EFaB@Bayes250 ◇ Tutorials

Sunday 15th December 2013

Hedibert Lopes (University of Chicago)

Sequential Monte Carlo

I will review sequential Monte Carlo (SMC) methods, or particle filters (PF), with special emphasis on its potential applications in financial time series analysis and econometrics. I start with the well-known normal dynamic linear model, also known as the normal linear state space model, for which sequential state learning is available in closed form via standard Kalman filter and Kalman smoother recursions. Two standard examples, the AR(1) plus noise and the SV-AR(1) plus noise models, will serve as motivating scenarios, with the latter playing the role of a whistle blower to the need of Monte Carlo (Markov chain-based or particle-filter based) tools when either or both normality and linearity are violated. Particle filters are then introduced as a set of Monte Carlo schemes that enable Kalman-type recursions when normality or linearity or both are abandoned. The seminal bootstrap filter (BF) of Gordon,Salmond and Smith (1993) is used to introduce the SMC jargon, potentials and limitations. We also review the literature on parameter learning, an area that started to attract much attention from the particle filter community in recent years. We give particular attention to the LiuWest filter (2001), Storvik filter (2002) and particle learning (PL) of Carvalho, Johannes, Lopes and Polson (2010). We argue that the BF and the auxiliary particle filter (APF) of Pitt and Shephard (1999) define two fundamentally distinct directions within the particle filter literature. We conclude the tutorial discussing the most recent contributions to the that combine MCMC and PF tools and provide a tentative projection to where the area is heading in the near future.

Deepak Agarwal (LinkedIn)

EFaB Tutorial: Statistical Challenges in Web Recommender Systems

Algorithmically matching items to users in a given context is essential for the success and profitability of large scale recommender systems like content optimization, computational advertising, search, shopping, movie recommendation, and many more. The objective is to maximize some utility (e.g. total revenue, total engagement) of interest over a long time horizon. This is a bandit problem since there is positive utility in displaying items that may have low mean but high variance. A key challenge in such bandit problems is the curse of dimensionality. Bandit problems are also difficult to work with for responses that are observed with considerable delay (e.g. return visits, confirmation of a buy). One approach is to optimize multiple competing objectives in the short-term to achieve the best long-term performance. For instance, in serving content to users on a website, one may want to optimize some combination of clicks and downstream advertising revenue in the short-term to maximize revenue in the long-run. In this talk, I will discuss technical challenges and solutions by focusing on a number of concrete applications - content optimization on the Yahoo! front page and LinkedIn, display advertising on Right Media ad exchange and LinkedIn, stream recommendation on LinkedIn. We show that large scale regression and latent factor models coupled with simple randomization schemes like Thompson sampling (drawing samples from the posterior distribution to rank items) works well in practice.

EFaB@Bayes250 ◇ Invited Talks

Sunday 15th & Monday 16th December 2013

Greg Allenby (Ohio State University)

Latent Topic Modeling of Consumer Reviews: Linking Text Evaluations to Customer Satisfaction and Brands

Firms collect an increasing amount of consumer feedback in the form of unstructured consumer reviews. These reviews contain text about consumer experiences with products and services that are different from surveys that query consumers for specific information. A challenge in analyzing unstructured consumer reviews is in making sense of the topics that are expressed in the words used to describe these experiences. We propose a new model for exploratory text analysis that makes use of the sentence structure contained in the reviews, and show that it leads to improved inference and prediction relative to existing models using data from www.expedia.com. The topics associated with different levels of satisfaction and hotel brands are found to be different, more distinguished and more interpretable that those emerging from alternative analysis.

