한국통계학회논문집
8권 기념호, 2001
pp. 5-12

STATISTICS PRESENT, NEAR FUTURE, AND BEYOND

Richard A. Johnson[1]

Abstract

We begin with a brief review of some important advances made in statistical theory over the last decade. The choice of topics is decidedly influenced by personal interests. Based on this review, we then propose some possible scenarios about the future of statistics.

1. Introduction

Before suggesting some possible scenarios for the future of statistics, we review some of the major advances in the last decade of the twentieth century. This review forms the basis for our predictions. There is no claim for comprehensiveness as that approach would likely result in a book length manuscript. Instead, the selection of topics can be considered as more of a personal view of the past ten years in statistics.

2. Advances in Large Sample Theory

The classical central limits theorems were established in the first half of the twentieth century and they provided a method to study the large sample properties of estimators and likelihood ratio statistics. Le Cam developed the theory of locally asymptotically normal families, contiguity approaches to large sample theory, and the comparison of experiments during the middle of the century. Most of the effort in the past decade concerns more specialized advances.

For instance, Yeo and Johnson(2000, 2001) while studying the asymptotic properties of inference procedures related to the new transformation

$$\psi(\lambda, x) = \begin{cases} ((x+1)^{\lambda} - 1)/\lambda & \text{for } x \geq 0, \lambda \neq 0, \\ \log(x+1) & \text{for } x \geq 0, \lambda = 0, \\ -((-x+1)^{2-\lambda} - 1)/(2-\lambda) & \text{for } x < 0, \lambda \neq 2, \\ -\log(-x+1) & \text{for } x < 0, \lambda = 2. \end{cases}$$

developed a uniform strong laws of large numbers for U-statistics, $U_n(\boldsymbol{\theta})$, which depend upon a parameter $\boldsymbol{\theta}$. Almost surely, the convergence is uniform in $\boldsymbol{\theta}$. Cho ital(2001) also give obtain the first uniform strong law in a regression setting. Uniform convergence leads directly to the conclusion that, the maximum of n^{-1}(*Log likelihood*) converges to the maximum of the limit thus establishing consistency.

Fu used the techniques of Markov chains to develop results for runs and patterns(for example Fu (1996)).

Horvath and co-workers used the KMT almost sure representations of Brownian bridges and Brownian motion to obtain numerous results including the asymptotic properties of smoothed quantile processes(see Csorgo and Horvath (1993, 1995)).

In another direction, Deheuvels has established numerous functional limit theorems as well as limiting results for kernel density estimators (see Deheuvels(2000) and Deheuvels, P. and Einmahl, J. H. J. (2001)).

[1]Department of Statistics, University of Wisconsin, Madison, WI 53706

Moving beyond the usual time series models, Roussas and co-workers developed methods of nonparametric inference for stochastic processes that satisfy $\alpha-$mixing or strong mixing conditions. (see Roussas(1990) and Roussas, Tran, and Ioannides(1992). These extensions to mixing processes are natural extensions of the earlier Roussas papers on Markov processes.

Inference for another kind of stochastic process, the associated processes, has also been developed recently. Bagai and Rao(1991) and Roussas(1991) are the first papers to address this issue and Roussas has remained a leader in this area of intensive activity (c.f. Roussas(2000)). A finite collection of random variables $\{X_t, t \in I\}$ is said to be positively associated if, for any real-valued coordinatewise increasing functions $F(\cdot)$ and $G(\cdot)$ defined on R^I

$$Cov\ (F(X_t, t \in I), G(X_s, s \in I) \geq 0$$

provided $E\ F^2(X_t, t \in I) < \infty$ and $E\ G^2(X_s, s \in B) < \infty$. If I is not finite, then the non-negativity must hold for all finite dimensional subsets.

An especially important topic of current interest is statistical inference for random fields which obey either a mixing or association dependence condition. Roussas(1994) studies asymptotic normality under positively associated or negatively associated random fields.

