Lifetime Data Analysis Interest Group

LIDA Newsletter

Volume 3, Number 1 – January 2018

ISSN 2473-5159

An interest group associated with the American Statistical Association

In Brief

* * * * * * * * * * * * *

Our website is:
`http://community.amstat.org/lifetimedataanalysislidainterestgroup/home`.
Published newsletters are archived under "Library".

* * * * * * * * * * * * *

Election 2017 Result
Dr. Jianwen Cai became Chair-elect on January 1, 2018.

LIDA Chater and Section Status

A Note from Industry: Estimands
Jonathan Siegel discusses on the regulatory addendum to the general guidance on "Statistical principless in clinical trials" (E9).

2019 Conference on Lifetime Data Science to be held in Pittsburgh
Local organization will be chaired by Drs. Ying Ding and Yu Cheng.

New Column: Book Review

New Column: Software Review

* * * * * * * * * * * * *

Membership in the LIDA-IG
If you have an interest in life history data you are invited to join.
Membership forms can be found on our website.

LIDA Officer

Chair:	Richard Cook
Chair-Elect:	Jianwen Cai
Past Chair:	Mei-Cheng Wang
Secretary 2015-2018:	Jonathan Siegel
Treasurer 2016-2019:	Chiung-Yu Huang
Program Chair (2018):	Yu Shen
Webmaster:	Weiliang Qiu
Newsletter Editor:	Jun Yan

Chair's Message

In your lifetime, if you are lucky, you will have a chance to work with people who are exceptional professionals, as well as delightful, dedicated and supportive colleagues. This is the fortunate position I find myself in as Chair of the LIDA IG executive committee for 2018. Much progress has been made to advance our collective agenda since the inception of our interest group by Mei-Ling Ting Lee. Under the leadership of Jack Kalbfleisch and Mei-Cheng Wang our membership has grown to 222 ASA members and 70 other members, a highly successful conference was held last year, and we have been active in the sponsorship of sessions at the Joint Statistical Meetings and elsewhere. These are exciting times!

As he steps off the executive committee I'd like to express sincere thanks to Jack for his influential leadership over the past few years. The progress has been amazing in a short space of time and much credit goes to Jack. The sentiments expressed by Mei-Cheng later in the newsletter are heartfelt and shared by all and we hope to benefit from your involvement in LIDA-IG activities in the future.

Mei-Cheng has been a remarkable force for the good of the LIDA-IG over the past year — with her considerable enthusiasm, experience and energy, and I'm very happy to work with her in the coming year. We congratulate Jianwen Cai on her election as Chair-Elect and welcome her to the committee. There is a lot to do in the coming year and I'm pleased to work with the members of the executive on our important initiatives.

Hearty thanks also go to Bin Nan for completing his term as Program Chair. He has ensured the profile of our group was high through sponsorship of sessions at the Joint Statistical Meetings and elsewhere. Yu Shen has taken over in 2018 and reports on the contribution to the JSM which is sponsored by our group. Another exciting LIDA-IG sponsored event this year is the Fourth International Workshop on Statistical Analysis of Multi-Outcome Data being held June 11–12, 2018. Lei Liu, who serves as Program Chair, provides more details in his entry later in this newsletter.

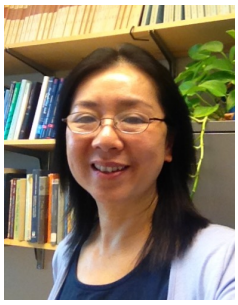
As noted in a later entry we find ourselves in the fortunate position of being ready to apply for approval as an official section of the American Statistial Association this year. A report is being drafted for submission in the coming months and we anticipate a decision will be made at the time of the Joint Statistical Meetings.

The enthusiasm felt for the 2017 Lifetime Data Science conference led to suggestions that this become a recurrent event. To this end plans are underway for a 2019 Lifetime Data Science conference. This will be hosted by the University of Pittsburgh and held May 29–June 1, 2019. Thanks to Ying Ding and Yu Cheng who are serving as co-Chairs of the Local Organizing Committee. Like the 2017 conference, the program will feature workshop, keynote speakers and invited sessions. The Scientific Program Committee will be formed during the next two months and more detail will be forthcoming in the near future. For now please mark these dates on your calendars — we hope to see you there!

Wishing you all the best for 2018!

Richard Cook, Chair 2018

Message from the Past Chair



As the Past-Chair of LIDA-IG, I feel very fortunate to work with a group of dedicated officers, all working hard toward achieving the goal to eventually become an ASA Section. In particular, I greatly appreciate the outstanding efforts and service of Jonathan Siegel as Executive Secretary, Bin Nan as Program Chair, Chiung-Yu Huang as Treasurer, Jun Yan as Editor of the Newsletter, and Weiliang Qiu as Web Master.

