} \;\Big|\; \cdots \underbrace{X_{n-1} \; X_n} \;\Big|$

$\left(\text{médias dos blocos} \right)$ $\quad Y_1 = \frac{1}{m} \sum_{i=1}^{m} X_i \qquad Y_2 = \frac{1}{m} \sum_{i=m+1}^{2m} X_i \qquad Y_3 = \frac{1}{m} \sum_{i=2m+1}^{3m} X_i \quad \cdots \quad Y_b$

▶ Intervalo de confiança via quantis das médias.

$$a_-(X_1^n, \delta) = q_{1/4}(Y_1^b) \quad , \quad a_+(X_1^n, \delta) = q_{3/4}(Y_1^b)$$

PORQUE FUNCIONA

- P/ cada Y_i, se $\frac{n}{b} \geq k$ & P k-regular

$$\mathbb{P}(Y_i \leq m_P)$$

$$\Rightarrow \mathbb{P}_P\left(q_{1/4}(Y_1^b) > m_P\right)$$

$$= \mathbb{P}_P(\text{mais de } \tfrac{1}{4} \text{ dos } Y_i\text{'s à direita de } m_P)$$

$$\leq \mathbb{P}_P\left(\mathrm{Bin}\left(b, \tfrac{1}{3}\right) \leq \tfrac{b}{4}\right) \leq \exp(-c_1 b)$$

- Do mesmo modo $\mathbb{P}_P\left(q_{3/4}(Y_1^b) < m_P\right) \leq \exp(-c_1 b)$.

PORQUE FUNCIONA (II)

- Usando Chebyshev,

$$\mathbb{P}\left(|Y_i - \mu_P| \geq L_0 \sigma_P \sqrt{\frac{1+\ln \frac{1}{\delta}}{n}}\right) \leq \frac{1}{8}$$

$$\Rightarrow \mathbb{P}_P\left(q_{3/4}(Y_1^b) - q_{1/4}(Y_1^b) > L_0 \sigma_P \sqrt{\frac{1+\ln\frac{1}{\delta}}{n}}\right) \leq e^{-c_1 b}$$

$\therefore$ Com prob $\geq 1 - 3e^{-c_1 b}$,

$$q_{3/4}(Y_1^b) - q_{1/4}(Y_1^b) \leq L_0 \sigma_P \sqrt{\frac{1+\ln\frac{1}{\delta}}{n}} \quad \& \quad \mu_P \in [q_{1/4}, q_{3/4}]$$

PARTE 4

Ideias mais complexas !

ESTIMADORES "MUITO SUBGAUSSIANOS"?

$$\mathcal{P}_{4,K} = \{ P : \mathbb{E}_P[(X - \mu_P)^4] \leq K \sigma_P^2 \}$$

$$\delta_{min} \stackrel{\sim}{=} \exp\left(-o\left(n^{1/6}\right)\right)$$

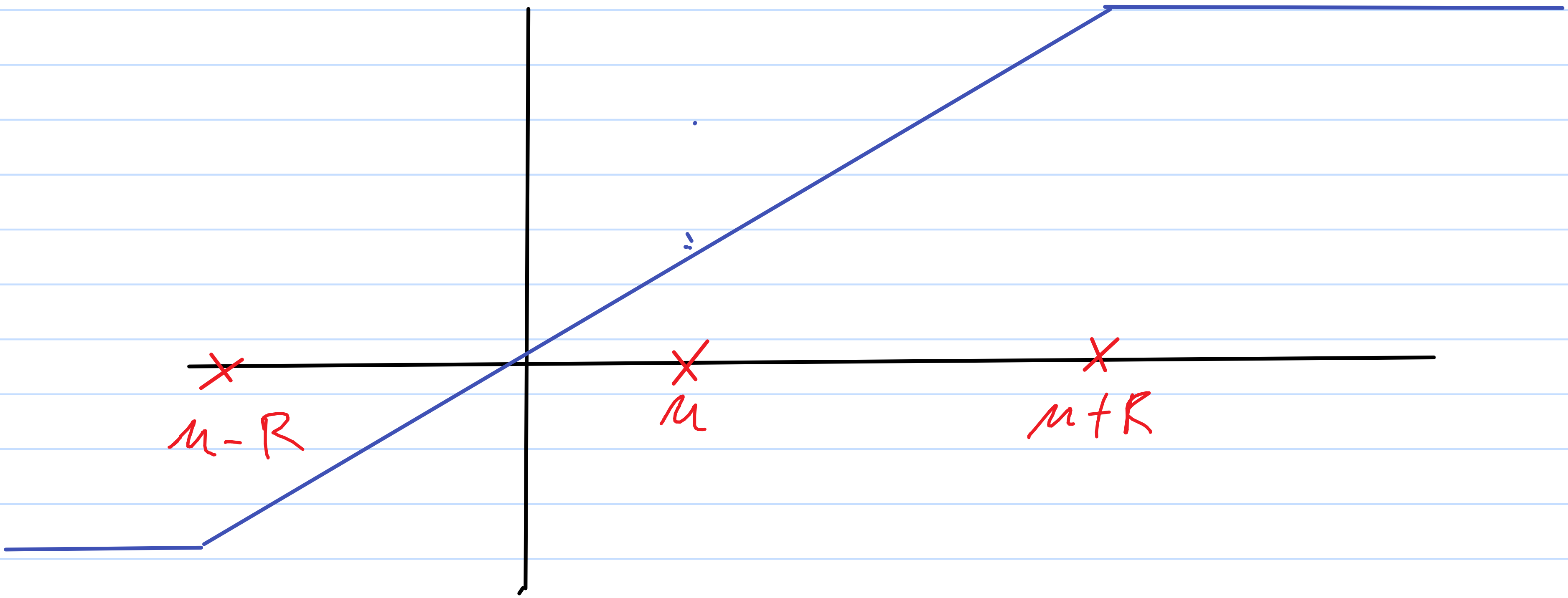
Teorema $\exists \hat{E}_n : \mathbb{R}^n \longrightarrow \mathbb{R} \quad \forall \delta \in (\delta_{min}, 1/2)$

