The Adolph C. and Mary Sprague

MILLER INSTITUTE

for Basic Research in Science

Newsletter
Fall 2015

Uncovering and Representing Structure in Data

Miller Fellow Focus: Elaine Angelino

Uncovering and Representing Structure in Data

As scientists, we strive to understand our observations of the world around us. We test our understanding of a specific system by making predictions about how it will behave. Probabilistic models provide a quantitative framework for representing systems and encoding uncertainty about system behavior using probability distributions. Estimating structure in data via probabilistic modeling is a process known as inference, and a central tool of modern data analysis.

As a motivating example, consider the requirements of real-time personalized medicine. For many different cancer types, we would like to build probabilistic models that quickly recommend treatments tailored to individuals, based on their high-dimensional genotype data. Can the model inference be computationally efficient, and can the resulting model make statistically accurate diagnoses?

My research confronts these kinds of computational and statistical questions that arise when we think about how to leverage modern datasets. This work follows two lines of attack: developing algorithms for large-scale inference and designing probabilistic models to analyze new kinds of data.

Algorithms for Large-Scale Inference

Inference is computationally expensive when the underlying functions are high dimensional and/or full of many local optima, as is typical with large datasets. In general, there are no analytic solutions to these problems; there are approximate and simulated approaches, but these are often slow and do not naturally leverage modern computing resources, such as compute clouds. In my thesis work and during the past year as a Miller Fellow, I have focused on developing new algorithms for large-scale Bayesian inference that use parallel computers.

Inference is dominated by two approaches. The first uses optimization procedures to find a single best-fit parameter setting based on observed data (Figure 1A). However, this model may not lead to the best predictions (Figure 1B). This motivates the Bayesian approach of integrating over all plausible parameter settings. Bayesian inference is a probabilistic framework for describing uncertainty via distributions, called posterior distributions, over model parameter values that explain observed data.

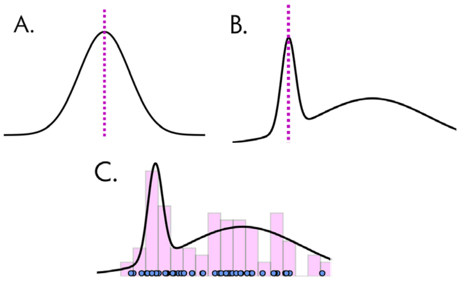


Figure 1
A. A simple posterior distribution where the best-fit parameter value (horizontal location of dashed line) is an informative summary of all reasonably likely values.
B. A more complicated posterior distribution. The bulk of its mass lies in the broad region to the right of the narrow peak around the best-fit parameter value. The true parameter value is more likely to be under the broad region than around the peak.
C. Posterior distributions are difficult to work with, but we can approximate them using samples (circles).

CONTINUED ON PAGE 6 >

Inside this edition:

Miller Fellow Focus 1, 6, 8
Call for Applications/Nominations 2-3
Next Steps & Announcements 4
19th Annual Symposium 5
60th Anniversary Announcement 7

Call for Nominations

Miller Research Fellowship Nominations

Deadline: Thursday, September 10, 2015

Miller Research Professorship Applications

Deadline: Thursday, September 17, 2015

Visiting Miller Professorship Nominations

Deadline: Friday, September 18, 2015

For more information on all our programs:
miller.berkeley.edu

In the News...

For updated news:
miller.berkeley.edu

"I was a Miller Fellow in 1976-1978, and my career benefited immensely from that brief period of freedom and ideas—unexpected ideas—from all sorts of directions"

- Adrian Bejan
J.A. Jones Distinguished Professor of Mechanical Engineering, Duke University. Honorary member of the American Society of Mechanical Engineers. Honarary Member of the Romanian Academy.

Call for Miller Professor Applications

Online Application Deadline: Thursday, September 17, 2015

Miller Research Professorship AY 2016-2017

The Miller Professorship program is looking with a view to the future in announcing the call for applications for terms in 2016-2017. The goal is to accommodate a greater range of campus faculty to participate in the vibrant Miller community. The objective of the Miller Professorship program is to provide opportunities for faculty to pursue new research directions on the Berkeley campus. For some, this may best be enabled by taking time off from teaching. This will continue to be an option. For others, the teaching obligations are critical to maintaining campus academic programs. There will thus be a new option for Miller Professors, allowing the continuation of campus service and teaching. Funds will be distributed differently depending upon the choice selected. Details of the terms and the application procedure are posted on the Miller Institute website. The primary purpose of the Miller Professorship program and the evaluation criteria will continue to be research excellence. Applicants will also be encouraged to describe their interest in participating in the Miller Institute community.

