COVID-19, Epidemiologia e Redes Sociais

Jose Storopoli, Alessandra Pellini e André Santos
josees@uni9.pro.br Universidade Nove de Julho - UNINOVE

Outline

1. LabCidades UNINOVE
2. Por que é dificil modelar e prever COVID-19?
3. Modelos Epidemiológicos Bayesianos
4. Classificador de Tweets com Aprendizagem de Máquina
5. Ferramentas e Métodos
6. Resultados Preliminares
7. Próximos Passos
8. Referências

Licença

LabCidades UNINOVE

- O Laboratório de Políticas Públicas e Cidades Inteligentes da UNINOVE possui como objetivo elaborar pesquisas para apoio à:
  - formulação de políticas públicas
  - tomada de decisão
- As principais técnicas são:
  - modelos estatísticos Bayesianos
  - algoritmos de aprendizagem de máquina
  - sistemas baseados em evidências
- Contexto de atuação:
  - Big Data
  - ferramentas *open source*
  - *Open science*: dados e código abertos para **replicabilidade** e **transparência**

LabCidades UNINOVE - Equipe

- Pesquisador Responsável: Jose Storopoli
- Pesquisador Associado: Alessandra Pellini
- Pesquisador Assistente: André Santos
- Pesquisador Assistente: Lorenzo Gottardi
- Alunos de Iniciação Científica:
  - João Vinícius Vieira Nóia
  - Elias Noda
  - Paula Fraga
  - Camila Brichta
  - Leandro dos Santos
  - Junior De Sousa Silva

Slides, códigos e dados disponíveis em `LabCidades/COVID-Classifier`

Por que é dificil modelar e prever COVID-19?

Primeiramente porque estamos fazendo errado (Ioannidis et al., 2020)

- input pobres de dados
- suposições incorretas de modelagem
- alta instabilidade de estimativas
- falta de incorporação de características epidemiológicas
- falta de transparência
- consideração de apenas uma ou algumas dimensões do problema em questão
- falta de *expertise*
- *groupthink* e efeitos *bandwagon*
- *selective reporting*

Por que é dificil modelar e prever COVID-19?

Mas alguns (não todos) desses problemas podem ser corrigidos (Ioannidis et al., 2020)

- Modelagem consciente de **distribuições preditivas** em vez de focar em estimativas pontuais
- considerando **dimensões múltiplas de impacto**
- reavaliando continuamente os modelos com base em seu **desempenho validado**

o Problema do R_t

A fórmula do R_t (número efetivo de reprodução) é:

$$R_t = R_0 \cdot \frac{S}{N}$$

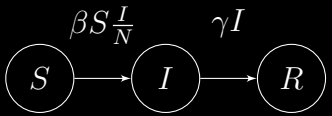
onde R_0 é o número básico de reprodução e $\frac{S}{N}$ é a proporção de suscetíveis S da população N.

Mas não sabemos os suscetíveis porque não temos testagem aleatórias para saber quantos N estão infectados e além disso aproximações de I por sintomas é complicado por conta da alta taxa de I assintomáticos (Gao et al., 2021):

- 30.8% - China, Wuhan (Nishiura et al., 2020)
- 51.7% Navio Diamond Princess (Mizumoto et al., 2020)

Modelos SIR

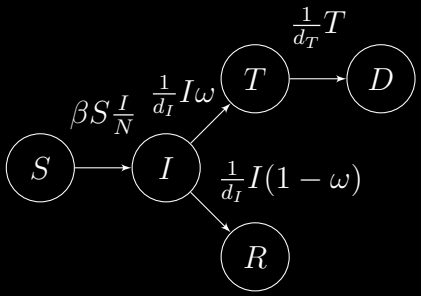
- Modelos compartimentais são usados para modelar a dinâmica de uma doença infecciosa em uma escala populacional
- Originados no início do século 20 com o modelo *Susceptible-Infectious-Recovered* (SIR) (Kermack & McKendrick, 1927)



Modelo SIR

Modelo SIRTD

- Algumas doenças infecciosas são fatais, portanto, para diferenciar entre os recuperados e os mortos, foi desenvolvido o modelo (SIRD) (Bailey et al., 1975)
- Como a COVID-19 pode superar rapidamente o sistema de saúde de uma nação (Pinto Neto et al., 2021), há a necessidade de incluir um estado que represente pacientes terminais T (SIRTD).