In 2017 the LIDA-IG had its first conference held at the University of Connecticut, Storrs, in May. The conference received very enthusiastic support from our profession and was well attended with 340 registered participants. Thanks to UConn organizing committee (chaired by Ming-Hui Chen and Jun Yan), the conference even had a budget surplus which will be used for the 2019 LIDA conference and other activities in the future. The LIDA-IG has sponsored three invited sessions and one topic-contributed session at the 2017 Joint Statistical Meetings (JSM) in Baltimore, and served as a co-sponsor for the 2017 Chinese Statistical Association (ICSA) China Conference in Jilin, China.

In 2017 the LIDA-IG had its second election for the chair-elect. The election was managed by the American Statistical Association and the election committee includes John D. Kalbfleisch (Past Chair), Douglas Schaubel (Past Program Chair) and Rebecca Betensky. We were very fortunate to have three strong and distinguished candidates: Jianwen Cai, Malka Gorfine and Ian MkKeague. There were 108 votes cast and Professor Jianwen Cai from the University of North Carolina has been elected as Chair-Elect for the Interest Group. She will be Chair-Elect in calendar 2018 and the Chair in calendar 2019.

As the Past-Chair, I will be responsible to organize the elections in the fall of 2018. A nominating committee of three members will identify candidates for election of Chair-Elect, Secretary and Program Chair for 2019. If you have ideal candidates in mind or like to be nominated for any of these positions please let me know and I will take the names forward to the nominating committee for discussion.

Wish all the LIDA-IG members a very happy and healthy New Year in 2018!

Mei-Cheng Wang, Chair 2017

Thanks and Appreciation to John D. (Jack) Kalbfleisch

Our past chair of 2017, Dr. John D. (Jack) Kalbfleisch is Emeritus Professor of Biostatistics and Statistics at the University of Michigan. We have been very fortunate to have him to lead the LIDA-IG in 2016. As the second chair of LIDA-IG, Jack has been involved in almost all the developments and activities, and has provided valuable and important guidance to the group. Jack received his Ph.D. in statistics from the University of Waterloo and began his career as assistant professor of statistics at the State University of New York at Buffalo. He was on faculty at the University of Waterloo and served as Chair of the Department of Statistics and Actuarial Science and as Dean of the Faculty of Mathematics. In 2002, he moved to the University of Michigan and served as Chair of the Department of Biostatistics at Michigan and later as Director of the Kidney Epidemiology and Cost Center from until his retirement in 2012. More recently, he has continued to work on various research projects and involve in professional activities (such as LIDA-IG). We thank Jack, deeply, for his many contributions and look forward to his continued involvement in LIDA-IG related activities in the future.

Mei-Cheng Wang, Chair 2017

Election 2017 Results

The Nominations Committee Consisted of Drs. Rebecca Betensky (Harvard), Douglas Schaubel (Michigan) and myself. As Past-Chair of the LIDA Interest Group, I served as Chair of the Nominations Committee. We met over the last part of the summer, 2017.

We had an outstanding slate of candidates comprising Dr. Jianwen Cai of the University of North Carolina, Dr. Malka Golfine of Tel-Aviv University, and Dr. Ian McKeaugue of Columbia University. The election was managed by the American Statistical Association and voting took place between October 4 and 27, 2017. Dr. Jianwen Cai was elected Chair-Elect effective January 1, 2018, and will serve as Chair in 2019.

Dr. Cai is Boshamer Distinguished Professor and Vice Chair of Department of Biostatistics at the University of North Carolina at Chapel Hill. Her research interests include development and application of statistical methods for survival outcomes, especially multivariate survival outcomes and recurrent events, design issues in clinical trials and cohort studies. She has had multiple statistical methods grants to support the methods development. She also collaborates extensively with other scientists working in various areas in health research. She is currently the Principal Investigator for the Coordinating Center for the Hispanic Community Health Study/Study of Latinos (HCHS/SOL), a multi-center longitudinal study of over 16,000 Hispanics/Latinos. She has served as Chair of the ASA Biometrics Section and on the Board of the International Chinese Statistical Association (ICSA) as well as Chair for COPSS FN David Award Committee and ASA Wilks Memorial Medal Committee. She served as President Elect/President/Past President of ENAR during 2015–2017. She has extensive editorial experience and is currently Associate Editor for Statistics in Biosciences, Lifetime Data Analysis, and Journal of the American Statistical Association – Application and Case Studies. She has served on many other statistical society committees and on NIH and NSF review panels. She was elected as Fellow of ASA in 2005 and Fellow of IMS in 2009.

The Nominations Committee would like to thank all three candidates for allowing their names to stand. A strong slate of candidates like this speaks very well for the health and future prospects of the LIDA Interest Group. We also extend our congratulations to Dr. Cai and look forward to working with her over her terms as Chair-Elect, Chair and Past-Chair.

Jack Kalbfleisch, Nomination Committee Chair

Figure 1: Participants of the annual interest group meeting at JSM 2017 in Baltimore, Maryland.