In another direction, Koul(1992) developed a comprehensive approach to large sample theory using weighted empirical processes. When Y_{ni}, $i = 1, ..., n$ are independent with Y_{ni} distributed as $F_{ni}(\cdot)$, the weighted empirical process depending on the fixed weights d_{ni}, $i = 1, ..., n$ is given by

$$\sum_{i=1}^{n} d_{ni}\ I(Y_{ni} \leq y)$$

The weights need not be non-negative. In a regression setting, where Y_{ni} depends on the values $\boldsymbol{x}_{ni} = (x_{ni1}, ...x_{nip})'$ of p predictor variables, the weighted empirical process takes the form

$$V_j(y, \boldsymbol{t}) = \sum_{i=1}^{n} x_{nij}\ I(Y_{ni} \leq y + \boldsymbol{x}_{ni}'\boldsymbol{t}) \quad \boldsymbol{t} \in R^p,$$

for $j = 1, ..., p$. A significant advantage of viewing regression and autoregressive models via certain weighted empirical processes is that it leads to important and more efficient inference procedures. In particular, this approach proved very useful in advancing the development of the asymptotic theory of robust inference procedures corresponding to non-smooth score functions from linear models to nonlinear regression and autoregressive models, as well as the theory of regression quantiles under linear AR(p) models.(see Koul and Ossiander (1994), Koul and Saleh(1995) and Koul (1996)).

3. Advances in Nonparametric and Semi-parametric Modeling

One useful variant of the classical multivariate c sample problem, called a quantile test, was posed by Johnson, Sim, etal (199). In the two-sample setting, it concerns the comparison of two treatments where one treatment can produce a few large readings in one or all responses. For each variable, the statistic only counts the number of observations from the first treatment that exceed the combined sample 90-th percentile. When comparing a control area with a contaminated site, there may be a few samples with high heavy metal readings from the contaminated site. The quantile test effectively detects this type of difference, a mixture of two densities.

Widespread attention has been given to developing efficient estimation procedures for semi-parametric models such as

$$Y_i = \boldsymbol{\beta}' \boldsymbol{x}_i + h(\boldsymbol{z}) + \epsilon_i$$

where the linear part is of interest and a smooth function $h(\cdot)$ models the effect of addition variables $\boldsymbol{z}$. See Bickel ital(1993) for details of the statistical theory. One major goal is to determine whether the parametric part can be estimated with high efficiency in the presence of the nonparametric term.

More generally, the full mean $m(\boldsymbol{x})$ of Y can be modeled as $h(\boldsymbol{x})$. The special additive model takes the from

$$m(\boldsymbol{x}) = h_1(x_1) + h_2(x_2) + \cdots + h_q(x_q)$$

where the $h_i(\cdot)$ are unspecified except that they belong to a smooth class of functions. This leads to smoothing based on splines (see Wahba (1990)) and Wavelet techniques.

Another issue of importance is determining the structural dimension about Y that is available in a p =variate vector of predictor variables $\boldsymbol{X}$. Consider the model where a possibly transformed response $Y^{(\lambda)}$, in obvious notation, satisfies the model

$$Y^{(\lambda)} = m(\boldsymbol{\beta}' \boldsymbol{x}) \ + \ \sqrt{v(\boldsymbol{\beta}_2{}' \boldsymbol{x})}\ \epsilon$$

where the mean $M(\cdot)$ and variance $v(\cdot)$ depend on the predictor variables in a linear fashion. Cook(2001) defines dimension as the number of linear combinations needed to explain all the information about Y that is available in a $p-$variate predictor vector $\boldsymbol{X}$. The model above has dimension 1 and the model

$$Y^{(\lambda)} = m(\boldsymbol{\beta}_1' \boldsymbol{x}, \boldsymbol{\beta}_2' \boldsymbol{x}) \ + \ \sqrt{v(\boldsymbol{\beta}_1' \boldsymbol{x}, \boldsymbol{\beta}_2' \boldsymbol{x})}\ \epsilon$$

has structural dimension 2. The idea is to select a few linear combinations and thereby reduce the dimension, without loss of information. Models that have structural dimension 3 or smaller can represent the data graphically so that visual inspection is based on full information. The theory uses ideas from sliced inverse regression due to Li(1991).

4. Advances in Linear, Log-linear and Generalized Linear Models and Inference

The multivariate general linear model

$$\boldsymbol{Y} = \boldsymbol{X}\boldsymbol{\beta} + \boldsymbol{\epsilon}$$

where $\boldsymbol{\epsilon}$ is distributed as a normal distribution with mean $\mathbf{0}$ and covariance $\boldsymbol{\Sigma}$ is well understood.