Modelo SIRTD

Modelo SIRTD

A dinâmica do SIRTD é governada por um sistema de equações diferenciais ordinárias:

- População $N = S + E + I + R + T + D$
- variação da quantidade por unidade de tempo dt
- taxa constante de contágio β
- taxa de mortalidade constante de indivíduos infectados ω
- tempo médio que os indivíduos estão infectantes d_I e em estado terminal d_T

$$\frac{dS}{dt} = -\beta S \frac{I}{N}$$
$$\frac{dI}{dt} = \beta S \frac{I}{N} - \frac{1}{d_I} I$$
$$\frac{dR}{dt} = \frac{1}{d_I} I \left(1 - \omega\right)$$
$$\frac{dT}{dt} = \frac{1}{d_I} I\omega - \frac{1}{d_T} T$$
$$\frac{dD}{dt} = \frac{1}{d_T} T$$

Especificação do Modelo[1]

Distribuições *a priori* dos parâmetros do modelo e função de verossimilhança que condiciona o compartimento D à quantidade de mortos observada:

$$\beta \sim \text{Normal}^{+}(\mu_\beta, \sigma_\beta)$$
$$\omega \sim \text{Beta}(\alpha_\omega, \beta_\omega)$$
$$d_I \sim \text{Normal}^{+}(\mu_{d_I}, \sigma_{d_I})$$
$$d_T \sim \text{Normal}^{+}(\mu_{d_T}, \sigma_{d_T})$$
$$\phi \sim \text{Exponencial}(\lambda_\phi)$$
$$\text{Mortos} \sim \text{Binomial Negativa}\left(\text{compart. } D, \frac{1}{\phi}\right)$$

[1] usamos o workflow de modelagem Bayesiana para modelos epidemiológicos de Grinsztajn et al. (2021)

Incorporação de Sintomas de Redes Sociais no Modelo

Incluímos dois novos parâmetros com suas distribuições *a priori* e uma nova função de verossimilhança que condiciona o compartimento I à quantidade de sintomas mencionados observados:

$$\begin{aligned}
\text{Taxa de Tweets} &\sim \text{Exponencial}(1) \\
\phi_{\text{tweets}} &\sim \text{Exponencial}(\lambda_{\phi_{\text{tweets}}}) \\
\text{Tweets} &\sim \text{Binomial Negativa}\left(\text{compart. } I \cdot \text{Taxa de Tweets}, \frac{1}{\phi_{\text{tweets}}}\right)
\end{aligned}$$

Podemos também partir do pressuposto que uma pessoa tuíta apenas uma vez por dia seu sintoma e colocar uma *priori* em Taxa de Tweets como uma Beta$(1, 1)$ e tornar isso uma proporção

Aprendizagem de Máquina

Aprendizagem de Máquina é uma área de estudo que fornece aos computadores a habilidade de aprender sem serem explicitamente programados (Mitchell, 1997).

Para medir o desempenho de um algoritmo de aprendizagem de máquina é preciso de uma medida de desempenho:

- Valor Preditivo Positivo[2]: $\frac{\text{Verdadeiro Positivos}}{\text{Verdadeiro Positivos} + \text{Falso Positivos}}$
- Sensibilidade: $\frac{\text{Verdadeiro Positivos}}{\text{Verdadeiro Positivos} + \text{Falso Negativos}}$
- Score F1[3]: $2 \cdot \frac{\text{Precisão} \cdot \text{Sensibilidade}}{\text{Precisão} + \text{Sensibilidade}}$

[2] conhecido também como precisão em aprendizagem de máquina

[3] média harmônica de precisão e sensibilidade

Aprendizagem de Máquina

Além disso separamos os dados em treino e teste, ou seja, dados que serão usados para treinar o algoritmo e dados que serão usados para testar o algoritmo.

Por que Redes Sociais?

Sabemos que é possível extrair informações sobre COVID-19 de Redes Sociais:

- Zong et al. (2020) apresentaram um corpus anotado de 7.500 tweets para eventos de COVID-19 demonstrando a possibilidade de identificar com precisão eventos de COVID-19 no Twitter
- Kaushal e Vaidhya (2020) treinaram um modelo de aprendizagem profunda de processamento de linguagem natural (PNL) para detectar eventos relacionados ao COVID-19 no Twitter
- Santosh et al. (2020) treinaram um modelo de aprendizagem profunda de processamento de linguagem natural (PNL) para detectar menção de sintomas do COVID-19 no Twitter

Todos sem extensões para a dinâmica da doença COVID-19 ou esforços de modelagem

Por que Twitter?

Simples, é o único com dados abertos que podemos coletar...

