# Sally Paganin, Ph.D.

CONTACT INFORMATION

Department of Biostatistics
Harvard T.H. Chan School of Public Health Building 2, Room 451
655 Huntington Avenue, Boston, Massachusetts 02115.
✉ spaganin@hsph.harvard.edu     Personal website https://salleuska.github.io/
ORCID   0000-0002-3451-1795     �

 salleuska   🎓 Google Scholar

RESEARCH INTERESTS

Bayesian nonparametrics, Computational Statistics, Hierarchical models, Item Response Theory, Partition Models, Models for Latent Variables, Data visualization.

CURRENT POSITION

**Research Fellow** at Harvard T.H. Chan School of Public Health (since March 2021) Department of Biostatistics, working with Jeff Miller on statistical methods for early cancer detection.

PAST POSITIONS

**Postdoctoral Researcher** at UC Berkeley (September 2019 – March 2021) Department of Statistics, Department of Environmental Science, Policy & Management statistical methodology, applications and algorithms development in NIMBLE, collaborating with Perry de Valpine and Chris Paciorek.

EDUCATION

**Università degli Studi di Padova**, Padova, Italy
Department of Statistical Sciences

2019  Ph.D. in Statistical Sciences

Thesis: Prior-driven cluster allocation in Bayesian mixture models
Advisor: Bruno Scarpa; Co-Advisor Amy H. Herring

2016  M.S. in Statistical Sciences

Thesis: Analysis of an insurance company sales structure: a Bayesian nonparametric model for a network of networks
Advisor: Bruno Scarpa; Co-advisor: Daniele Durante

2012  B.S. in Statistical Sciences and Information Technologies

Thesis: Bayesian nonparametric models based on Dirichlet Process
Advisor: Livio Finos; Co-advisor: Dario Solari

AWARDS

- Young researcher travel award, 2022 ISBA world meeting

- 2022 PDA Research Symposium honorarium recipient (HSPH)

- Bayesian Analysis Discussion Paper of March 2021 issue (Editorial Board's choice)

- Young researcher travel award, 2018 ISBA world meeting

- Travel award for COBAL V (2017), Guanajuato, Mexico

- Young researcher travel award, 2016 ISBA world meeting

PUBLICATIONS

(*) authors listed in alphabetical order
**Publications in refereed journals**

4. **Paganin, S.**, Paciorek, C., Wehrhahn, C., Rodriguez, A., Rabe-Hesketh, S., de Valpine, P. (2022) Computational methods for Bayesian semiparametric item response theory models using NIMBLE. Accepted at *Journal of Educational and Behavioral Statistics* (arXiv preprint.)

3. de Valpine, P., **Paganin, S.**, Turek, D. (2022). compareMCMCs: An R package for studying MCMC efficiency. *Journal of Open Source Software*, 7(69), 3844, doi:10.21105/joss.03844

2. **Paganin, S.**, Herring, A.H., Olshan A.F., Dunson, D.B. and The National Birth Defect Prevention Study. (2021) Centered Partition Process: Informative Priors for Clustering (with Discussion). *Bayesian Analysis*, 16(1), 301–370. doi:10.1214/20-BA1197

1. Durante, D., **Paganin, S.**, Scarpa, B. and Dunson, D.B. (2017) Bayesian modeling of networks in complex business intelligence problems. *Journal of the Royal Statistical Society: Series C* 66, 555–580

### Edited Books

(*) Argiento, R., Camerlenghi, F., Paganin, S. [Eds.] (2022) New Frontiers in Bayesian Statistics. *BaYSM 2021, Online, September 1–3*. Springer Proceedings in Mathematics and Statistics, Springer.

### Book Chapters and Discussions

(*) Aliverti, E., Paganin, S., Rigon, T. and Russo, M. (2019) Contributed discussion to "Latent nested nonparametric priors". *Bayesian Analysis* 14, 4, 1303–1356

(*) Aliverti E., Forastiere L., Padellini T., Paganin S. and Wit E. (2018). Hierarchical Graphical Model for Learning Functional Network Determinants. *"Studies in Neural Data Science",* Springer Proceedings in Mathematics & Statistics