Siddhartha Chib (Washington University)

Estimating and Comparing Affine Term-Structure Models

Arbitrage-free affine term-structure models are widely used in the dynamic modeling of the yield curve of default-free bonds. From a statistical perspective, these models have the form of Gaussian state space structure that is linear in unobserved and unobserved factors but nonlinear in the deep parameters of the model, such as those that measure the market price of factor risks. On account of the nonlinearity of the parameters, the likelihood function can be quite irregular, and the consequent fitting can be fraught with difficulties. In this talk, building on prior work in Chib and Ergashev (2009), we describe ways in which the Bayesian perspective can be helpful in surmounting those difficulties. One way is through careful modeling of the prior distribution to reflect meaningful economic information, such as the belief that the yield curve is upward sloping on average. Another is through careful application of MCMC fitting methods. In this connection, we demonstrate the value of the Tailored Randomized Block Metropolis-Hastings (TaRB-MH) method in the fitting process and contrast the gains that accrue from this approach with those from other M-H variants, such as the fixed block M-H method, and the Hamiltonian Monte Carlo method. Finally, the Bayesian perspective is helpful in the comparison of alternative affine models. The ideas are illustrated with both simulated and real data on affine models that contain up to five factors and ten yields. Joint work with Kyu Ho Kang.

Sylvia Frühwirth-Schnatter (Vienna University of Economics & Business)

Time-Varying Parameter Models – Achieving Shrinkage and Variable Selection

Time-varying parameter (TVP) models are a popular tool for handling data with smoothly changing parameters. However, in situations with many parameters the flexibility underlying these models may lead to overfitting models and, as a consequence, to a severe loss of statistical efficiency. This occurs, in particular, if only a few parameters are truely time-varying, while the remaining ones are constant or even insignificant. As a remedy, hierarchical shrinkage priors have been introduced for TVP models to allow shrinkage both of the initial parameters as well as their variances toward zero. The present paper contributes to the literature in two ways. First, we investigate shrinkage for TVP models based on the Normal-Gamma prior which has been introduced recently by Griffin and Brown (2010) for standard regression models. Our approach extends Belmonte, Koop, and Korobilis (2011) who considered the Bayesian LASSO prior, a special case of the Normal Gamma prior. While both priors reduce the risk of overfitting and increase statistical efficiency, they do not allow for variable selection. Hence, as a second contribution, we follow Frhwirth-Schnatter and Wagner (2010) and consider TVP models with spike-and-slab priors which explicitly incorporate variable selection both with respect to the initial parameters as well as their variances. Following Belmonte et al. (2011), hierarchical shrinkage priors as well as spike-and-slab priors are applied to EU area inflation modelling based on the generalized Phillips curve. Since the corresponding time series are relatively short, variable selection through the spike-and-slab priors is particularly sensitive to the choice of hyperparameters.

Mark Jensen (Atlanta Federal Reserve)

Mutual Fund Performance: Learning About the Generic Alpha

This paper provides a generalization of Jones and Shanken (2005) "Mutual fund performance with learning across funds." Jones and Shanken adopt a hierarchical prior for the alphas of mutual funds that allows for learning about the hyperparameters and consequent shrinkage of the posterior distributions. They show how posterior distributions for specific cases depends on the priors for the hyperparameters. In that paper the primary estimands are the posterior distributions for the specific cases. This paper theme

could be stated as, "Taking the prior (even) more seriously." This paper investigates the features of the prior that affect the rate at which learning occurs. It relaxes the restrictions imposed by the parametric form for the conditional prior for alpha, thus allowing for multiple modes and fat tails. This paper adopts the notion of the generic case where; (i) the generic alpha applies to any mutual fund for which one has no data, (ii) it is the generic alpha to which the prior applies, (iii) the prior can be interpreted as the density for the generic case with no data, (iv) we provide what could be called density estimation with noisy measurements. To that end we adopt a nonparametric prior that allows for great flexibity and also allows one to control the rate at which learning occurs.

Sylvia Kaufmann (Study Center Gerzensee, Switzerland)