Além disso oferece três vantagens interessantes do ponto de vista metodológico (Gomez-Carrasco et al., 2020):

- as mensagens compartilhadas são públicas
- as postagens são limitadas ao número de caracteres, facilitando a identificação do conteúdo
- as funções são limitadas, o que simplifica a compreensão dos processos de comunicação

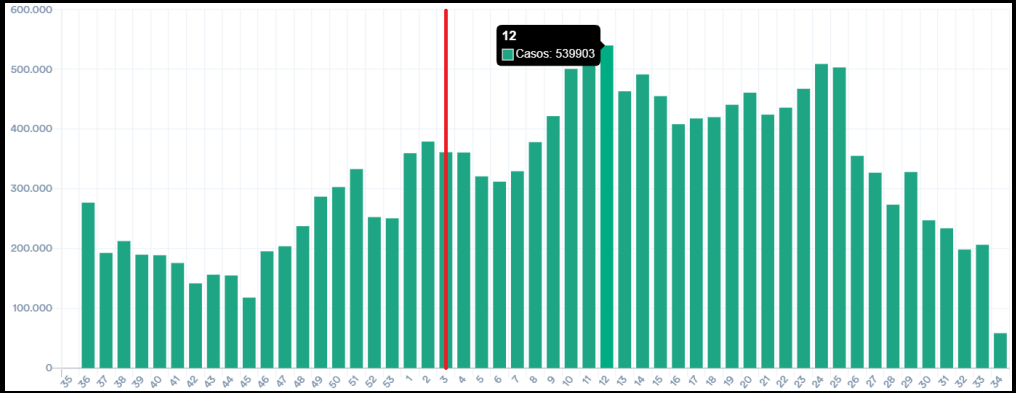
Anotação de Tweets

- Capacitação de alunos voluntários de iniciação científica
- Diretrizes e manual de rotulação formalizados em um documento online
- Acompanhamento semanal
- Três categorias (duas das quais foram baseadas no Guia de Vigilância Epidemiológica de COVID-19 (Brasil, 2021)):
  - `Síndrome Gripal (SG)`
  - `Síndrome Respiratória Aguda Grave (SRAG)`
  - `Sintomas generalizados`

Ferramentas Utilizadas

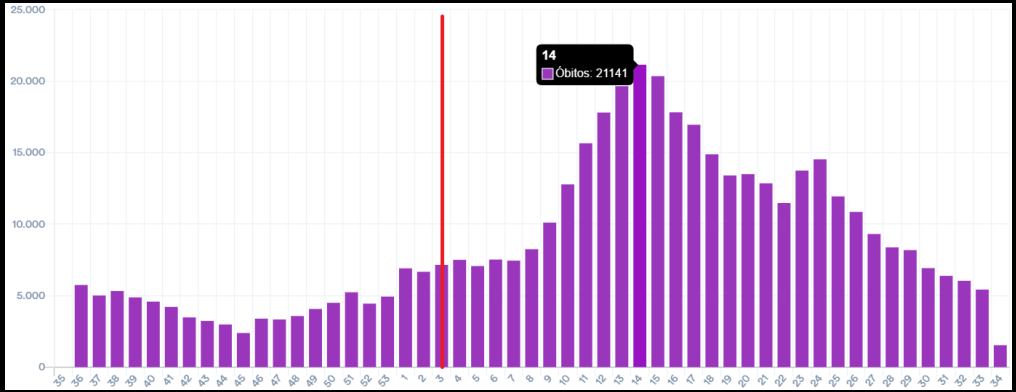
- Código: `GitHub`
- *Preprints*: `arXiv`
- Linguagens: `Julia` (Bezanson et al., 2017) e `Python`
- Modelos Bayesianos: `Stan` (Carpenter et al., 2017) e `Turing.jl` (Ge et al., 2018)
- Aprendizagem de Máquina: `MLJ.jl` (Blaom et al., 2020)
- Aprendizagem Profunda: `Flux.jl` (Innes, 2018)
- Equações Diferenciais: `DifferentialEquations.jl` (Rackauckas et al., 2021)
- Interface de Equações Diferenciais com Aprendizagem Profunda: `DiffEqFlux.jl` (Rackauckas et al., 2020)