Report from the Annual General Meeting at JSM 2017

As an interest group within the American Statistical Association, LIDA holds its annual business meeting during the Joint Statistical Meetings conference. Our meetings are listed in the JSM program. This year we held it at JSM in Baltimore on Monday, July 31.

We had about 40 attendees, about the same as last year.

Our Chair, Mei-Cheng Wang, opened the meeting. Our Past Chair, Jack Kalbfleisch, provided introductory remarks.

Jonathan Siegel, our secretary, provided a bit of LIDA history and explained our charter and the process of getting it approved within ASA. He also provided a membership report. We reported 201 verified ASA members and 62 other members, up from 185 verified ASA members and 47 other members reported at the 2016 annual meeting.

Jonathan discussed the Fall elections. Elections are run through ASA. Last year, Richard Cook was elected the 2017 chair-elect who became chair in 2018, and the Secretary and Treasurer (Jonathan Siegel and Chiung-Yu Huang) ran unopposed. This year, only the 2018 chair-elect was open for election. Jack Kalbfleisch, the Past Chair, serves as chair of the Nominating Committee.

William Notz, our Council of Sections liaison, discussed the Council of Sections, explained the process of becoming a section, provided insight into the Council of Sections perspective, and answered the membership's questions.

Mei-Cheng discussed our successful conference on Lifetime Data Science at the University of Connecticut in Storrs, CT, May 25–27 2017. The conference exceeded expectations, with approximately 340 attendees, and exciting sessions and discussions. It resulted in a budget surplus.

Bin Nan spoke about the opportunities available for providing sessions and presenting papers at JSM through LIDA. LIDA receives one topic contributed session and can apply for one of a pool of invited sessions that are awarded competitively.

Richard Cook discussed planning for the next conference on lifetime data science. The next conference will be in 2019.

Jin Yan discussed our newsletter and website.

Chiung-Yu Huang gave a brief treasurer's report. Interest Groups do not normally have treasurers or handle their own funds, but we created the position to support our activities. The University of Connecticut handled the finances for this conference. The conference resulted in a budget surplus, still being accounted for as of the annual meeting. In addition, ASA recently offered grants of $1000 per year for use for eligible expenses.

Our meeting was adjourned, with an opportunity to socialize (Figure 1).

Jonathan Siegel, Secretary

LIDA Charter and Section Status

The Lifetime Data Analysis Interest Group (LIDA-IG) began as a small informal group in 2013. It was formally approved by the American Statistical Association Committee on Sections in 2015. As an interest group, we get certain privileges within ASA including sponsoring our own topic-contributed session at JSM, having a non-voting representative at Council of Sections meetings, and receiving ASA assistance and facilities for our annual meeting at JSM, elections, and other activities. An interest group is open to anyone.

The founding Chair of the Group was Dr. Mei-Ling Ting Lee. Dr. Ross Prentice was also closely involved in organizing the group. Each year we elect a new Chair-elect who becomes the Chair in the following year. Mei-Ling Ting Lee served as Chair in 2015, John (Jack) Kalbfleisch, Mei-Cheng Wang and Richard Cook succeeded the Chair position in 2016, 2017 and 2018. In the Fall of 2017 we elected Jianwen Cai as 2018 Chair-Elect. Dr. Cai will become the 2019 Chair. Other members of the executive committee are Jonathan Siegel (Secretary), Chiung-Yu Huang (Treasurer), and Yu Shen (Program Chair). In addition, Jun Yan serves as the editor for the Newsletter and Weiliang Qiu is the web master.

Our charter required some time to develop. Our initial version was presented at our annual meeting at JSM in 2016, but on review the Council of Sections requested some changes. After we went through a couple of rounds of negotiations and refinements with COS, they accepted our May 15 2017 version, which is currently in effect. Our charter provides for elections for chair-elect, secretary, and treasurer, with other officers filled by rotation or appointment by the chair. Our program chair is responsible for conferences and other educational programs. Elections are held in the fall of each year between September and November through ASA's ballot system. All officers have 1-year terms except Secretary and Treasurer, which have 3-year terms.

We are seeking to be approved as a section within ASA. A section has additional abilities including collecting annual dues assessed along with ASA membership fees, using ASA bank accounts and accountants to handle money, sponsoring invited sessions at JSM, and other features. ASA requirements for becoming a section, approved at the 2015 annual COS meeting at JSM, include having at least 200 ASA members and functioning as an Interest Group for at least 3 years. We will formally apply for section status in 2018.

The LIDA-IG currently has 222 members who are also members of the American Statistical Association. In addition, we have 70 members who are not members of the ASA. Many of these are members of sister organizations such as the Statistical Society of Canada or the International Chinese Statistical Association. To further expand membership, we plan to advertise through our Newsletter and other outreach with the membership to encourage those members to join the ASA as well.

