IN2P3 School of Statistics 2021

Yann Coadou
for the SOS organising committee

CPPM Marseille
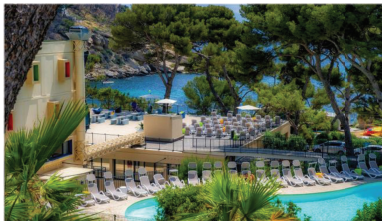
IN2P3/IRFU Machine Learning workshop
Zoom, 17 March 2021

SOS 2020

We hope to welcome you in May at Carry !

Further information and registration date will be sent soon and will appear on the indico page

https://indico.in2p3.fr/e/sos2020

Sabine Crépé-Renaudin   SOS 2020   22 janvier 2020   8

SOS 2020

We hope to welcome you in May at Carry !

Further information and registration date will be sent soon and will appear on the indico page

https://indico.in2p3.fr/e/sos2020

Sabine Crépé-Renaudin   SOS 2020   22 janvier 2020   8

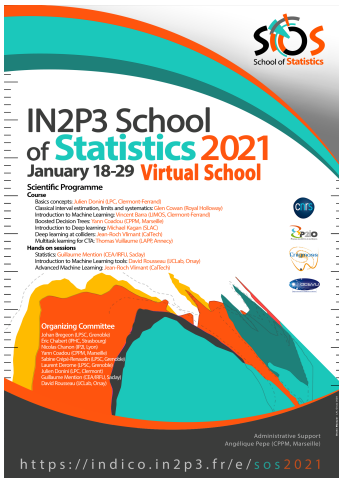- And then, COVID-19 happened. . .

7th edition of the IN2P3 School of Statistics

- 18–29 January 2021 (two weeks, mornings only)
- Using Zoom and Discord (for permanent chat)
- Attendance: PhD students, postdocs, undergrads, staff
- In English, also attracting foreigners
- Indico: ▸ https://indico.in2p3.fr/e/SOS2021
- General web site: ▸ http://sos.in2p3.fr
- Financial support (not necessary in the end):

- **Origin:**
  - strong demand from IN2P3 community for statistical analysis training
  - usually not much developed during initial training
  - nowadays key ingredient of scientific results production
- **Goal:** *describe, explain and manipulate statistical concepts and tools necessary to perform statistical analyse of data in particle physics, astroparticle physics and cosmology*
  - proper usage of tools
  - real impact in day-to-day work
  - competitiveness and visibility in our big collaborations
- **Target audience:** PhD students, postdocs, staff (junior and senior) interested in basic concepts and new tools
- 7th edition, first (last?) fully online one

2008	Strasbourg (67)
2010–2016	Autrans (38)
2018	La Londe-les-Maures (83)
2020 → 2021	Zoom

- Johan Bregeon (LPSC, Grenoble)
- Éric Chabert (IPHC, Strasbourg)
- Nicolas Chanon (IP2I, Lyon)
- Yann Coadou (CPPM, Marseille) –chair–
- Sabine Crépé-Renaudin (LPSC, Grenoble)
- Laurent Derome (LPSC, Grenoble)
- Julien Donini (LPC, Clermont-Ferrand)
- Romain Madar (LPC, Clermont-Ferrand)
- Guillaume Mention (CEA/IRFU, Saclay)
- David Rousseau (IJClab, Orsay)

Administrative support

- Angélique Pèpe (CPPM, Marseille)

School programme (mornings for two weeks)

- **Three main components:**
  - fundamental concepts (consolidate basic knowledge, understand ML)
  - multivariate analysis and machine learning (increased in recent sessions)
  - hands-on sessions (also increased)
- **Time for discussion:**
  - one of the core concepts of the school: more difficult online...
    $\Rightarrow$ used Zoom and Discord server

	Mon 18	Tue 19	Wed 20	Thu 21	Fri 22
9:00-10:30	Basic concepts I	Intervals & limits I	Hands-on: basics (starts at 10:00)	Machine learning	Boosted decision trees
10:30-10:45	Break				
10:45-12:15	Basic concepts II	Intervals & limits II			

	Mon 25	Tue 26	Wed 27	Thu 28	Fri 29
9:00-10:30	Hands-on: basics machine learning	Deep learning I	Deep learning at colliders	Hands-on: advanced machine learning	Multitask learning
10:30-10:45		Break			Farewell
10:45-12:15		Deep learning II			

All times are morning CET

Similar to 1-week schedule with morning/afternoon

SOS 2021 teachers

Lectures

- Julien Donini (LPC, Clermont): Basic concepts
- Glen Cowan (Royal Holloway): Intervals & limits
- Vincent Barra (LIMOS, Clermont-Ferrand): Machine learning
- Yann Coadou (CPPM, Marseille): Boosted decision trees
- Michael Kagan (SLAC): Deep learning
- Jean-Roch Vlimant (CalTech): Deep learning at colliders
- Thomas Vuillaume (LAPP, Annecy): Multitask learning