Applications from UC Berkeley faculty for Miller Research Professorship terms in the 2016-17 academic year will be accepted online beginning in May 2015. The purpose of the Professorship is to release members of the faculty from teaching and administrative duties and allow them to pursue research on the Berkeley campus. Appointees are encouraged to follow promising leads that may develop in the course of their research.

Applications are judged competitively and are due by Thursday, September 17, 2015. It is anticipated that between five to eight awards will be made.

Call for Visiting Miller Professor Nominations

Online Nomination Deadline: Friday, September 18, 2015

Visiting Miller Research Professorship AY 2016-2017

The Advisory Board of the Miller Institute for Basic Research in Science invites Berkeley faculty to submit online nominations for Visiting Miller Research Professorships and the Gabor A. and Judith K. Somorjai Visiting Miller Professorship Award for terms in Fall 2016 or Spring 2017. The purpose of these Visiting Miller Professorships is to bring promising or eminent scientists to the Berkeley campus on a short-term basis for collaborative research interactions. It is required that awardees are in residence at Berkeley during their appointment term. Faculty members or research scientists from any place in the world are eligible to be considered for sponsorship. Non-US citizens must be eligible for J-1 Scholar visa status. Faculty members at other UC campuses are eligible to be nominated for this program. The Miller Institute, as the sponsor and administrative department, will extend an invitation after advising the department of its selection.

Any questions regarding this program may be directed to the Institute's office by phone at (510) 642-4088 or by emailing the Miller Institute.

:: **Nomination & Application details**: miller.berkeley.edu

:: **Questions?** Kathryn Day: 510-642-4088 | millerinstitute@berkeley.edu

The Adolph C. and Mary Sprague

Miller Institute for Basic Research in Science

Call for Nominations: Miller Research Fellowship

2016-2019 Term

Nomination Deadline: Thursday, September 10, 2015

The Miller Institute for Basic Research in Science invites department chairs, faculty advisors, professors and research scientists at institutions around the world to submit nominations for Miller Research Fellowships in the basic sciences. The Miller Institute seeks to discover and encourage individuals of outstanding talent, and to provide them with the opportunity to pursue their research on the Berkeley campus. Fellows are selected on the basis of their academic achievement and the promise of their scientific research. Miller Fellows also have a keen curiosity about all science and share an appreciation for an interdisciplinary experience. The Miller Institute is the administrative home department for each Miller Fellow who is hosted by an academic department on the Berkeley campus. All research is performed in the facilities provided by the UC Berkeley academic department. A list of current and former Miller Research Fellows can be found at: http://miller.berkeley.edu/fellowship/members/all-mf-by-name

Miller Research Fellowships are intended for exceptional young scientists of great promise who have recently been awarded, or who are about to be awarded, the doctoral degree. Normally, Miller Fellows are expected to begin their Fellowship shortly after being awarded their Ph.D. A short period as a post-doctoral fellow elsewhere does not exclude eligibility. However, applicants who have already completed substantial postdoctoral training are unlikely to be successful except in unusual circumstances. **A nominee cannot hold a paid or unpaid position on the Berkeley campus at the time of nomination or throughout the competition and award cycle**. Nominees who are non-US citizens must show eligibility for obtaining J-1 Scholar visa status for the duration of the Miller Fellowship. Non-US citizens will be required to prove English language proficiency prior to award. The Miller Institute does not support H1B visa status. The Fellowship term must commence between July 1 and September 1, 2016. Eligible nominees will be invited by the Institute to apply for the Fellowship. Direct applications and self-nominations are not accepted.