Becoming a section, if the ASA COS accepts our application, will have some consequences for our members. Under the ASA rules, members of a section have to be members of the ASA. This means that if our application for section status is accepted, non-ASA members will no longer be able to be full LIDA members. We value all our members and want to find a way to keep everyone involved. To this end, we are planning an associate status to continue to enable all our current members to receive our newsletter and participate in conferences, educational programs, etc. However, under ASA rules, only ASA members will be able to participate in section elections under the ASA ballot system. Sections are also able to charge membership dues, which are handled by the ASA membership process.

If you'd like more information about our charter, please email our Secretary, Jonathan Siegel, at `jonathan.siegel@bayer.com`, and we will be happy to send you a copy or answer any questions. If you are interested in volunteering, please contact our current Chair, Richard Cook, at `rjcook@uwaterloo.ca`. If you are interested in serving on the LIDA leadership team next year, please contact Mei-Cheng Wang, our Past Chair and the chair of the Nominating Committee for this fall's elections, at `mcwang@jhu.edu`.

Jonathan Siegel, Secretary

A Note from Industry: Estimands

The FDA, EMA, and other regulatory authorities plan to issue an addendum to the general guidance on "Statistical principles in clinical trials" (E9), which will be the first update to this guidance since it was issued in 1998. The EMA has released a draft of this guidance (guideline in EU nomenclature) for comment, entitled "ICH E9 (R1) Addendum on estimands and sensitivity analyses in clinical trials to the guideline on statistical principles for clinical trials" (`http://goo.gl/vgiu8W`).

The guidance focuses on the concept of estimands.

Traditional time-to-event analysis methods rely on an assumption that the reasons events are not observed in censored patients are non-informative, i.e., that the reasons for drop-out or non-observation are independent of the effects being estimated. Clinical trialists have increasingly come to conclude that these types of assumptions are often inappropriate to the clinical trial context. Far from being non-informative, withdrawal from clinical trials or inability to come to the clinic to be assessed is often highly related to treatment and disease effects of interest. It can be highly informative.

The draft guidance introduces a nomenclature to describe informative effects that result in the non-observation of the event of interest. Informative withdrawal and similar events are classified as intercurrent events. Intercurrent events are conceptualized as qualitative treatment outcomes which should be regarded as representing data about the treatment, and hence not appropriately classified as "missing data" or "censoring" at all, at least not within the traditional meaning of these terms. An estimand, a term derived from causal inference literature, describes the effect of interest taking intercurrent effects appropriately into account.

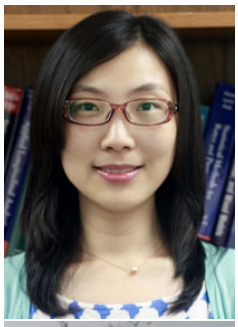
The draft guidance recognizes that there is currently no completely satisfactory methodology available to implement this conceptual framework. Current potential methods discussed include continuing to observe patients through the intercurrent events; defining composite events (e.g., the earlier of the event of interest or the intercurrent event); a "while on treatment" strategy (regarding only events prior to the intercurrent events as of interest), and casual inference methods. But all of these methods have flaws. Measurement through intercurrent events is often impossible if e.g. assessment requires a clinic visit and the events prevent the patient from coming to the clinic. Composite events permit measuring what is observable, but not necessarily what is of interest, especially when informativeness is partial. A "while on treatment" strategy simply ignores the informativeness of intercurrent events. And causal inference methods often rely on strong assumptions that may not be applicable to a study context.

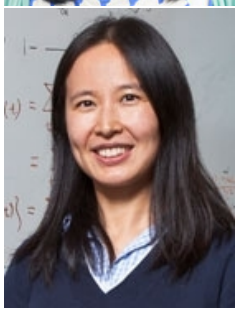
The draft guidance emphasizes sensitivity analyses. As each potential approach relies on strong and potentially inapplicable assumptions, sensitivity analyses provide a way to check their appropriateness and defensibility in a study context.

The framers of the draft guidance recognize the inadequacies of current methods. While seeking to maximize reliable estimation within currently available approaches, the proposed conceptual framework also provides an opportunity to the academic community to develop creative new methods that fully make the conceptual switch from regarding intercurrent events as random noise to regarding them as qualitative informative data, and perhaps provide better ways to solve the problem.

Jonathan Siegel, Secretary

2019 Conference on Lifetime Data Science in Pittsburgh





The 2019 conference on Lifetime Data Science will be held at the University of Pittsburgh on May 29–June 1, 2019 (Wednesday to Saturday). The first conference was successfully convened at the University of Connecticut (2017). The 2019 conference will feature 4 workshops, 3 plenary talks, approximately 40 invited sessions, and a poster session. The conference will cover a broad range of topics in lifetime data science such as biased sampling schemes, biomarkers, causal inference, clinical trials, competing risks, dependence models, epidemiological methods, frailty models, high dimensional data, infectious disease, interval censoring, joint modeling, measurement error, missing data, mixture models, multistate processes, prediction, pregnancy studies, recurrent events, transplantation, and truncation.