Hands-on sessions

- Guillaume Mention (CEA/IRFU, Saclay): basics of statistics
- David Rousseau (IJClab, Orsay): basics machine learning
- Jean-Roch Vlimant (CalTech): advanced machine learning

Machine learning @ SOS 2021

- Complete overview of founding principles
- Boosted decision trees / TMVA (current work horse of LHC analyses)
- Deep learning (with very useful technical details):
  - basics of neural networks: empirical risk, gradient descent, automatic differentiation, backpropagation
  - DNN
  - CNN
  - RNN / LSTM
  - autoencoders
  - generative models (GAN, VAE)
- Deep learning at colliders: handling data (acquire, compress, clean, etc.), generative models, image/graph representation, decorrelation, anomaly search
- Multitask learning: CTA example, learn energy/direction/classification simultaneously
- Hands-on!

  - Advice: watch video recording of lectures
    - https://indico.in2p3.fr/e/SOS2021
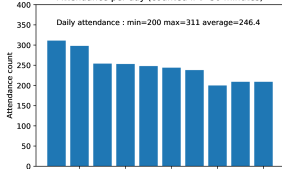
Hands-on sessions

- All done with Jupyter notebooks, all available
- Can run locally (install anaconda) or on Google Colab
- **Basics of statistics:**
    - basic python packages (NumPy, SciPy, MatPlotLib)
    - probabilities
    - histograms
    - covariance
    - fitting
- **Basics machine learning**
    - dataset handling, features
    - BDT with XGBoost, LightGBM
    - optimisation
    - variable selection (feature vs. permutation importance)
- **Advanced machine learning**
    - model quantisation
    - anomaly detection
    - tracking with graph NN
    - likelihood free inference

- 450 registered participants!
- 355 connected to Zoom at some point
- 250 on Zoom on average (slow decrease over time, Zoom fatigue)
- Most people connected for full Lecture
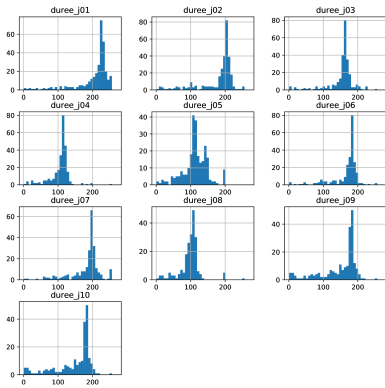- Zoom statistics used to deliver certificate of attendance

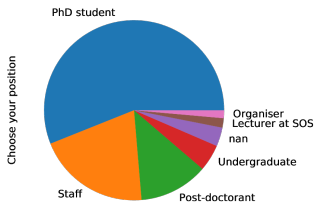Distribution of connection duration (limited to 3h per day)

Total attendance :355 persons
Total duration :27.8 h
Mean attendance duration : 17.0 h
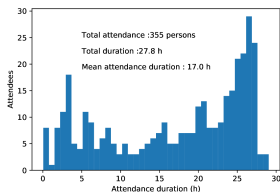
Attendance per day (counted if t>30 minutes)

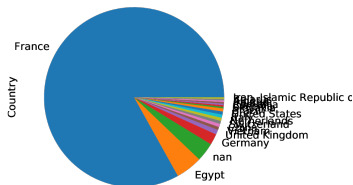Daily attendance : min=200 max=311 average=246.4
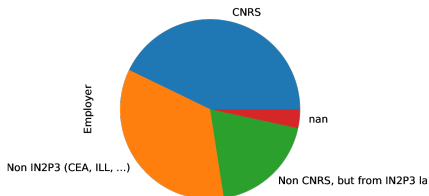

attendance (minutes) each day

position

Connecting from country

Employer

- 83% from France, 5% from Egypt (?), 2% Germany
- 56% PhD, 20% staff, 12% postdoc, 5% undergrad
- 42% CNRS, 35% Non-IN2P3 (mostly CEA), 19% non-CNRS but IN2P3

- Originally planed in May 2020
- Hope/dream for some time to run after first lockdown
- Finally held fully online in January 2021 over two weeks (mornings only)
- Very successful despite circumstances: 450 registrants, 355 people connected, 200 certificates
- A lot of useful feedback from survey (150 answers)
- All material (slides, hands-on Jupyter notebooks, video recordings) publicly available at ▸ https://indico.in2p3.fr/e/SOS2021
- People seemed pleased:

• Thanks! It was an interesting journey, I knew maybe 10% of these concepts! It was very fruitful
• I get back home (virtually, at least) with many new tools I want to play with.
• I would like to express my profound appreciation for all the organizers for putting up such a wonderful school. I learnt a lot!
• I really appreciate the time and insight that went into planning and holding the school.
• Tous les présentateurs ont su être clairs et pédagogues malgré le challenge du grand nombre de participants, ce que j'ai beaucoup apprécié.

- Most likely going back to original model:
  - in person
  - one week in a relatively isolated place to foster discussions
  - probably in May 2022

- Further information and registration date to be sent in a few months

School of **Statistics**

School of **Statistics**