In spatial applications, the response $\boldsymbol{Y}(\boldsymbol{u})$, the predictor variables $\boldsymbol{x}(\boldsymbol{u})$, and the errors $\boldsymbol{\epsilon}(\boldsymbol{u})$ all depend on the location $\boldsymbol{u}$. Even in the single response case, the covariance matrix $V(\boldsymbol{\theta})$ for the response has entries $cov(y(u_i), y(u_k))$. Some progress has been made on creating $V(\boldsymbol{\theta})$ with useful structure and developing inference procedures but much more remains to be done.

Slightly more general are the hierarchical models. A vector $\boldsymbol{Y}$ of observations depends on an unobserved random effects vector $\boldsymbol{U}$. The joint distribution is assumed to be of a parametric form that depends on the unknown parameters $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \boldsymbol{\theta}_2)$. Further, the conditional distribution of the observable vector $\boldsymbol{Y}$, given the unobservable random vector

$\boldsymbol{u}$ has the density, or probability mass function, $f(\boldsymbol{y}|\boldsymbol{u};\boldsymbol{\theta}_1)$. At the next level of hierarchy, we specify a parametric density $g(\boldsymbol{u};\boldsymbol{\theta}_2)$ for $\boldsymbol{U}$.

Starting with independent random variables $(\boldsymbol{Y}_i, \boldsymbol{U}_i)$, $i = 1, 2, ..., n$, where $(\boldsymbol{Y}_i, \boldsymbol{U}_i)$ is distributed as $f_i(\boldsymbol{y}_i|\boldsymbol{u}_i;\boldsymbol{\theta}_1)$ and $\boldsymbol{U}_i$ is distributed as $g_i(\boldsymbol{u}_i;\boldsymbol{\theta}_2)$. That is, both the conditional and marginal distributions can depend on i. Then, the likelihood is given by

$$L(\boldsymbol{\theta}; \boldsymbol{y}_1, ..., \boldsymbol{y}_n) = \prod_{i=1}^{n} \int f_i(\boldsymbol{y}_i|\boldsymbol{u}_i;\boldsymbol{\theta}_1) g_i(\boldsymbol{u}_i;\boldsymbol{\theta}_2) d\, \boldsymbol{u}_i$$

Under the related Bayesian approach, a prior density $\pi(\boldsymbol{\theta})$ is assumed for the parameters. Then, the posterior density

$$\pi(\boldsymbol{\theta} \mid \boldsymbol{y})_1, ..., \boldsymbol{y}_n) \propto \pi(\boldsymbol{\theta}) L(\boldsymbol{\theta}; \boldsymbol{y}_1, ..., \boldsymbol{y}_n)$$

$$= \pi(\boldsymbol{\theta}) \prod_{i=1}^{n} \int f_i(\boldsymbol{y}_i|\boldsymbol{u}_i;\boldsymbol{\theta}_1) g_i(\boldsymbol{u}_i;\boldsymbol{\theta}_2 d\, \boldsymbol{u}_i$$

Some rather complicated models have been fit using Markov chain Monte Carlo methods. These models are quite general and interesting applications have been given (see Gilfand ital (1990)).

5. Advances in the Markov Chain Monte Carlo Methods

The 1990's witnessed a revolutionary approach to model fitting and the complexity of models that can be posited. Although Markov chain Monte Carlo(MCMC) methods have existed for over 50 years, use of these methods has exploded in the last decade. The book by Robert and Casella(1999) gives many details and applications. In the context of a Bayesian setting, the key idea of the MCMC approach is to generate a sequence of random vectors $\boldsymbol{\theta}^{(t)} = (\boldsymbol{\theta}_1^{(t)}, ..., \boldsymbol{\theta}_r^{(t)})$. The sequence of $\boldsymbol{\theta}^{(t)}$ should follow a Markov chain model that is irreducible, ergodic, and stationary where the stationary distribution, $p(\cdot)$, is the distribution of interest. Then, by the ergodic theorem,

$$\lim_{T\to\infty} \frac{1}{T} \sum_{t=1}^{T} h(\boldsymbol{\theta}^{(t)}) = \int h(\boldsymbol{\theta})\, p(\boldsymbol{\theta}) d\boldsymbol{\theta}$$

for any integrable $h(\cdot)$. That is, the integral can be approximated by an average just as in the case of independent samples.