*All nominations must be submitted using the Online Nomination System at: http://**miller.berkeley.edu**

Nominators will need the following required information to complete the online nomination process:

- Nominee's complete full and legal name (do not use nicknames)
- Nominee's current institution
- Nominee's complete, current, and active E-mail address, current mailing address with postal code and telephone number
- Nominee's Ph.D. Institution and (expected) Date of Ph.D. (month & year required)
- Letter of recommendation and judgment of nominee's promise by the nominator (saved in PDF format). Letter must be specific to the Miller Fellowship, have a current date, and be on institutional letterhead. The Executive Committee finds it helpful in the recommendation letter to have the candidate compared with others at a similar stage in their development.
- Nominator's current active E-mail address, title, and professional mailing address (include zip code/campus mail code)

The Institute provides a stipend of $65,000 with annual 2% increases and an annual research fund of $10,000, for total initial compensation of $75,000. There is provision for travel to Berkeley for Miller Fellows and their immediate families and a maximum allowance of $3,000 for moving personal belongings. Benefits, including medical, dental, vision and life insurance are provided with a modest contribution from the Miller Fellow. All University of California postdocs are represented by the UAW. Fellowships are awarded for three years, generally beginning August 1, 2016 and ending July 31, 2019. Approximately eight to ten Fellowships are awarded each year. Candidates will be notified of the results of the competition starting in mid-December, and a general announcement of the awards will be made in the spring.

We are grateful for your thoughtful participation in this process and hope that you regard the time you may devote to this effort justified by the contribution you will be making to the careers of distinguished young scientists.

MILLER INSTITUTE FOR BASIC RESEARCH IN SCIENCE
2536 Channing Way, Berkeley, CA 94720-5190
ph: 510-642-4088 | fax 510-643-7393
http://miller..berkeley.edu

From the Executive Director

This coming year promises to be an especially exciting one for current and former members of the Miller Institute – we are celebrating our 60th anniversary! We do so with an array of events, from hosted department colloquia throughout the year to a weekend symposium combined with excursions and social events. The 60th anniversary symposium will feature the exciting cross-disciplinary selection of world leaders that you have come to expect from the June retreats. We do hope you can attend, but please note that registration is required http://www.miller.berkeley.edu/60years and some events have a strict limit on the number of participants.

This year the Miller Institute updated one of its core programs, the Miller Professorship. Miller Professors may still apply to pursue new ideas and be relieved of teaching responsibilities. However, it will also be possible to request support for pursuing this research while maintaining normal academic responsibilities. In either case, the novelty and importance of the proposed research will continue to be the primary consideration and all Miller Professors will get to participate in the weekly Institute activities.

Summer is always a season of transition at the Miller Institute. New members arrive, and Miller Fellows move on. The Executive Committee is also changing with Rich Saykally and Kathy Collins wrapping up their appointments. Their sense of humor and passion will be very much missed. Steve Leone joins the committee and Jasper Rine returns after spending the past year organizing the 60th anniversary events. After 6 years of sharing his wisdom and creative ideas, Harry Kroto finished his second term on the Advisory Board. Joining the Advisory Board is Steve Block from Stanford who, as an attendee at several symposia, is no stranger to the Miller Institute.

Even if you are unable to attend the 60th anniversary events, please let us know if and when you visit campus as it may be possible to join us at lunch or other social events.

Sincerely, Michael Manga
Executive Director, Miller Institute &
Professor of Earth and Planetary Science

Birth Announcements

Chris Lowe (Miller Fellow 1998-2001) and husband Chad Harris welcomed daughter Isabella Harris-Lowe, born May 3, 2015.

Andy Suarez (Miller Fellow 2001-03) and Elissa Suhr welcomed son Taysom Suhr Suarez on May 14, 2015.

Eva Schmid (Miller Fellow 2008-11) and Oliver Hoeller welcomed son, Jakob Hoeller on May 7, 2015.

Rebecca Schulman (Miller Fellow 2008-11) and Marc Kamionkowski (Visiting Miller Professor 2010) are pleased to announce the birth of their daughter, Ada Vera Kamionkowski, born 16 Feb 2015.

Phil Hopkins (Miller Fellow 2008-11) and Jessie Christiansen welcomed twins, Clara and Hugo on January 29th, 2015.

Alice Shapley (Miller Fellow 2003-05) and Edwin Schauble welcomed son, Jacob John Schauble on August 21, 2014.