### National conference proceedings (with peer-review)

- **Paganin, S.** (2021). Semiparametric IRT models for non-normal latent traits (pp. 178 - 181) *CLADAG 2021 Book of Abstracts and Short Papers*
- **Paganin, S.**, Paciorek C., de Valpine P. (2020). Bayesian IRT models in NIMBLE. (pp. 644 - 649) *Book of Short Papers SIS 2020*
- **Paganin, S.** (2019). Domain knowledge based priors for clustering. *Proceedings of the Conference of the Italian Statistical Society. "Smart Statistics for Smart Applications",* Pearson
- **Paganin, S.** (2017). Modeling of complex network data for targeted marketing. *Proceedings of the Conference of the Italian Statistical Society. "Statistics and data Sciences: new challenges, new generations".* Firenze University Press

### Working papers (available upon request)

- **Paganin, S.**, Russo, M., Scarpa, B. (2022+). A generalized partial credit model for network dependent latent traits with an application on modeling students' ability.
- **Paganin, S.**, de Valpine P. (2022+). Computational methods for fast Bayesian model assessment via calibrated posterior p-values.

---

SOFTWARE

- COMPAREMCMCs, *R package, version 0.5.0.* An R package for running, managing, and comparing results from different MCMC packages.
- NIMBLE, *R package, version 0.12.2.* Contributed with reversible jump MCMC.

---

TEACHING & MENTORING EXPERIENCE

### UC Berkeley

- Mentor for the Undergraduate research mentorships in statistics (Spring 2020).
  Mentee: Tianchi Liu. Project: *Building models and case studies using NIMBLE.*

### Università degli Studi di Padova

- Thesis co-supervision: Claudia Stocchi (2021), master student in Statistical Sciences.
- *Statistical Methods for Insurance Marketing.* Invited lecture for the course of "Business Statistics" taught by Prof. Mariangela Guidolin. (April 24, 2018 & online on November 24, 2020).
- Mentor at *Scegli con noi.* Orientation events for high schools students.
  (2014, 2015, 2016 editions) Campus Agripolis di Legnaro, Padova, Italy.
- *Advanced statistical inference* (M.Sc.), Year 2013/2014, Academic Tutor, Department of Statistical Sciences.
- *Introduction to real analysis* (B.Sc.), Year 2013/2014, Academic Tutor, Department of Statistical Sciences.

### Short-courses

- *Programming With Hierarchical Statistical Models Using NIMBLE* (May 23, 2022). Full-day, hybrid short course during the 35th New England Statistics Symposium hosted at UConn.

PRESENTATIONS    **Invited Seminars**

- Fast Bayesian model assessment via calibrated posterior predictive p-values
  *University of Nottingham* (Statistics and Probability seminar), April 7, 2022.

- Centered Partition Processes: Informative Priors for Clustering
  *Millenium Nucleus Center for the Discovery of Structures in Complex Data*
  *Pontificia Universidad Católica de Chile* (online) October 8, 2021.

- Centered Partition Processes: Informative Priors for Clustering
  *UC Davis, Department of Statistics* (Spring seminars series), May 6, 2021.

- Centered Partition Process: Informative Priors for Clustering.
  *Bayesian Analysis - Discussion paper Webinar*
  [https://www.youtube.com/watch?v=V0zCB8doqlo]
  March 29, 2021.

**Invited talks**

- Semiparametric IRT models for non-normal latent traits
  *Cladag 2021*, Online conference, September 9, 2021.

- Informative model-based clustering via Centered Partition Processes.
  *YoungStatS "One World webinar" - Developments in Bayesian Nonparametrics*
  virtual, April 21 2021.

- Modeling health surveys via semiparametric IRT models.
  *Bayesm:O.* Session on "Advances in Bayesian methods for medical data",
  virtual, November 17, 2020.

- Prior-Driven Cluster Allocation in Bayesian Mixture Models.
  *JSM 2020.* j-ISBA organized session, virtual conference, August 03, 2020

- Domain knowledge based priors for clustering.
  *SIS 2019.* ySIS organized session. Milano, Italy. June 19, 2019

- Modeling of complex network data for targeted marketing.
  *SIS 2017.* Firenze, Italy. June 29, 2017

**Contributed talks**

- A hierarchical Hidden Markov Model for cancer detection.
  *JSM 2022.* Washington D.C., August, 10 2022

- A hierarchical Hidden Markov Model for cancer detection.
  *ISBA 2022.* Montreal, Canada. July 1, 2022

- Statistics for early cancer detection.
  *PDA Research Symposium.* HPSH, Harvard University, online. February 22, 2022