$$\sup_{P \in \mathcal{P}_{4,K}} \mathbb{P}_P\left(|\hat{E}_n(X_1^n) - \mu_P| \geq \sigma_P \sqrt{2+\varepsilon} \sqrt{\frac{\ln 1/\delta}{n}} \right) \leq \delta$$

constante $\sqrt{2}$ é a melhor possível

IDEIAS DA PROVA

$$\Psi_{\mu,R}(X) = \begin{cases} \mu - R & , \quad X < \mu - R \\ X & , \quad X \in [\mu-R, \mu+R] \\ \mu + R & , \quad X > \mu + R \end{cases}$$



IDEIAS DA PROVA (II)

$$\left[\ \hat{E}_n(X_1^n) := \frac{1}{n} \sum_{i=1}^{n} \Psi_{\hat{\mu}_n,\, n^{1/6}\hat{\sigma}_n}(X_i)\ \right]$$

onde $\hat{\mu}_n$, $\hat{\sigma}_n$ são estimadores preliminares da média e variância.

▶ Fundamental Se $A := \left\{ \begin{array}{l} |\hat{\mu}_n - \mu_P| > \sigma_P/2 \text{ ou} \\ |\hat{\sigma}_n - \sigma_P| > \sigma_P/2 \end{array} \right\}$

$$\mathbb{P}_P\left(|\hat{E}_n(X_1^n) - \mu_P| \geq \sqrt{2+\varepsilon}\ \sigma_P \sqrt{\frac{\ln \frac{1}{\delta}}{n}}\right) \leq \delta - \delta^2 + \mathbb{P}_P(A).$$

A IDEIA É DECOMPOR O ERRO ...

$$\hat{M}_n(X_1^n, a, b) := \frac{1}{n} \sum_{i=1}^{n} \Psi_{a,b}(X_i)$$

$$\left| \hat{M}_n(X_1^n, \hat{\mu}_n, \hat{\sigma}_n n^{1/6}) - \mu_p \right|$$

(quando A não vale!)

$$\leq \left| \hat{M}_n(X_1^n, \mu_p, \sigma_p n^{1/6}) - \mu_p \right| +$$

$$\sup_{\substack{|\sigma - \sigma_p| \leq \sigma_p/2 \\ |\mu - \mu_p| \leq \sigma_p/2}} \left| \hat{M}_n(X_1^n, \mu, \sigma n^{1/6}) - \hat{M}_n(X_1^n, \mu_p, \sigma_p n^{1/6}) \right|$$

sempre que vale $A^c = \left\{ \begin{array}{l} |\hat{\mu}_n - \mu_p| \leq \sigma_p/2 \\ |\hat{\sigma}_n - \sigma_p| \leq \sigma_p/2 \end{array} \right\}$.

... E USAR ENCADEAMENTO

$$\sup_{\substack{|\sigma-\sigma_p|\leq \sigma_p/2 \\ |\mu-\mu_p|\leq \sigma_p/2}} \left| \hat{M}_n(X_1^n, \mu, \sigma n^{1/6}) - \hat{M}_n(X_1^n, \mu_p, \sigma_p n^{1/6}) \right|$$

controlado usando uma discretização para o espaço de μ's e σ's.

PARTE 5

Referências e problemas em aberto

ESTIMAÇÃO EM 1D

- Olivier Catoni, "Challenging the empirical mean..." Ann IHP Prob & Stat (2012)

  https://projecteuclid.org/euclid.aihp/1353098444

- O nosso artigo "Sub-Gaussian Mean Estimators" Ann Stat (2016)

  http://projecteuclid.org/euclid.aos/1479891632

Em aberto: constantes ótimas + "exatidão" na cauda

EXTENSÕES A VETORES

- Emilien Joly, Gábor Lugosi & R.I.O.
  "On the estimation of the mean of a random vector"
  Elec. Journal of Statistics (2017)
  http://projecteuclid.org/euclid.ejs/1488423803

- Gábor Lugosi, Shahar Mendelson
  "Sub-Gaussian estimates of the mean of a random vector"
  https://arxiv.org/abs/1702.00482

Em aberto: Estimadores para múltiplos δ, eficiência computacional.

REGRESSÃO

- Christian Brownlees, Emilien Joly & Gábor Lugosi
  "Empirical risk minimization for heavy-tailed losses"
  Annals of Statistics (2015)

  http://projecteuclid.org/euclid.aos/1444222083

- Gábor Lugosi, Shahar Mendelson
  "Risk minimization by median-of-means tournaments"

  https://arxiv.org/abs/1608.00757

- Guillaume Lecué & Matthieu Lerasle
  "Learning from MoM's principles" (sendo escrito)