Drs. Ying Ding and Yu Cheng from the University of Pittsburgh will be the co-chairs of the local organizing committee with additional members Drs. Chung-Chou H. Chang, Wei Chen, and Jong H. Jeong. More information about the conference will come. We encourage you all to mark your calendar and plan to attend the event!

Ying Ding and *Yu Cheng*, Co-Chairs, Local Organization

JSM 2018 Program Update

We thank the invited session organizers who selected LIDA-IG as either primary sponsor or a co-sponsor for JSM 2018. One invited session, "Non- and Semiparametric Methods to Accommodate Dependency and Heterogeneity in Complex Data," organized by Dr. Naisyin Wang was co-sponsored by LIDA-IG.

Those proposals not selected as invited sessions were entered into the competition for topic-contributed sessions. This year LIDA-IG received one allocation for a topic-contributed session, the same as last year. We had multiple such proposals in the competition pool, in which one proposal organized by Dr. Qingning Zhou was selected as the topic-contributed session with LIDA-IG as the primary sponsor. The title of this session is "Recent Advances in Design and Analysis of Two-Phase Studies". Another topic-contributed session approved for the JSM entitled "First-hitting-time based Threshold Regression and Applications," was organized by Dr. Mei-Ling Ting Lee and co-sponsored by LIDA-IG.

If you missed the deadline this year, please consider starting early for next year and choose LIDA-IG as one of the sponsors.

Yu Shen, Program Chair 2018

News from Lifetime Data Analysis

Every year, Springer summarizes a publisher's report for the articles published in *Lifetime Data Analysis*. From the report of 2017, the top-downloaded article in 2016 for publication years 2014–2016 was "The versatility of multistate models for the analysis of longitudinal data with unobservable features" by Vernon T. Farewell and Brian D. M. Tom, Volume 20, Number 1 (January 2014). The most-cited 2013–2014 articles for Impact Factor Year 2015 was "Applying competing risks regression models: An overview" by Bernhard Haller, Georg Schmidt, and Kurt Ulm, Volume 19, Number 1 (January 2013).

The January 2018 issue (Volume 24, number 1) of *Lifetime Data Analysis* is a special issue dedicated to Jack Kalbfleisch (LIDA-IG Chair 2016). The journal can be accessed at `https://link.springer.com/journal/10985`. Articles appearing in this issue include:

- Nonparametric estimation of the multivariate survivor function: the multivariate KaplanMeier estimator. Ross L. Prentice, Shanshan Zhao. Pages 3-27
- Two-phase outcome-dependent studies for failure times and testing for effects of expensive covariates. J. F. Lawless. Pages 28-44
- Conditional screening for ultra-high dimensional covariates with survival outcomes. Hyokyoung G. Hong, Jian Kang, Yi Li. Pages 45-71
- Variable selection and prediction in biased samples with censored outcomes. Ying Wu, Richard J. Cook. Pages 72-93
- Joint analysis of interval-censored failure time data and panel count data. Da Xu, Hui Zhao, Jianguo Sun. Pages 94-109
- Alternating event processes during lifetimes: population dynamics and statistical inference. Russell T. Shinohara, Yifei Sun, Mei-Cheng Wang. Pages 110-125
- Joint modeling of survival time and longitudinal outcomes with flexible random effects. Jaeun Choi, Donglin Zeng, Andrew F. Olshan, Jianwen Cai. Pages 126-152
- Modeling of semi-competing risks by means of first passage times of a stochastic process. Beate Sildnes, Bo Henry Lindqvist. Pages 153-175
- Modeling restricted mean survival time under general censoring mechanisms. Xin Wang, Douglas E. Schaubel. Pages 176-199

Mei-Ling Ting Lee, Editor-in-Chief, Lifetime Data Analysis

The Fourth International Workshop on Statistical Analyses of Multi-Outcome Data

The Fourth International Workshop on the Statistical Analyses of Multi-outcome Data (SAM 2018) will be held in the Washington University in St. Louis on June 11–12, 2018 (Monday & Tuesday). The first three meetings were successfully convened in the University of Paris VI (2012), Cambridge University (2014), and Renmin University of China in Beijing (2016). We also have a European sister meeting in the University of Liverpool in 2017 and the University of Manchester in 2019.

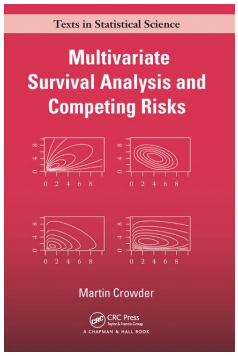
Twenty invited speakers (including two keynote speakers Drs. Jeremy Taylor and Richard Cook) will present interesting topics on joint models, high dimensional outcomes, dynamic prediction modeling, and recurrent event data. There will be a tutorial on the Statistical Analyses of Multi-Outcome Data before the meeting. Student posters are welcome which will be entered into a competition. Winners will receive a prize. Please visit `https://sites.wustl.edu/sam2018/` for more details.