Casos Novos - COVID-19



Fonte: Painel Coronavírus, Ministério da Saúde, 2021

Óbitos Novos - COVID-19



Fonte: Painel Coronavírus, Ministério da Saúde, 2021

Dados Coletados

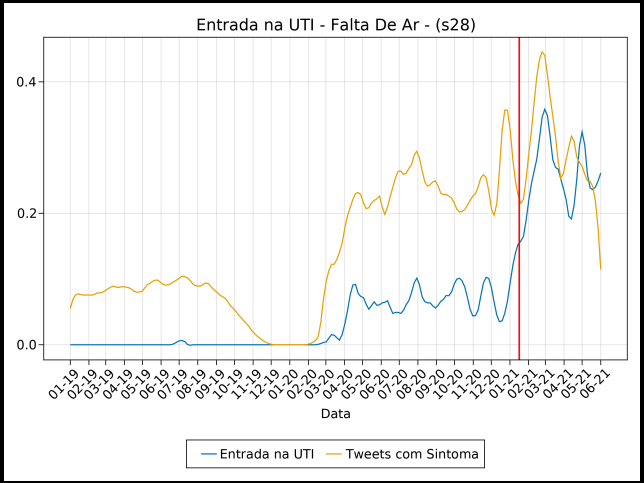
Twitter:

- Quase 14 milhões de tweets de Janeiro de 2019 à Julho de 2021 (dos Santos et al., 2021)
- Podem ser encontrados em 10.5281/zenodo.5073680
  - 2019: 4,043 milhões de linhas. 28 colunas. 1.1GB
  - 2020: 6,155 milhões de linhas. 28 colunas. 1.8GB
  - 2021: 3,657 milhões de linhas. 28 colunas. 1.1GB

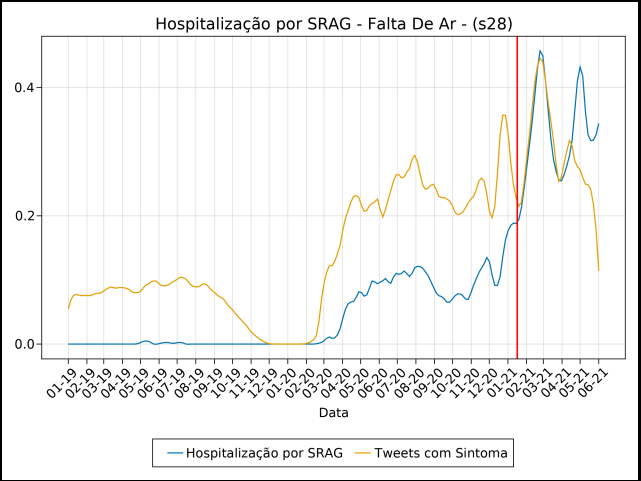
SRAG:

- As bases SRAG do SUS podem ser acessadas em `openDataSUS.saude.gov.br`:
  - 2019: 0,050 milhões linhas. 139 colunas. 23MB
  - 2020: 1,197 milhões de linhas. 154 colunas. 642MB
  - 2021: 1,330 milhões de linhas. 162 colunas. 711MB

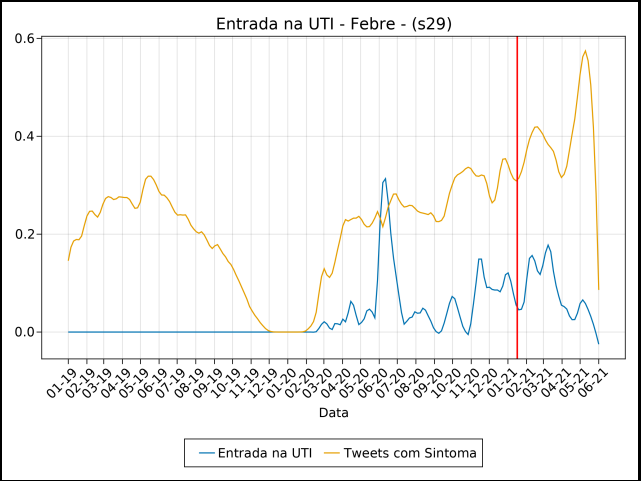
Tweets vs Base SRAG-SUS



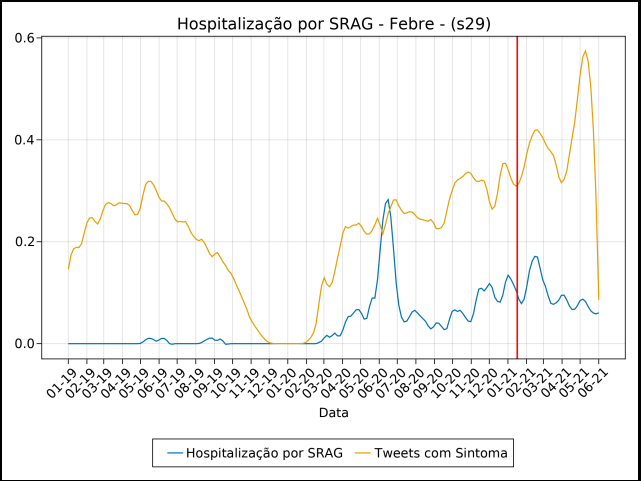
Tweets vs Base SRAG-SUS



Tweets vs Base SRAG-SUS



Tweets vs Base SRAG-SUS



Prova de Conceito - Paper no arXiv (Storopoli et al., 2021)