Angelos Michaelides (Somorjai Visiting Miller Professor 2013) and his wife Michelle welcomed daughter, Josephine Anna Michaelides born on 16th of August 2014.

Obituaries

Ernest S. Kuh, Miller Professor 1965-66, passed away on June 29, 2015.

Next Steps

The Miller Institute congratulates the following Miller Fellows on their next endeavors.

Thomas Bodin is a CNRS Researcher at the Laboratoire de Geologie de Lyon, France.

Qian Chen is an Assistant Professor at the University of Illinois Urbana-Champaign.

Xie Chen is an Assistant Professor at Caltech.

Francesco d'Eramo is moving to UC Santa Cruz in a postdoctoral position.

Timofey Frolov is assuming a research position at the Lawrence Livermore National Lab.

Justin Kim is moving to Stanford University to continue his work with Professor Carolyn Bertozzi.

Chen Li will move to Johns Hopkins University as an Assistant Professor.

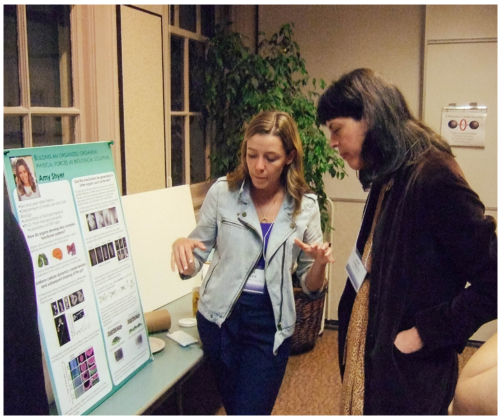
Milo Lin is moving to Dallas as an Assistant Professor at the University of Texas, Southwestern Medical Center.

Christopher Martin is now an Assistant Professor at the University of North Carolina.

Steven Sam is an Assistant Professor at the University of Wisconsin, Madison.

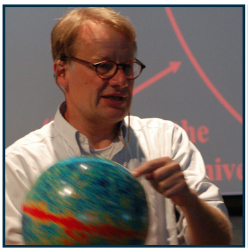
Sa Kan Yoo is an Associate Chief Researcher at Riken Research, Japan.

19th Annual Interdisciplinary Symposium


Miller Fellow Amy Shyer and Zoe Corbyn


Ruth Murray Clay


Maitreya Dunham


Lyman Page


Adam Cohen


James Sethian


Donna Blackmond


Cynthia Kenyon


Miller Fellow Carson Bruns explains his work


Miller Fellow Chris Martin


Miller Senior Fellow Barbara Meyer with Miller Fellows Rachel Pepper, Amy Shyer and Brooke Gardner


Miller Fellow Chen Li


Jasper Rine, Yun Song and Alberto Grunbaum


Miller Fellow Kestutis Cesnavicius


Miller Fellow Itay Budin with Howard Stone


Mikhail Shapiro with Miller Fellow Da Yang

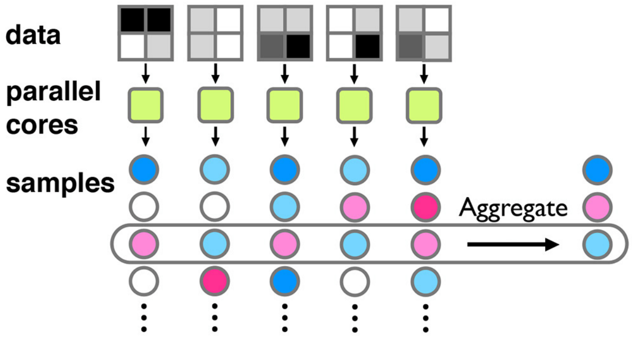
> CONTINUED FROM PAGE 1 [Miller Fellow Focus]

Many posterior distributions are difficult to work with, both mathematically and computationally. In practice, we either use optimization techniques to focus on maxima, or we generate samples, i.e., representative examples of parameter values, given observed data (Figure 1C). From these samples, we can form a histogram that approximately summarizes the entire posterior distribution. Markov chain Monte Carlo (MCMC) algorithms are general procedures for both optimization and sampling problems. Traditional MCMC algorithms are serial and thus slow, and cannot exploit parallel computers. In particular, each iteration of the algorithm can be slow when it depends on a large dataset, and the algorithm can require prohibitively many iterations to converge.