The workshop is co-sponsored by the Division of Biostatistics, Washington University in St. Louis, ASA Lifetime Data Interest Group, and the Midwest Chapter of the International Chinese Statistical Association. We are looking forward to seeing you in St. Louis!

Lei Liu, Washington University in St. Louis

Book Review

Multivariate Survival Analysis and Competing Risks



Martin Crowder. Publisher: CRC Press, Taylor & Francis Group, Boca Raton, FL 2016. ISBN 9781439875216.

Dr. Crowder's book provides a comprehensive overview of various distributions and models that have been proposed for multivariate survival data and competing risks data. The book is divided into four parts. The first part provides necessary background in univariate survival analysis, and introduces many parametric and semi- or non-parametric methods for continuous and discrete event times. The second part focuses on multivariate survival data, and covers numerous advanced topics such as frailty and random effects, recurrent events, and multi-state processes. The third part of the book is devoted to the discussions of competing risks problems, the identifiability issues, different quantities to describe competing risks, and various models and estimation methods. The last part of the book provides intuitive introduction to counting processes without too many technical details.

The material discussed in each topic is succinct, which provides enough detail to follow and includes relevant references for interested audience to dig deeper. The accompanying real-world data examples and R code are extremely handy for anyone who would like to explore those methods. The data and R code are conveniently available at `www.crcpress.com/product/isbn/9781439875216`. The exercises and the hints immediately after each section are great reinforcement of the material.

Despite the advanced topics it covers, the book is a pleasant read, as the presentation style is entertaining. For example, in the preface of Part III on competing risks, Dr. Crowder describes the two different viewpoints of competing risks data with such vivid language that you feel the pain of the losing underlying processes, for which nothing is recorded, with a helpful reminder "History tends to ignore the runners up."

All together the book is very suitable for an advanced survival analysis course for its broad range of topics.

Yu Cheng
Department of Statistics, University of Pittsburgh

Software Review

Accelerating the Convergence of EM Algorithm in Survival Analysis using SQUAREM

The EM algorithm (Dempster 1977) is a ubiquitous approach for obtaining maximum likelihood estimates in a wide array of applications. Many estimation problems in survival analysis with censored data can be solved using the self-consistency principle, which results in the EM algorithm for nonparametric or semi-parametric maximum likelihood estimation (Mykland 1996; Tarpey 1996).

Primary reasons for the popularity of EM are its simplicity and stability. The stability of EM comes from the fact that it provides a monotonic increase of likelihood. However, in many applications the EM algorithm exhibits agonizingly slow, linear rate of convergence, often requiring hundreds, and even thousands, of iterations to converge.

A decade ago, we developed a class of iterative schemes, called squared iterative methods (SQUAREM), to accelerate the convergence of EM (Varadhan 2008). SQUAREM is especially attractive in high-dimensional problems due to its minimal storage requirements. We discuss the `R` package `SQUAREM` for accelerating EM algorithms. Package `SQUAREM` is easy to use. It is stable like the parent EM algorithm. Thus, we can obtain fast convergence without sacrificing the hallmark stability of EM.

The goal of this article is to introduce the readers of LIDA to this remarkable technique for speeding up convergence of EM algorithms, and to promote its use in survival analysis estimation problems.

A pseudocode of SQUAREM acceleration is shown below in Algorithm 1 gives an indication of its simplicity.

Algorithm 1 Pseudocode for SQUAREM

Require: initial value θ_0
1: **repeat**
2: $\theta_1 = \text{EMupdate}(\theta_0)$
3: $\theta_2 = \text{EMupdate}(\theta_1)$
4: $r = \theta_1 - \theta_0$
5: $v = (\theta_2 - \theta_1) - r$
6: Compute steplength $\alpha = \frac{||r||}{||v||}$
7: If necessary, modify α (for global convergence)
8: $\theta' = \theta_0 - 2\,\alpha\, r + \alpha^2\, v$
9: $\theta_0 = \text{EMupdate}(\theta')$
10: **until** convergence

Although one can naively implement the pseudocode given above, it is strongly advised to use the `R` package `SQUAREM` since it has built-in safeguards for ensuring global convergence (step 7).

Brief Description of R Package SQUAREM Package `SQUAREM` has been available on CRAN since 2010. It can be downloaded via `https://cran.r-project.org/web/packages/SQUAREM/index.html`. It works for any smooth, contraction mapping with a linear convergence rate. The main function is `squarem()` which implements EM acceleration:

```
library(SQUAREM)
str(squarem)

## function (par, fixptfn, objfn, ...,
##     control = list())
```

This function accelerates any smooth, contractive, fixed-point iteration algorithm including EM/MM and other EM-like algorithms. The main arguments include `par`, `fixptfn`, and `objfn`. The first argument `par` denotes a vector of starting values for the parameters. The argument `fixptfn` defines a function representing the fixed-point iteration: $\theta^{k+1} = F(\theta^k)$, which a single iteration step of the EM algorithm. In fact, it is quite easy to create this function from the existing `R` code for EM algorithm that the user has already written. All one has to do is extract the part of `R` code that is contained inside the iterative loop for EM algorithm (this code should compute a single EM step), and write it as a separate function.