Com 9.600 tweets anotados, conseguimos acurácia de 90% no conjunto de dados de teste (quebra de 80%/20%)

rótulo	precisão[4]	sensibilidade[5]	score F1
0	0.94	0.93	0.93
1	0.75	0.80	0.78

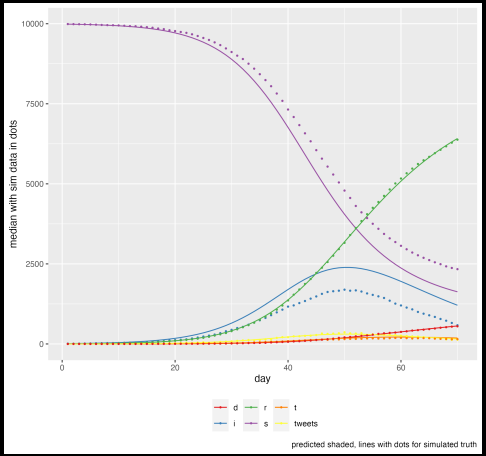
Resultados do Classificador de Sintomas de Tweets Brasileiros

[4]também chamada de valor preditivo positivo

[5]também conhecida como revocação ou *recall*

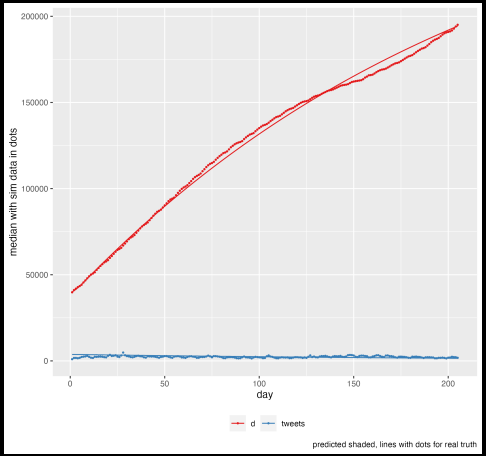
UNINOVE

Prova de Conceito - Paper no arXiv (Storopoli et al., 2021)



Dados Simulados versus Estimativas do Modelo

Prova de Conceito - Paper no arXiv (Storopoli et al., 2021)



Dados do Brasil 2020 versus Estimativas do Modelo

Próximos Passos - Classificador de Tweets

- Aprendizagem de máquina com diferentes modelos (Teorema do Almoço Grátis[6] (Wolpert, 1996))
- Aprendizagem profunda com redes neurais e técnicas de Processamento de Linguagem Natural (PLN) como por exemplo o BERT[7] (Devlin et al., 2018) já usado por Kaushal e Vaidhya (2020) e Santosh et al. (2020) em redes sociais no contexto de COVID-19.

[6] *No Free Lunch Theorem*

[7] Bidirectional Encoder Representations from Transformers

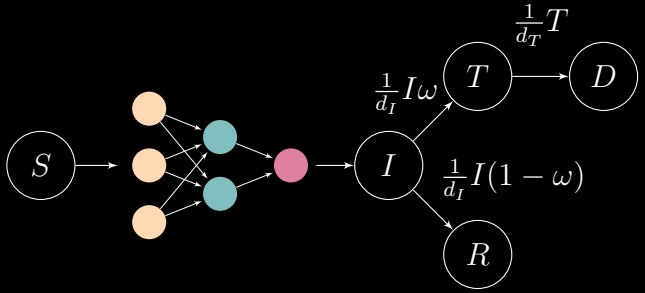
Próximos Passos - Modelo Epidemiológico Bayesiano

- incorporar sintomas do Twitter
- taxa de contágio variável β (exemplo em Jagan et al. (2020))
- tempo médio que os indivíduos estão em estado terminal variável d_T
- inserção de vacinação (ver seção 4.3 de Gleeson et al. (2021))
- hierarquização do modelo para estados/regiões do Brasil (complicados pressupostos de estratificação de tweets)

Próximos Passos - Rede Neural no Modelo Epidemiológico

Redes Neurais são aproximadores universais de funções (Zubov et al., 2021)

- Rede neural para prever a relação não-linear e complexa entre S e I (Pereira et al., 2021)
- Rede neural para prever a relação entre I e menções de sintomas



Modelo SIRTD com Rede Neural

Referências I

Bailey, N. T. et al. (1975). *The mathematical theory of infectious diseases and its applications*. Charles Griffin & Company Ltd, 5a Crendon Street, High Wycombe, Bucks HP13 6LE.