Maxim Rabinovich, Professor Michael I. Jordan, and I recently introduced variational consensus Monte Carlo (VCMC), a novel class of data-parallel MCMC algorithms [1]. VCMC is a generalization of the consensus Monte Carlo (CMC) algorithm [2], which was designed to fit into the widely used MapReduce framework originally developed at Google. CMC divides the data into partitions and uses parallel cores to generate samples from each partition separately (Figure 2). These samples are only reflective of their respective data partitions, but they can be aggregated to form approximate samples reflective of the entire dataset. CMC aggregates via a weighted average, where the weights are derived for the special case of Gaussian posteriors. VCMC generalizes this approach to non-Gaussian posteriors by finding the best weights through optimization, leading to significantly smaller errors in computations involving the approximate samples.

Figure 2

Consensus Monte Carlo enables efficient, data-parallel MCMC.



Probabilistic Graphical Models

Recently, I have been studying structure in data through the lens of probabilistic graphical models. For example, I am working with cell biologists who are exploiting advances in high-throughput assays to collect high-dimensional imaging data of cells during cell division. What determines whether the cell divides evenly? What sub-cellular structures influence the sizes and shapes of other structures? Graphical models provide a framework for understanding such systems of interacting parts.

Probabilistic graphical models are an expressive class of probabilistic models with powerful predictive capabilities. These models use graphs (networks) to succinctly express relationships between system components. Nodes correspond to variables and edges between nodes represent interactions. For instance, in an Ising model from statistical physics, each node is a binary random variable and corresponds to a spin; each spin directly interacts only with its (physical) neighbors. Importantly, variables can influence each other via indirect interactions.

Modeling these indirect interactions can have critical predictive consequences. For example, with an interdisciplinary team of scientists, I am studying a system of interacting genes and proteins that regulate fruit fly embryo development. We want to predict quantities like P, the expression level of a protein. Suppose P depends directly on A and B, the concentrations of two transcription factor proteins bound to the regulatory region of the gene encoding the protein of interest (Figure 3A). We can model P directly as a function of A and B, e.g., via linear regression. However, the entire system contains many other genes and proteins that can indirectly influence P, e.g., suppose another protein Y interacts with A, which interacts with P. A graphical model represents all these interactions (Figure 3B) and allows us to make complex predictions, e.g., regarding what happens to P if we activate the expression of protein X while knocking down the expression of Y.

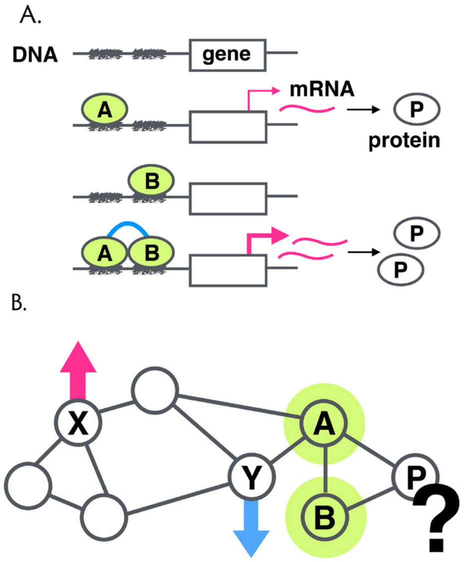


Figure 3

A. Transcription factor proteins **A** and **B** (green ovals) regulate the expression of a gene encoding protein **P**. Upstream of the gene, the transcription factors can bind to the DNA, as well as to each other (blue link), leading to different levels of expression.

B. A graphical model encodes the structure of how proteins and genes (nodes) in a system interact (edges). A linear regression model for gene expression level **P** depends only on the direct interactions of transcription factors **A** and **B** (green circles), which can also directly interact with each other. In contrast, a graphical model can represent indirect interactions (e.g., protein **Y** interacts with **A**, which interacts with **P**) and allows us to make complex predictions (e.g., what happens to **P** when we activate the expression of **X** while knocking down that of **Y**).