Capturing Changing Dynamics at the Zero Lower Bound

Since the outbreak of the financial crisis, the policy rates of major central banks have reached the zero lower bound and have remained at historical low levels so far. In this situation, even if monetary authorities are able to accomodate new incoming shocks by implementing unconventional measures, it is conceivable that the effect of shocks and the effectiveness of policy measures may change dramatically. The effect of shocks which are of a transitory nature when interest rates are away from the zero lower bound, might become permanent when they decrease to very low levels. On the other hand, the transmission of policy impulses to the real economy might change, effects might be delayed, or effects might become smaller or stronger. One possibility to capture these changes in policy-relevant models, is to use models incorporating a time-varying parameter process. Here, the parameters are driven by a latent switching process. The latent process itself is driven by time-varying state probabilities determined by a set of variables that are typically considered to be the major determinants of a central bank's policy like the interest rate level, the inflation rate and the output gap. The Bayesian model specification also takes into account explicitly that the interest rate is bounded below by zero. In fact, the setup can be adapted to situations in which unconventional monetary policy targets directly prices in specific asset markets, e.g. government bonds, mortgages or currency markets. The model estimate then allows to determine at which level the interest rate entered the critical zero lower bound region. The estimate also allows to compute state-contingent impulse responses, and due to the specification with time-varying state probabilities, we may compute forecasts about the probability to leave the zero lower bound region again.

Robert Kohn (University of New South Wales)

Importance Sampling Squared for Bayesian Inference in Latent Variable Models

We consider Bayesian inference by importance sampling when the likelihood is analytically intractable but can be unbiasedly estimated. We refer to this procedure as importance sampling squared (IS2), as we can often estimate the likelihood itself by importance sampling. We provide a formal justification for importance sampling when working with an estimate of the likelihood and study its convergence properties. We analyse the effect of estimating the likelihood on the resulting inference and provide guidelines on how to set up the precision of the likelihood estimate in order to obtain an optimal tradeoff between computational cost and accuracy for posterior inference on the model parameters. We illustrate the IS2 method in empirical applications for a generalised multinomial logit model and a stochastic volatility model. Our results show that the IS2 method can lead to fast and accurate posterior inference when optimally implemented.

Fei Liu (Queen's University, NYC)

A Bayesian Markov-switching Model for Sparse Dynamic Network Estimation

Inferring Dynamic Bayesian Networks (DBNs) from multivariate time series data is a key step towards the understanding of complex systems as it reveals important dependency relationship underlying such systems. Most of the traditional approaches assume a "static" DBN. Yet in many relevant applications, such as those arising in biology and social sciences, the dependency structures may vary over time. In this paper, we introduce a sparse Markov-switching vector autoregressive model to capture the structural changes in the dependency relationships over time. Our approach accounts for such structural changes via a set of latent state variables, which are modeled by a discrete-time discrete-state Markov process. Assuming that the underlying structures are sparse, we estimate the networks at each state through the hierarchical Bayesian group Lasso, so as to efficiently capture dependencies with lags greater than one time unit. For computation, we develop an efficient algorithm based on the Expectation-Maximization method. We demonstrate the strength of our approach through simulation studies and a real data set concerning climate change.

Marco del Negro (New York Federal Reserve)

Time-varying Prediction Pools: An Investigation of Financial Frictions and Forecasting Performance

We provide a methodology for estimating time-varying weights in optimal prediction pools, building on the work by Geweke and Amisano (2011, 2012). We use this methodology to investigate the relative forecasting performance of DSGE models with and without financial frictions from 1992 to 2011 at various forecast horizons. Our results indicate that models with financial frictions produce superior forecasts only in periods of financial distress. This may explain why macroeconomists may have neglected models with financial frictions prior to the Great Recession.

Nick Polson (Chicago Booth)

Waiting Time Methods in Financial Econometrics

We provide a brief survey of waiting time modeling in financial econometrics from a statistical perspective. Alternative types of waiting time definitions are juxtaposed and various modeling trade-offs are considered. In particular, we emphasize the merits of waiting time modeling as a natural data reduction technique for ultra high frequency financial data.

Steve Scott (Google)

Consensus Monte Carlo, Bayes, and Big Data

A useful definition of "big data" is data that is too big to fit on a single machine, either because of processor, memory, or disk bottlenecks. Graphics processing units can alleviate the processor bottleneck, but memory or disk bottlenecks can only be alleviated by splitting "big data" across multiple machines. Communication between large numbers of machines is expensive (regardless of the amount of data being communicated), so there is a need for algorithms that perform distributed approximate Bayesian analyses with minimal communication. Consensus Monte Carlo operates by running a separate Monte Carlo algorithm on each machine, and then averaging the individual Monte Carlo draws. Depending on the model, the resulting draws can be nearly indistinguishable from the draws that would have been obtained by running a single machine algorithm for a very long time. Examples of consensus Monte Carlo will be shown for simple models where single-machine solutions are available, for large single-layer hierarchical models, and for Bayesian additive regression trees (BART).