Bezanson, J., Edelman, A., Karpinski, S. & Shah, V. B. (2017). Julia: A Fresh Approach to Numerical Computing. *SIAM review*, *59*(1), 65–98. https://doi.org/10.1137/141000671

Blaom, A. D., Kiraly, F., Lienart, T., Simillides, Y., Arenas, D. & Vollmer, S. J. (2020). MLJ: A Julia package for composable machine learning. *Journal of Open Source Software*, *5*(55), 2704. https://doi.org/10.21105/joss.02704

Referências II

Brasil. (2021). Ministério da Saúde. Secretaria de Vigilância em Saúde. Guia de vigilância epidemiológica: emergência de saúde pública de importância nacional pela doença pelo coronavírus 2019 – COVID-19 / Ministério da Saúde, Secretaria de Vigilância em Saúde. – Brasília : Ministério da Saúde. 86 p. : il.. https://www.gov.br/saude/pt-br/coronavirus/publicacoes-tecnicas/guias-e-planos/guia-de-vigilancia-epidemiologica-covid-19/view

Carpenter, B., Gelman, A., Hoffman, M. D., Lee, D., Goodrich, B., Betancourt, M., Brubaker, M., Guo, J., Li, P. & Riddell, A. (2017). Stan : A Probabilistic Programming Language. *Journal of Statistical Software*, *76*(1). https://doi.org/10.18637/jss.v076.i01 2846 citations (Semantic Scholar/DOI) [2021-02-13]

Referências III

Devlin, J., Chang, M.-W., Lee, K. & Toutanova, K. (2018, outubro 10). *BERT: Pre-Training of Deep Bidirectional Transformers for Language Understanding*. arXiv: 1810.04805. Obtido 13 setembro 2020, de http://arxiv.org/abs/1810.04805
9991 citations (Semantic Scholar/arXiv) [2021-02-13]

dos Santos, A. L. M. F., Pellini, A. C. G., Noda, E. S., Nóia, J. V. V., Fraga, P., Brichta, C., dos Santos, L. R., Silva, J. D. S. & Storopoli, J. E. (2021, julho 1). *Brazilian Portuguese COVID-19 Tweets*. Zenodo. https://doi.org/10.5281/zenodo.5073680

Gao, Z., Xu, Y., Sun, C., Wang, X., Guo, Y., Qiu, S. & Ma, K. (2021). A systematic review of asymptomatic infections with COVID-19. *Journal of Microbiology, Immunology and Infection*, *54*(1), 12–16. https://doi.org/10/ggx2zb

Referências IV

Ge, H., Xu, K. & Ghahramani, Z. (2018). Turing: A Language for Flexible Probabilistic Inference. *International Conference on Artificial Intelligence and Statistics*, 1682–1690. Obtido 20 fevereiro 2021, de http://proceedings.mlr.press/v84/ge18b.html

Gelman, A., Carlin, J. B., Stern, H. S., Dunson, D. B., Vehtari, A. & Rubin, D. B. (2013). *Bayesian Data Analysis*. Chapman and Hall/CRC.

Gleeson, J. P., Murphy, T. B., O'Brien, J. D., Friel, N., Bargary, N. & O'Sullivan, D. J. P. (2021, junho 8). *Calibrating COVID-19 SEIR Models with Time-Varying Effective Contact Rates*. arXiv: 2106.04705 [physics, q-bio]. Obtido 25 junho 2021, de http://arxiv.org/abs/2106.04705

Referências V

Gomez-Carrasco, P., Guillamon-Saorin, E. & Osma, B. G. (2020). Stakeholders versus Firm Communication in Social Media: The Case of Twitter and Corporate Social Responsibility Information. *European Accounting Review*, *30*(1), 31–62. https://doi.org/10.1080/09638180.2019.1708428

Grinsztajn, L., Semenova, E., Margossian, C. C. & Riou, J. (2021, fevereiro 4). *Bayesian Workflow for Disease Transmission Modeling in Stan*. arXiv: 2006.02985 [q-bio, stat]. Obtido 5 junho 2021, de http://arxiv.org/abs/2006.02985

Innes, M. (2018). Flux: Elegant Machine Learning with Julia. *Journal of Open Source Software*. https://doi.org/10.21105/joss.00602