5.1 Gibbs Sampling

Gelfand and Smith(1990) sets the Gibbs sampler within the framework of Markov chain Monte Carlo techniques in a manner that applies to estimation of very general statistical models.

The Gibbs sampler is particularly applicable to evaluating posterior distributions. It applies to situations where the form of the density is known up to a multiplying constant. That is, the joint density

$$f(\boldsymbol{\theta}) = \frac{g(\boldsymbol{\theta})}{\int g(\boldsymbol{\theta}) d\boldsymbol{\theta}}$$

where $g(\cdot)$ is known. All simulation approaches to investigating the form and properties of the joint distribution, $f(\cdot)$, proceed by generating very large samples from the distribution.

Then, any feature of the distribution will be well approximated by the corresponding feature of the sample. Non-iterative methods of generating samples seem to suffer the curse of dimensionality and do not work well for high dimensional joint densities.

The Gibbs sampler is an iterative method that divides the high dimensional problem into lower dimensional sub-problems.

The Gibbs sampler proceeds as follows:

Partition $\boldsymbol{\theta}$ into k blocks so $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_r)$. At stage t, let $\boldsymbol{\theta}^{(t)} = (\boldsymbol{\theta}_1^{(t)}, ..., \boldsymbol{\theta}_r^{(t)})$. We then make a transition to $\boldsymbol{\theta}^{(t+1)}$ by sampling from appropriate conditional distributions.

$$
\begin{array}{llll}
\text{draw} & \boldsymbol{\theta}_1^{(t+1)} & \text{from} & f(\boldsymbol{\theta}_1 |\, \boldsymbol{\theta}_2^{(t)} ..., \boldsymbol{\theta}_r^{(t)}) \\
\text{draw} & \boldsymbol{\theta}_2^{(t+1)} & \text{from} & f(\boldsymbol{\theta}_2 |\, \boldsymbol{\theta}_1^{(t)}, ..., \boldsymbol{\theta}_3^{(t)} ..., \boldsymbol{\theta}_r^{(t)}) \\
 & & \vdots & \\
\text{draw} & \boldsymbol{\theta}_r^{(t+1)} & \text{from} & f(\boldsymbol{\theta}_r |\, \boldsymbol{\theta}_1^{(t)}, ..., \boldsymbol{\theta}_{r-1}^{(t)})
\end{array}
$$

This gives one complete update of the random vector by the Gibbs sampler. The distributions $f(\boldsymbol{\theta}_i |\, \boldsymbol{\theta}_1, ..., \boldsymbol{\theta}_{i-1}, \boldsymbol{\theta}_{i+1}, ..., \boldsymbol{\theta}_r)$ are called the complete, conditional distributions.

The key to the procedure is that replaces the sampling of a complete vector $\boldsymbol{\theta}$ by successively sampling the lower dimensional blocks $\boldsymbol{\theta}_i$.

Convergence of the algorithm remains an issue. It is best to get samples of $\boldsymbol{\theta}$ from separate independently started iterations. The sample can then use standard procedures for estimating the features of the joint density. Alternatively, some researches generate only one long sequence of random vectors by iteration and, after a burn-in period, use, say, every $m-$th term where m is large. There will be some dependence in this sample and it should be checked by calculating the sample autocorrelation functions. Smith and Roberts(1993) give some relatively simple sufficient conditions on $g(\cdot)$ so that Gibbs sampler converges.

There are some difficulties with the application of Gibbs sampling, and MCMC methods in general. They have been applied to models that are far too complex for the amount of data available and with inadequate attention to convergence questions.

One software package(Speigelhalter, Thomas, Best and Gilks(1995)), called BUGS, is quite general and quite widely used for teaching and research.