Justin Tobias (Purdue University)

Priors and Posterior Computation in Linear Endogenous Variable Models

Estimation in models with endogeneity concerns typically begins by searching for instruments. This search is inherently subjective and identification is generally achieved upon imposing the researcher's strong prior belief that such variables have no conditional impacts on the outcome. Results obtained from such analyses are necessarily conditioned upon the untestable opinions of the researcher, and such beliefs may not be widely shared. In this paper we, like several studies in the recent literature, employ a Bayesian approach to estimation and inference in models with endogeneity concerns by imposing weaker prior assumptions than complete excludability. When allowing for instrument imperfection of this type, the model is only partially identified, and as a consequence, standard estimates obtained from the Gibbs simulations can be unacceptably imprecise. We thus describe a substantially improved "semi-analytic" method for calculating parameter marginal posteriors of interest that only requires use of the well-mixing simulations associated with the identifiable model parameters and the form of the conditional prior. Our methods are also applied in an illustrative application involving the impact of Body Mass Index (BMI) on earnings.

Toshiaki Watanabe (Hitotsubashi University)

Bayesian Analysis of Multiple Structural Changes in ARFIMA Models with an Application to Realized Volatility

The analysis of structural change in the time series data which has the long-memory property is important because the long-memory property may possibly be caused by structural change. This article develops a method for the analysis of multiple structural changes in such data. The irreversible Markov switching (MS) model proposed by Chib (1998) is incorporated in the autoregressive fractionally integrated moving average (ARFIMA) model. The Hamilton filter is no longer applicable to the resuting model because the ARFIMA model is an AR model with infinite lag-length. A Bayesian method is developed where the parameters and the state variable representing the number of structural changes up to each period are sampled from their posterior distribution using the particle Markov chain Monte Carlo (PMCMC) method proposed by Andrieu, Doucet and Holenstein (2010). The marginal likelihood is used for selecting the number of change-points and analyzing whether the long-memory property is spuriously caused by structural change or not. Several researchers have documented that the realized volatility (RV) defined as the sum of squared intraday returns has the long memory property. Our method is illustrated by applying to the RVs of several financial returns.

EFaB@Bayes250 ◇ New Researcher Talks

Monday 16th December 2013

Kyu Ho Kang (Korea University)

Forecasting the Term Structure of Interest Rates with Potentially Misspecified Models

Since Diebold and Li (2006) showed the outstanding performance of a dynamic Nelson-Sieglel model (DNSM) in forecasting the yield curve, the DNSM has been widely used in many macro and finance area. Because of its parsimonious but flexible model specification the Bayesian model-averaging method based on the Bayes factor typically gives a weight of nearly one on the DNSM excluding a standard arbitrage-free affine term structure model (ATSM). Nevertheless, the ATSM has been also commonly used because it provides plenty of economically interpretable outcomes such as term premium and model-implied term structure of real interest rates. Meanwhile, the random-walk (RW) is often used as a benchmark in out-of-sample forecasting comparison. Despite the popularity of these three frameworks, none of them dominates the others across all maturities and forecast horizons. This fact indicates that those models are potentially misspecified. In this paper we investigate whether combining the possibly misspecified models in a linear form suggested by Geweke and Amisano (2011) and Waggoner and Zha (2012) help improve the predictive accuracy. For this we compare out-of sample prediction performance from the merged models with a constant model weight with those of the three individual prediction models and the merged models with a Markov-switching model weight for eight different maturities and forecast horizons of 1, 3, 6 and 12 months. We find that overall the constant mixture model is most supported. In particular, the constant mixture model consistently forecasts better than the individual prediction models across all maturities and forecast horizons.