- Flexible model assessment via approximate calibrated posterior predictive p-values.
  *ISBA 2021.* Online conference. June 29, 2021

- Centering Exchangeable Partition Models.
  *IBC 2018.* Barcelona, Spain. July 10, 2018

- Bayesian modeling of networks in complex business intelligence problems.
  *COBAL V.* Cimat, Guanajuato, Mexico. June 8, 2017

**Poster presentations**

- Flexible model assessment via approximate calibrated posterior predictive p-values .
  *ISBA 2021.* Online conference. June 29, 2021

- Computational methods for Bayesian semiparametric Item Response Theory models.
  *WiDS Cambridge, 2021.* Online conference. March 11, 2021

- Informative Bayesian Clustering for Mixture Models.
  *Advanced Statistics for Physics Discovery.* Padova, Italy. September 24, 2018

- Centering Exchangeable Partition Models.
  *ISBA 2018.* Edinburgh, United Kingdom. June 27, 2018

- Bayesian modeling of networks in complex business intelligence problems.
  *ISBA 2016.* Forte Village Resort Convention Center Sardinia, Italy. June 15, 2016

| | |
|---|---|
| **VISITING PERIODS** | Mar. 2019 – Jun. 2019. *Research visit* at Department of Environmental Science, Policy & Management, UC Berkeley. |
| | Jan. 2018 – Mar. 2018. *Visiting Research Scholar* at the Department of Statistics, Duke University (NC, USA) under the supervision of Prof. Amy H. Herring |
| | Oct. 2016 – Jun. 2017. *Visiting Research Scholar* at the Department of Biostatistics University of North Carolina at Chapel Hill (NC, USA) under the supervision of Prof. Amy H. Herring |

**EDITORIAL ACTIVITY**

**Associate Editor**

from 06/2021 The New England Journal of Statistics in Data Science (Software section)

**Reviewer**

(2022) Statistical Papers, Journal of Machine Learning Research.

(2021) BMJ, `R` Journal (2), Journal of Computational and Graphical Statistics, Harvard Data Science Review, Applied Psychological Measurements, Duke Dathaton.

(2020) BMJ, Statistical Papers, NSF proposal review, JOSS.

**SERVICE TO PROFESSION**

**Positions in scientific societies**

- j-ISBA board, Treasurer (2020-2022)

**Organization of scientific events**

- Scientific committee of BayesComp 2023 (Levi, Finland).
- Session organizer, (with j-ISBA) at JSM 2022 "Advances in scalable Bayesian methods for spatial data".
- Session organizer, (with j-ISBA) at ISBA 2022 "Advances in Bayesian methods for complex data".
- Organizing committee of BAYSM 2021 (virtual conference).
- Organizing committee of ISBA 2021 World Meeting (virtual conference).
- assistance to organization, IT set-up and conduction of the NIMBLE short course, June 3–5 2020, UC Berkeley (online workshop).

**Outreach**

- regular contributor to the ISBA Bulletin (2021)
- volunteer for *Venetonight – La notte dei ricercatori*, Padova, Italy. (2015 – 2017) developed an online app interfacing with Twitter to track and display the event sentiment in real-time.
- volunteer for *StatisitcAll*, Treviso, Italy. (2015) statistical games and activities to show the magic of statistics to kids and adults.

**Memberships**

- International Society for Bayesian Analysis (ISBA); j-ISBA; New England Statistical Society (NESS).

**WORK EXPERIENCE**

- *Statistical consultant* for Azienda ULSS n.4. Analysis of relapsing rate after being in rehab facilities.
- Jul. 2014 – Aug. 2015. *Marketing Analyst* at Generali Italia SPA (Mogliano Veneto) Data quality and gathering. Customer satisfaction analysis: definition of the sampling plan, questionnaires analysis, text analysis of online survey opinions. Reporting.

**PROGRAMMING & LANGUAGES**

**Programming**

- Programming languages: R (advanced) practice of Rcpp, plyr/dplyr, ggplot2, shiny libraries. C/C++, SQL (good). JAVA, Python, Julia (base).
- O.S.: Linux (Fedora/Ubuntu), OSx, Windows and relative software.
- Other: practice of LaTeX, Sweave, Markdown for reporting, Git and Github as revision control system, Gimp/Illustrator for image processing.

**Languages**

- Italian: native; English: fluent; French: basic; Spanish: basic.