Referências VI

Ioannidis, J. P. A., Cripps, S. & Tanner, M. A. (2020). Forecasting for COVID-19 has failed. *International Journal of Forecasting*. https://doi.org/10.1016/j.ijforecast.2020.08.004

Jagan, M., deJonge, M. S., Krylova, O. & Earn, D. J. D. (2020). Fast estimation of time-varying infectious disease transmission rates. *PLOS Computational Biology*, *16*(9), e1008124. https://doi.org/10/gmg75f

Kaushal, A. & Vaidhya, T. (2020). Winners at W-NUT 2020 Shared Task-3: Leveraging Event Specific and Chunk Span information for Extracting COVID Entities from Tweets. *Proceedings of the Sixth Workshop on Noisy User-generated Text (W-NUT 2020)*. https://doi.org/10.18653/v1/2020.wnut-1.79

Referências VII

Kermack, W. O. & McKendrick, A. G. (1927). A contribution to the mathematical theory of epidemics. *Proceedings of the royal society of london. Series A, Containing papers of a mathematical and physical character*, *115*(772), 700–721.

McElreath, R. (2020). *Statistical Rethinking: A Bayesian Course with Examples in R and Stan*. CRC press.

Mitchell, T. M. (1997). *Machine Learning*. McGraw Hill.

Mizumoto, K., Kagaya, K., Zarebski, A. & Chowell, G. (2020). Estimating the asymptomatic proportion of coronavirus disease 2019 (COVID-19) cases on board the Diamond Princess cruise ship, Yokohama, Japan, 2020. *Eurosurveillance*, *25*(10), 2000180.

Referências VIII

Nishiura, H., Kobayashi, T., Miyama, T., Suzuki, A., Jung, S.-m., Hayashi, K., Kinoshita, R., Yang, Y., Yuan, B., Akhmetzhanov, A. R. et al. (2020). Estimation of the asymptomatic ratio of novel coronavirus infections (COVID-19). *International journal of infectious diseases*, *94*, 154.

Pereira, F. H., Schimit, P. H. T. & Bezerra, F. E. (2021). A deep learning based surrogate model for the parameter identification problem in probabilistic cellular automaton epidemic models. *Computer Methods and Programs in Biomedicine*, *205*, 106078. https://doi.org/10/gmg58g

Referências IX

Pinto Neto, O., Kennedy, D. M., Reis, J. C., Wang, Y., Brizzi, A. C. B., Zambrano, G. J., de Souza, J. M., Pedroso, W., de Mello Pedreiro, R. C., de Matos Brizzi, B., Abinader, E. O. & Zângaro, R. A. (2021). Mathematical Model of COVID-19 Intervention Scenarios for São Paulo—Brazil. *Nature Communications*, *12*(1), 418. https://doi.org/10.1038/s41467-020-20687-y

Rackauckas, C., Anantharaman, R., Edelman, A., Gowda, S., Gwozdz, M., Jain, A., Laughman, C., Ma, Y., Martinuzzi, F., Pal, A., Rajput, U., Saba, E. & Shah, V. B. (2021, maio 12). *Composing Modeling and Simulation with Machine Learning in Julia*. arXiv: 2105.05946 [cs]. Obtido 28 junho 2021, de http://arxiv.org/abs/2105.05946

Referências X

Rackauckas, C., Ma, Y., Martensen, J., Warner, C., Zubov, K., Supekar, R., Skinner, D., Ramadhan, A. & Edelman, A. (2020, agosto 6). *Universal Differential Equations for Scientific Machine Learning*. arXiv: 2001.04385 [cs, math, q-bio, stat]. Obtido 23 março 2021, de http://arxiv.org/abs/2001.04385

Santosh, R., Schwartz, H. A., Eichstaedt, J. C., Ungar, L. H. & Guntuku, S. C. (2020). Detecting Emerging Symptoms of COVID-19 using Context-based Twitter Embeddings.

Storopoli, J., dos Santos, A. L. M. F., Pellini, A. C. G. & Baldwin, B. (2021, junho 22). *Simulation-Driven COVID-19 Epidemiological Modeling with Social Media*. arXiv: 2106.11686 [stat]. Obtido 24 junho 2021, de http://arxiv.org/abs/2106.11686

Referências XI

Wolpert, D. H. (1996). The Lack of a Priori Distinctions between Learning Algorithms. *Neural Computation*, *8*(7), 1341–1390. https://doi.org/10.1162/neco.1996.8.7.1341 691 citations (Semantic Scholar/DOI) [2021-02-13]