Juan Carlos Martínez-Ovando (Banco de Mexico)

An Extended Time- and State-Dependent Price-Setting Model with Stickiness

A main premise of macroeconomic models is that the microeconomic behavior of the optimizing agent mimics the dynamics of macroeconomic agents. In recent years, micro-data evidence has exhibited that such a premise might be wrong; e.g. individual firms typically delay their decision to adjust their prices in response to macroeconomic stimulus. In this paper, we develop a micro-founded model to describe the dynamics of firms price adjustments within sectors. The model is time-dependent in the sense that it acknowledges heterogeneity among firms within sectors; i.e. durations and price adjustments are both random. Our model is derived as a marked duration model for durations and magnitudes of price adjustments. The model is also state-dependent, in the sense that it allows the incorporation of variables of the current state of the economy as regression variables for their dynamics. The model specification is micro-founded using micro-price data of Mexican firms. We link the stochastic rule of firms price adjustment in a standard DSGE model to assess its implication for monetary policy.

Jouchi Nakajima (Bank of Japan)

Bayesian Latent Threshold Dynamic Models: Identifying Conventional and Unconventional Monetary Policy Shocks

This work proposes a new estimation framework for identifying monetary policy shocks in both conventional and unconventional policy regimes using a structural vector autoregression (VAR) model. We extend recently introduced latent threshold dynamic models to include a recursive identification switching with a time-varying constraint for the interest rate zero lower bound, with an ability to induce time-varying sparsity into time-varying parameters in the VAR. We empirically analyze Japans monetary policy to illustrate the proposed approach for modeling regime-switching between conventional and unconventional monetary policy periods. The estimation results show that increasing bank reserves lowers long-term interest rates in the unconventional policy periods, and that the impulse responses of inflation and the output gap to a bank reserve shock appear to be positive but highly uncertain.

Xia Wang (University of Cincinnati)

Investigating the Impact of Customer Stochasticity on Firm Price Discrimination Strategies Using a New Bayesian Mixture Scale Heterogeneity Model

Price discrimination strategies by firms often involve calibration of price sensitivity of consumers through heterogeneous discrete choice models. The role of the scale coefficient of one such widely used model, the 'mixed logit' model, has come under recent scrutiny in the marketing literature. Not normalizing the scale coefficient to one, as is customary in the mixed logit model, can lead to better predictive model fit and insights about the level of customer stochasticity. An open question not studied in the above papers is: how does customer stochasticity, as parameterized by a heterogeneous scale coefficient, impact firm price discrimination

strategies. In this paper we develop a new model that employs a mixture model approach to incorporate both parameter heterogeneity and customer stochasticity into a novel discrete choice method. We cast the model in a fully Bayesian hierarchical framework, where estimation of model parameters is accomplished through Markov Chain Monte Carlo (MCMC) simulations from their posterior distributions. We term this new model the Bayesian Mixture Scale Heterogeneity (BMSH) model, discuss identifiability of the model, and also show simulation evidence that the model is identified using artificial data. Our BMSH model has comparable/better prediction properties as compared to the widely used mixed logit model, when calibrated on an empirical scanner dataset from the yogurt category. We show through targeting simulations that a firm's price discrimination strategies can be made more efficient using the new model. We also show how the BMSH model can be extended to incorporate learning effects of a consumer in a category.

Hongxia Yang (IBM Watson)

Dynamic Latent Class Model Averaging for Online Prediction

We consider the problem of online prediction when it is uncertain what the best prediction model to use is. We develop a method called Dynamic Latent Class Model Averaging, which combines a state-space model for the parameters of each of the candidate models of the system with a Markov chain model for the best model. We propose a polychotomous regression model for the transition weights to assume that the probability of a change in time depends on the past through the values of the most recent time periods and spatial correlation among the regions. The evolution of the parameters in each submodel is defined by exponential forgetting. This structure allows the "correct model to vary over both time and regions. In contrast to existing methods, the proposed model naturally incorporates clustering and prediction analysis in a single unified framework. We develop an efficient Gibbs algorithm for computation, and we demonstrate the value of our framework on simulated experiments and on a real-world problem: forecasting IBMs corporate revenue.