CONTINUED ON PAGE 8 >

The Adolph C. and Mary Sprague

MILLER INSTITUTE

for Basic Research in Science

Register Online

miller.berkeley.edu

Roger Bilham
Visiting Miller Professor Fall 2004

Kenneth Eisenthal
Visiting Miller Professor Fall 1989

Alice Guionnet
Visiting Miller Professor Fall 2006

L. "Maha" Mahadevan
Visiting Miller Professor Spring 2006

Geoff Marcy
Miller Professor 2011 - 2012
2015 Miller Senior Fellow

Ryosuke Motani
Miller Fellow 1997 - 1999

Sarah Otto
Miller Fellow 1992 - 1994

Vijay Pande
Miller Fellow 1996 - 1998

More Info: miller.berkeley.edu

Join us to *Celebrate* *60 years* of Science

Friday, January 15 - Sunday, January 17, 2016

Friday evening: Reception - Alumni House

Saturday - Stanley Hall

8:30 – 8:45	Registration
8:45 – 9:00	Welcome
9:00 – 11:45	Talks
12:00 – 1:00	Lunch
1:00 – 3:15	Talks
3:15 – 3:45	Break
3:45 – 6:00	Talks
6:30	Hors D'oeuvres Dinner

Sunday - Optional Tours

8:00 – 5:00	Geology Tour of Marin County
10:00	Advanced Light Source Astronomy Department – Telescope UC Botanical Garden
11:00	UC Botanical Garden UC Museum of Vertebrate Zoology
12:00	UC Museum of Vertebrate Zoology

Your tax deductible gift
will continue to support
basic science research:

miller.berkeley.edu/Gift

With generous support from:
The College of Chemistry
The College of Engineering
The College of Natural Resources
The Dean of Biological Sciences
and The Vice Chancellor for Research

Non-Profit
Organization
U.S. Postage
PAID
University of
California,
Berkeley

University of California, Berkeley
Miller Institute for Basic Research in Science
2536 Channing Way, #5190
Berkeley, CA 94720-5190
510.642.4088
miller.berkeley.edu

Miller Institute News - Fall 2015
Please send address corrections to:
miller_adm@berkeley.edu

1-36299-24810-44

Miller Fellow Focus

> CONTINUED FROM PAGE 6

Before we can leverage graphical modeling tools in our system, we need to specify a concrete model. In many systems of interest, we do not know the graph structure a priori, and we might not know all the relevant variables. For instance, if we are studying a network of genes and proteins, we do not necessarily know all their binding interactions, and there might be unknown regulatory mechanisms at play. Still, we can measure those variables of which we are aware. In certain cases, we can recover the graph structure from these observations alone. In general, however, we will be uncertain about the true underlying graph structure. Bayesian approaches provide a principled framework for working with this uncertainty, but we need efficient and scalable inference algorithms to make these approaches feasible. Fitting graphical models to data presents many challenges that I hope to understand by studying real systems.

There is always a gap between the observations we can make and the underlying systems we want to probe, and inference is the central tool for bridging this gap. Solving inference problems at scale and for new and complex systems will require innovation at the interface of computer science and statistics research.

Elaine Angelino (Miller Fellow 2014-2017) completed her A.B. in applied mathematics, M.A. in systems biology, and Ph.D. in computer science at Harvard. During grad school, she enjoyed working with Margo Seltzer and Eddie Kohler (computer systems) and Ryan P. Adams (machine learning). She is indebted to her undergrad advisor, Michael P. Brenner, for telling her about the Miller Institute. Now, she is bringing together the various components of her academic training, and is thankful that her faculty host, Michael I. Jordan, has encouraged her to beef up on statistics. Outside of work, Elaine is becoming more Californian by the day, e.g., she owns climbing shoes and actually uses them.

Contact: elaine@berkeley.edu

---

[1] Maxim Rabinovich, Elaine Angelino, Michael I. Jordan. Variational consensus Monte Carlo. Preprint on arXiv:1506.03074 [stat.ML]. 2015.
[2] Steven L. Scott, et al. Bayes and big data: The consensus Monte Carlo algorithm. Bayes 250. 2013.

New Website

Please visit our new website at miller.berkeley.edu

Online Newsletter

The Miller Institute invites you to enjoy our previous e-newsletters by visiting: **miller.berkeley.edu**