The optional argument `objfn` defines an objective function that we are going to minimize. In the case of EM algorithms, it would be the negative log-likelihood function of data. Even though an objective function is not necessary, and `squarem` performs equally well without it, global convergence is not guaranteed without it.

A Toy Example: Poisson Mixtures We consider the famous data from The London Times during the years 1910–1912 reporting the number of deaths of women 80 years and older. We construct dataset `dat` containing the number of days n_i, (`freq`) in which i (`ndeath`) deaths occured.

```
dat <- data.frame(
  ndeath = 0:9,
  freq = c(162, 267, 271, 185, 111, 61, 27, 8, 3, 1))
```

A two-component mixture of Poisson distribution provides a good fit to the data, whereas a single Poisson distribution had a poor fit. The likelihood for this problem can be written as

$$\prod_{i=0}^{9} \left[p f(i;\mu_1) + (1-p) f(i;\mu_2) \right]^{n_i},$$

where $f(\cdot;\mu)$ is the probability mass function of a Poisson distribution with mean μ. We have to estimate three parameters, $\theta = (p, \mu_1, \mu_2)$. The EM algorithm is as follows:

$$p^{(k+1)} = \sum_i n_i \, \hat{\pi}_{i1}^{(k)} \Big/ \sum_i n_i,$$

$$\mu_j^{(k+1)} = \sum_i i \, n_i \, \hat{\pi}_{ij}^{(k)} \Big/ \sum_i n_i \, \hat{\pi}_{ij}^{(k)}, \; j = 1, 2,$$

where $\hat{\pi}_{ij}^{(k)} = p^{(k)} f(i;\mu_j^{(k)}) / [p^{(k)} f(i;\mu_1^{(k)}) + (1-p^{(k)}) f(i;\mu_2^{(k)})]$, $j = 1, 2$.

The MLEs are: $(\hat{p}, \hat{\mu_1}, \hat{\mu_2}) = (0.3599, 1.256, 2.663)$. The EM converges very slowly for this problem. Now, I will demonstrate how easy it is to set up Squarem acceleration of any EM-like algorithm using this example. Imagine that we implement basic EM algorithm using the function `EM.poisMix` below.

```
EM.poisMix <- function(par, maxiter = 5000,
                       tol = 1e-08, dat) {
    conv <- FALSE
    for (iter in 1:maxiter) {
        pnew <- poisMix1step(par, dat)
        if (sqrt(crossprod(pnew - par)) < tol) {
            conv <- TRUE; break
        }
        par <- pnew
    }
    list(par = pnew, fpevals = iter,
         convergence = conv)
}
```

The function `poisMix1step` called in the EM iteration above performs a single EM step, which is to be fed to `fixptfn` in calling the `squarem` function:

```
poisMix1step <- function(par, dat) {
    pnew <- rep(NA,3)
    i <- dat$ndeath; y <- dat$freq
    z1 <- par[1] * dpois(i, par[2])
    z2 <- (1 - par[1]) * dpois(i, par[3])
    zi <- z1 / (z1 + z2)
    pnew[1] <- sum(y * zi) / sum(y)
    pnew[2] <- sum(y * i * zi) / sum(y * zi)
    pnew[3] <- sum(y * i * (1-zi)) / sum(y * (1-zi))
    pnew
}
```

The objective function, which is the negative loglikelihood, has a closed-form for this example.

```
poisMixNLLK <- function(par, dat) {
    i <- dat$ndeath; y <- dat$freq
    llk <- y * log(par[1] * dpois(i, par[2]) +
                   (1 - par[1]) * dpois(i, par[3]))
    - sum(llk)
}
```

We use the starting value $(p, \mu_1, \mu_2) = (0.3, 1, 5)$ and a parameter convergence tolerance of 10^{-8}. The classic EM will be compared with `squarem`, which will be called with and without specification of `objfn`. We use package `microbenchmark` to summarize the timing performance of 100 repetitions.

```
p0 <- c(0.3, 1, 5) # initial value
library(microbenchmark)
res <- microbenchmark(
    f1 <- EM.poisMix(par = p0, dat = dat),
    f2 <- squarem(par = p0,
       fixptfn = poisMix1step,
       control = list(tol = 1.e-08), dat = dat),
    f3 <- squarem(par = p0,
       fixptfn = poisMix1step, objfn = poisMixNLLK,
```

```
       control = list(tol = 1.e-08), dat = dat))

## mean time in milliseconds
summary(res)[, "mean"]

## [1] 71.739377  3.502687  3.997744

## parameter estimates are almost the same
cbind(f1$par, f2$par, f3$par)

##           [,1]      [,2]      [,3]
## [1,] 0.3598864 0.3598859 0.3598859
## [2,] 1.2560968 1.2560960 1.2560960
## [3,] 2.6634055 2.6634050 2.6634050

## the number of EM steps are very different
c(f1$fpevals, f2$fpevals, f3$fpevals)

## [1] 2696   54   54
```