5.2 Perfect Simulation

One variation of an MCMC algorithm, called perfect simulation or exact sampling, provides in finite time actual independent draws from the stationary distribution of a Markov chain. To date, it works best with finite state space Markov chains. The technique is based on a backward coupling method due to Propp and Wilson (1996). When the goal is to generate samples from $f_X(x)$, the distribution can be augmented by introducing a random variable Y, taking only a finite number of values, and considering the joint distribution $f(x, y)$. The second variable should be chosen in such a way that both conditional distributions $f(x |\, y)$ and $f(y |\, x)$ are easy to sample. A two step Gibbs sampler induces a Markov chain on Y

Because Y has a finite state space, in some cases the Propp and Wilson coupling method can be used to obtain independent and identically distributed observations from $f_Y(y)$. For each $Y = y$ generated, an X can be selected from $f(x \mid y)$. Meng(2000) gives an extension of the algorithm to multistage backward coupling.

6. Where Are We Going?

The history of recent major advances suggests some pessimism. Statisticians were not the developers of some of the novel computational algorithms. The Gibbs sampler was developed primarily by physicists and applied mathematicians trying to model complex physical, chemical, or biological systems. Computer intensive inference procedures have become a primary research area. Although statisticians have a better understanding of uncertainty and errors, they may lack access to the fastest computers common to computer science departments.

We suggest three possible scenarios for the future of statistics. The most pessimistic scenario is given first.

*S*cenario 1. The preponderance of research will permanently shift to intensive computer based modeling and inference. The line between statistics and a sub-field of computer science becomes even more blurred. Many departments of statistic will vanish and be replaced by small groups within computer science. If electrical and computer engineering is in the same unit, advances may even be made when hot new algorithms can be immediately converted to "hard wired" circuits consisting of a specialized chip or even a biological circuit. The point being that these other subject areas may have the competitive advantage with computing hardware of the latest type.

*S*cenario 2. Things continue as they currently are for a few decades. The new statistical theory being developed is becoming very specialized. Faster computers allow for enough replicates that error bounds can be added to cases where only point estimators are currently available because of the complexity of the biological or physical system under investigation. Advances in statistics will take place in the important applications areas of health, computational biology, and genetics. Statisticians become even more valuable as members of collaborative research teams attacking major problems in the sciences. However, it may be difficult to maintain an active intellectual stimulation until the end of the 21-st century unless major breakthroughs in statistical theory continue to occur.

*S*cenario 3. In the 21-st century, statisticians will come closer to a unified theory of inference. This unified theory is likely to have heavy Bayesian component and it will need to include a decision theory component. Frequency based ideas will be necessary to evaluate the properties of new procedures. Whether the major language remains frequentist or reverts more to Bayesian with an element of likelihood theory may still not be decided. For the Bayesian side, it is well known that optimal statistical procedures are Bayesian.

New techniques for asymptotic analysis will be developed that pertain to studying the properties of complicated new statistical algorithms. Mathematical expansions should be able to suggest some bounds or estimate of error in forecasts and parameter estimates.

With these new tools, statistics will remain a field of active research and experts in statistics will help in most scientific investigations across all experimental applications.

Rather than make a choice between the three scenarios, as a statistician I prefer to put a probability distribution on the probabilities of the scenarios. For instance, one subjective prior distribution is a Dirichlet distribution with mean .2 for p_1, .4 for p_2 and .3 for p_3 with perhaps .1 for an entirely new Scenario.

A profession that trains experts to make decisions under substantial uncertainties and to access the amount of uncertainty in parameter estimates and uncertainties in forecasts will undoubtedly be viable at the end of the 21-st century. Statistics will continue to be a fascinating subject and, with its endless opportunities it participate in collaborative research across all fields of science, will continue to draw bright students.

REFERENCES

Bagai, I and Prakasa Rao, B. L. S. (1991), Estimation of the Survival Function for Stationary Associated Processes, *Statistics and Probability Letters*, **12**, 385-391.

Bickel, P. Klaassen, C. Ritov, Y. and J. Wellner (1993), *Efficient and Adaptive Estimation for Semiparametric Models* Johns Hopkins University Press, Baltimore, Md.

Cho, K. Yeo, I. Johnson, R. A., Loh, W.Y. (2001), Asymptotic Theory for Box-Cox Transformations in Linear Models, *Statistics and Probability Letters* **51**, 337-343.

Cook, R. D.(2001), Dimension Reduction and Graphical Exploration in Regression. To appear *Biometrics.*

Csorgo, M. and Horvath, L. (1993), *Weighted Approximations in Probability and Statistics*, John Wiley, New York

Csorgo, M. and Horvath, L. (1995), On the Distance Between Smoothed and Empirical Quantile Processes, *Annals Statistics*, **23**, 113-131.