EFaB@Bayes250 ◇ Posters

Monday 16th December 2013

Brenda Betancourt (University of California, Santa Cruz)

Modelling & Prediction of Financial Trading Networks: A Case Study in the NYMEX Natural Gas Futures Market

Over the last few years there has been a growing interest in using financial trading networks to understand the microstructure of financial markets. However, most of the methodologies developed so far have been based on descriptive summaries of the networks such as the average node degree and the clustering coefficient. This paper develops novel statistical methods for modeling sequences of financial trading networks. Our approach uses a stochastic blockmodel to describe the structure of the network during each period, and then links multiple time periods using a hidden Markov model. This structure allows us to identify extreme events that affect the structure of the market and make accurate short-term prediction of future transactions. The methodology is illustrated using data from the NYMEX natural gas futures market from January 2005 to December 2008.

Angela Bitto (Vienna University of Economics and Business)

Time-Varying Parameter Models – Achieving Shrinkage and Variable Selection

The present paper contributes to the literature in two ways. First, we investigate shrinkage for Time-Varying Parameter (TVP) models based on the Normal-Gamma prior which has been introduced by Griffin and Brown (2010) for standard regression models. Our approach extends Belmonte, Koop, and Korobilis (2011) who considered the Bayesian LASSO prior, a special case of the Normal Gamma prior. While both priors reduce the risk of over fitting and increase statistical efficiency, they do not allow for variable selection. Hence, as a second contribution, we follow Frhwirth-Schnatter and Wagner (2010) and consider TVP models with spike-and-slab priors which explicitly incorporate variable selection both with respect to the initial parameters as well as their variances. Following Belmonte et al. (2011) we choose EU area inflation modelling based on the generalized Phillips curve as our application. Comparing the predictive evaluation, the Normal Gamma prior significantly outperforms the Bayesian LASSO prior.

Andrea Eross (University of Southampton)

Tracing Liquidity Contagion in the Short-Term Interbank Market using MCMC Analysis

Interbank markets are channels of contagion due to the overlapping claims that banks have on one another. If liquidity dries up in the overnight market as during the Financial Crisis of 2007-08, domino effects transmit liquidity shocks to other markets. Exploring the LIBOR-OIS spread reveals an unusual pattern, as the time series seems to change its behaviour by switching from a stationary I(0) to a non-stationary process. Modelling this behaviour with standard econometric models is deemed to failure; hence, a novel Bayesian approach is developed. We present a two-state Markov chain model that determines the crisis and tranquil periods in the time series. Our model contains the permanent and transitory components without using a decomposition method into liquidity and failure elements. Our approach traces the liquidity risk that built up at the height of the Financial crisis of 2007-08 in the short-term interbank market, and subsequently its contagion to other markets. We show that the variability in the LIBOR-OIS spread shows structural breaks consistent to liquidity crashes and thus acts as an early-warning indicator of an imminent liquidity shortage. Depending on which state the system is in, we model our series either as a first-order autoregressive process or as a Gaussian white noise process. The transition between the states is described by a Markov process and the probability of being in a crisis or non-crisis period is estimated using Bayesian inference.

Gokhan Esen (Washington DC)

Sequential Monte Carlo Analysis of Bivariate Common Factor Model with Particle Learning Filters

This article describes the Sequential Monte Carlo (SMC) analysis of the bivariate common factor model using recently proposed Bayesian filtering technique of Particle Learning (PL) (Carvalho et al., 2010). In state space context, common factor models can be considered as the dynamic generalization of factor analysis (Harvey 1989). They have been shown to successfully extract the common trend in bivariate time series using MCMC (Petris et al., 2010). SMC analysis is developed by extending the Markov Chain Monte Carlo (MCMC) solution of bivariate factor model extracting common stochastic trend (Petris et al., 2010) into a PL framework. Application of both MCMC and PL solutions are entertained in simulated time series as well as US 30-year conventional mortgage rate and 10-year Treasury constant maturity rate.

Marco Ferreira (University of Missouri)

Bayesian Reference Analysis for Exponential Power Regression Models with Applications to Business

We develop Bayesian reference analyses for linear regression models when the errors follow a exponential power distribution. Specifically, we obtain explicit expressions for reference priors for all the six possible orderings of the model parameters. In addition, we show that associated with these six parameters orderings there are only two reference priors. Further, we show that both of these reference priors lead to a proper posterior distribution. Furthermore, we show that the proposed reference Bayesian analyses compare favorably to an analysis based on a competing noninformative prior. Finally, we illustrate our Bayesian reference analysis for exponential power regression models with applications to two datasets. In the first application we analyze excess returns for a publicly traded company. In the second application we study the relationship between sold home videos versus profits at the box office.