The output shows the dramatic improvement for SQUAREM over basic EM algorithm. SQUAREM outperforms basic EM for this case by a factor of 50 in terms of the number of EM evaluations (54 EM steps compared to 2696 EM steps). Also, note that both algorithms converged to essentially the same parameters and the same maximum log-likelihood.

Semiparametric Regression in Survival Analysis Here we show an example of the usefulness of `squarem` for accelerating EM algorithms in lifetime data models. A flexible, semiparametric proportional hazards regression model was presented in Wang et al. (2016) for interval censored data. We refer to the paper for all the details. We ran 100 simulations, with a convergence tolerance of 10^{-7}, for the last scenario shown in Table 1 of the paper (we are grateful to Chris McMahan of South Carolina University for sharing his R code). Results are summarized in Table 1. The speed up by SQUAREM is very impressive — a 15-fold acceleration. A sample plot (Figure 2) clearly shows the faster convergence of SQUAREM. Those interested in running the `R` code for this example can contact me.

Table 1: Averages of 100 replicates in the acceleration of EM algorithm for interval censoring semiparametric regression model (Wang et al. 2016).

	EM	SQUAREM
# EM Steps (IQR)	2324 (1263, 2554)	155 (113, 163)
log-likelihood	−388.1589	−388.1589
CPU (sec)	2.35	0.18

SQUAREM was also applied in a most recent work on semiparametric estimation of accelerated means model for panel count data (Chiou et al., 2017), with impressive results where it increased the convergence rate by a factor of 5.

Summary SQUAREM is a powerful computational tool for speeding up the convergence of EM algorithms. It is as robust as the original EM algorithm in terms of providing reliable convergence to the fixed-point. Hence, we can obtain fast convergence

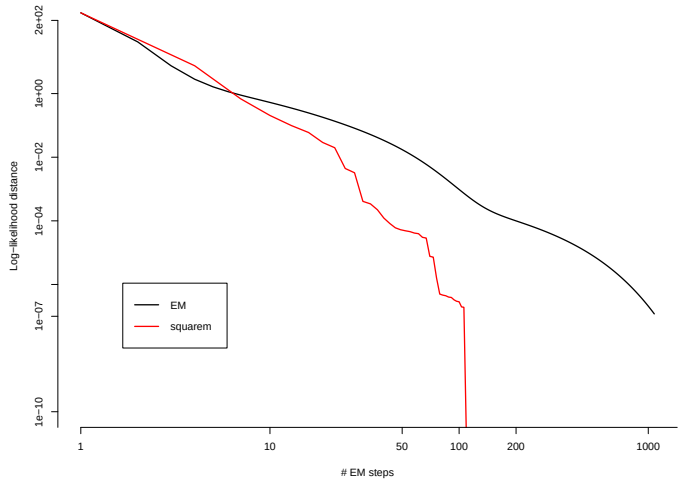


Figure 2: Convergence of EM and SQUAREM to the maximum log-likelihood. On the Y-axis is the distance from the converged maximum log-likelihood value, and on the X-axis is the number of EM evaluations.

without sacrificing the hallmark stability of EM. In other words, *you can have the cake and eat it too!*

I encourage the modelers in life-time data analysis to take advantage of this powerful and easy-to-use tool to speed up their EM algorithm.

References

Chiou S, Xu G, Yan J, and Huang CY (2017). Semiparametric estimation of accelerated mean model with panel count data under informative examination times. *Biometrics*, In press. `http://dx.doi.org/10.1111/biom.12840`

Dempster AP, Laird NM, and Rubin DB (1977), Maximum likelihood from incomplete data via the EM algorithm (with discussion), *Journal of the Royal Statistical Society B*, 39: 1–38.

Mykland PA, and Ren JJ. (1996), Algorithms for computing self-consistent and maximum likelihood estimators with doubly censored data, *The Annals of Statistics*, 24: 1740–1764.

Tarpey T, and Flury B. (1996), Self-consistency: a fundamental concept in statistics, *Statistical Science*, 11: 229-243.

Varadhan R, and Roland C., Simple and globally convergent methods for accelerating the convergence of any EM algorithm. *Scandinavian Journal of Statistics,* 35: 335–353.

Wang L, McMahan CS, Hudgens MG, and Qureshi ZP (2016). A flexible, computationally efficient method for fitting the proportional hazardds model to interval-censored data. *Biometrics*, 72: 222–231.

Ravi Varadhan
Johns Hopkins University School of Medicine