Deheuvels, P.(2000), Uniform Limit Laws for Kernel Density Estimators on Possibly Unbounded Intervals, *Recent Advances in Reliability-Methodology, Practice and Inference*, ed. Limnios, L and M Nikulin, Birkhauser, Boston. 477-492

Deheuvels, P. and Einmahl, J. H. J.(2001), Functional Limit Laws for Increments of Kaplan-Meier Product-Limit Processes and Applications, to appear *Annals Probability.*

Fu, J. (1996), Distribution of Runs and Patterns Associated with a Sequence of Multi-state Trials, *Statistica Sinica*, **6**, 957-974

Gelfand, A. E. and Smith, A. F. M. (1990), Sampling-Based Approaches to Calculating Marginal Densities, *J. American Statistical Association*, **85**, 398-409.

Johnson, R. A. , Sim, S., Klein, B. and R. Klein(1998), A Multivariate MultiSample Quantile Test for Ordered Alternatives, **93**, 807-818.

Koul, H. L. (1992), *Weighted Empiricals and Linear Models* , **21**, Lecture Notes-Mongraph Series, Institute of Mathematical Statistics, Hayward, California.

Koul, H. L. (1996), Asymptotics of some estimators and sequential residual empiricals in non-linear time series. *Annals Statistics*, **24**, 380-404.

Koul, H. L. and Ossiander, M. (1994), Weak convergence of randomly weighted dependent residual empiricals with application to autoregression. *Annals Statistics*, **22**, 540–562.

Koul, H. L. and Saleh, A.K.Md.E. (1995), Autoregression quantiles and related rank-scores processes. *Annals Statistics*, **23**, 670-689.

Li, K. C.(1991), Sliced Inverse Regression for Dimension Reduction(with discussion). *Journal of the American Statistical Association*, **86**, 316-342.

Meng, X-L. (2000), Toward a More General Propp-Wilson Algorithm: Multistage Backward Coupling, *Fields Institute Communications Series Vol. 26: Monte Carlo Methods*, American Mathematical Society.

Propp, J. G., and Wilson,D. B. (1996), Exact Sampling with Coupled Markov Chains and Applications to Statistical Mechanics, *Random Structures and Algorithms*, **9**, 223-252.

Robert, C. P., and Casella, G. (1999), *Monte Carlo Statistical Methods*, Springer-Verlag, New York.

Roussas, G. G. (1990), Nonparametric Regression Estimation Under Mixing Conditions, *Stochastic Processes and Their Applications*, **36**, 107-116.

Roussas, G. G. (1991), Kernel Estimates under Association: Strong Uniform Consistency, *Statistics and Probability Letters*, **12**, 392-403

Roussas, G. G. (1994), Asymptotic Normality of Random Fields of Positively and Negatively Associated Processes, *Journal of Multivariate Analysis*, **50**, 152-173.

Roussas, G. G. (2000), Asymptotic Normality of the Kernel Estimate of a Probability Density Function under Association, *Statistics and Probability Letters*, **50**, 1-12.

Roussas, G. G., Tran, L. T., and Ioannides, D. A. (1992), Fixed Design Regression for Time Series: Asymptotic Normality, *Journal of Multivariate Analysis*, **40**, 262-291.

Smith, A. F. M., and and Roberts, C. O. (1993), Bayesian Computation Via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods,*J. Royal Statistical Soc, Ser. B*, **55**, 3-23.

Spiegelhalater, D. J., Thomas, A., Best N., and Gilks, W. R.(1995), BUGS: Bayesian Inference Using Gibbs Sampling, Version 0.50, Medical Research Council, Biostatistics Unit, Cambridge, U. K

Yeo, I. and Johnson, R. A.(2000), A New Family of Power Transformations to Improve Normality or Symmetry, *Biometrika*, **87**, 954-959.

Yeo, I. and Johnson, R. A.(2001), A Uniform Strong Law of Large Numbers for U-Statistics with Application to Transforming to Near Symmetry, *Statistics and Probability Letters*, **51**, 63-39.

Wahba, G. (1990), *Spline Models for Observational Data*, Vol. 59 in the CBMS-NSF Regional Conference Series in Applied Mathematics. SIAM, Philadelphia