Zong, S., Baheti, A., Xu, W. & Ritter, A. (2020, junho 24). *Extracting COVID-19 Events from Twitter*. arXiv: 2006.02567 [cs]. Obtido 19 junho 2021, de http://arxiv.org/abs/2006.02567

Zubov, K., McCarthy, Z., Ma, Y., Calisto, F., Pagliarino, V., Azeglio, S., Bottero, L., Luján, E., Sulzer, V., Bharambe, A., Vinchhi, N., Balakrishnan, K., Upadhyay, D. & Rackauckas, C. (2021, julho 19). *NeuralPDE: Automating Physics-Informed Neural Networks (PINNs) with Error Approximations*. arXiv: 2107.09443 [cs]. Obtido 24 julho 2021, de http://arxiv.org/abs/2107.09443

Backup Slides

Estatística Bayesiana

A estatística Bayesiana é uma abordagem de **análise de dados baseada no teorema de Bayes, onde o conhecimento disponível sobre os parâmetros em um modelo estatístico é atualizado com as informações dos dados observados** (Gelman et al., 2013; McElreath, 2020). O conhecimento prévio é expresso como uma distribuição *a priori* e combinado com os dados observados na forma de uma função de verossimilhança para determinar a distribuição posterior. A posterior também pode ser usada para fazer previsões sobre eventos futuros.

Inferência Bayesiana

$$\underbrace{P(\theta \mid y)}_{\text{Posterior}} = \frac{\overbrace{P(y \mid \theta)}^{\text{Verossimilhança}} \cdot \overbrace{P(\theta)}^{\textit{Priori}}}{\underbrace{P(y)}_{\text{Constante Normalizadora}}}$$

- θ – parâmetro(s) de interesse
- y – dados observados
- ***Priori***: probabilidade prévia do valor do(s) parâmetro(s)
- **Verossimilhança**: probabilidade dos dados observados condicionados aos valores do(s) parâmetro(s)
- **Posterior**: probabilidade posterior do valor do(s) parâmetros após observamos os dados y
- **Constante Normalizadora**: $P(y)$ não faz sentido intuitivo. Essa probabilidade é transformada e pode ser interepretada como algo que existe apenas para que o resultado de $P(y \mid \theta)P(\theta)$ seja algo entre 0 e 1 – uma probabilidade válida.

Teorema de Bayes como Motor de Inferência

A estatísica Bayesiana nos permite **quantificar diretamente a incerteza** relacionada ao valor de um ou mais parâmetros do nosso modelo condicionado aos dados observados. Isso é a **característica principal** da estatística Bayesiana. Pois estamos estimando diretamente $P(\theta \mid y)$ por meio do teorema de Bayes. A estimativa resultante é totalmente intuitiva: simplesmente quantifica a intercerteza que temos sobre o valor de um ou mais parâmetro condicionado nos dados, nos pressupostos do nosso modelo (verossimilhança) e na probabilidade prévia que temos sobre tais valores.

O que muda da Estatística Frequentista?

- **Flexibilidade** - peças probabilísticas para construir um modelo[8]:
  - Conjecturas probabilísticas sobre os parâmetros:
    - *Priori*
    - Verossimilhança
- Melhor tratamento da **incerteza**:
  - Coerência
  - Propagação
  - Não se usa *"se amostrássemos infinitamente de uma população que não existe..."*
- Sem p**-valores**:
  - Todas as intuições estatísticas fazem **sentido**
  - 95% de certeza que o valor do parâmetro θ está entre x e y
  - Quase **impossível** fazer p-hacking.

[8] como se fosse LEGO

Estatística Bayesiana vs Frequentista

	Estatística Bayesiana	Estatística Frequentista
Dados	Fixos — Não Aleatórios	Incertos — Aleatórios
Parâmetros	Incertos — Aleatórios	Fixos — Não Aleatórios
Inferência	Incerteza sobre o valor do parâmetro	Incerteza sobre um processo de amostragem de uma população infinita
Probabilidade	Subjetiva	Objetiva (mas com diversos pressupostos dos modelos)
Incerteza	Intervalo de Credibilidade — $P(\theta \mid y)$	Intervalo de Confiança — $P(y \mid \theta)$

Vantagens da Estatística Bayesiana

- Abordagem Natural para expressar Incerteza
- Habilidade de incorporar Informações Prévias
- Maior Flexibilidade do Modelo
- Distribuição Posterior completa dos Parâmetros
- Propagação Natural da Incerteza

Principal Desvantagem: Velocidade lenta de estimativa de modelos[9]

[9] *e.g.* 30 segundos ao invés de 3 segundos na abordagem frequentista