Elena Goldman (Pace University)

Dynamic Analysis of 'Too Big to Fail' Risks

Following recent studies of systemic risks by Acharya et. al (2010) and Brownlees and Engle (2012) among others I introduce Bayesian estimation of the dynamic MES (marginal expected shortfall) and SRISK (the expected capital shortage of a firm conditional on a substantial market decline). The rankings for MES and SRISK are used to analyze the systemic risks of financial institutions and are daily reported by the Volatility Institute (http://vlab.stern.nyu.edu). However, this measures are reported without uncertainty around the estimates and thus one cannot distinguish if the rankings of large financial institutions are statistically different. Recent surveys of systemic risk analytics by Bisias et al. (2012) and Brunnermeier and Oehmke (2012) among others also do not show how to measure and incorporate uncertainty for systemic risk measures. First, the present paper shows how to derive the distribution of MES and SRISK using Bayesian Markov Chain Monte Carlo (MCMC) algorithms. Second, I introduce a generalized threshold conditional volatility model (GTARCH) and compare it to traditional asymmetric models of volatility. I propose Bayesian estimation of a GTARCH model and apply it for forecasting volatility of equity returns of financial institutions, market indices and log-differences of Credit Default Swaps (CDS) of banks' secured bonds. Using a new asymmetric GTARCH model and capturing uncertainty around the measures I found that MES and SRISK are statistically different for major financial institutions but the systemic risks are higher when the GTARCH model is used.

Rolando Gonzales (Bayes Foundation)

Modeling Hyperinflation Phenomena: A Bayesian Approach

Hyperinflations are short-lived episodes of economic instability in prices, characteristically lasting twenty months or less. Classical statistical techniques applied to these small samples could lead to incorrect inference problems. This paper describes a Bayesian approach for modeling hyperinflations which improves the modeling accuracy using small-sample inference based on specific parametric assumptions. A theory-congruent model for the Bolivian hyperinflation was estimated as a case study.

Rolando Gonzales (Bayes Foundation)

The Wage Curve, Once More with Feeling: Bayesian Model Averaging of Heckit Models

The sensitivity of the wage curve to model uncertainty is evaluated with Bayesian methods. Over 65 thousand Heckit wage-curves were estimated using data of a developing economy. After averaging the estimates with the posterior probability of each model being true, the wage-curve unemployment elasticity is closer, once more, to –0.1, suggesting that the statistical regularity of the wage curve is robust to the inclusion/exclusion of different control explanatory terms, even in developing economies.

Kaoru Irie (Duke University)

Online Updates of Models with Latent Thresholds

The multivariate dynamic model with latent thresholds is known to be useful to adjust the flexibility of the dynamic parameters and achieve the accurate prediction. However, in situation where the data are observed frequently, it is computationally hard to repeat employing MCMC every time we obtain an additional observation. For the reason above, we make use of approximate Bayesian computation method to update the posterior with the new observation, using the previous posterior estimated by MCMC as the prior. The ABC algorithm is expected to work well because (1) we have already learned a lot about the model with the training dataset and MCMC, and (2) our model has the form of compositional dynamic regression, which allow ABC to be applied to the 1-dimensional observation.

Glen Livingston Jr (University of Newcastle)

A Fully Bayesian Analysis of Smooth Threshold Autoregressive (STAR) Model: A Prior Sensitivity Analysis

The Smooth Threshold Autoregressive (STAR) model is a model used in economics and finance. The sensitivity to the specification of the priors in a fully Bayesian analysis of STAR models is considered. In this talk, we first present a complete Bayesian analysis of

the posterior distribution of STAR models which extends the Bayesian analysis of AR models by Vermaak et al. (2004). Theoretical derivation of the joint posterior distribution followed by the derivation of the full conditional distributions for each parameter are derived in detail. Following this, a posterior simulator using a combination of MCMC algorithms such as the MH algorithm, Gibbs sampler, RJMCMC and MTM algorithm will be described. Finally, we will present some initial simulation studies for the proposed MCMC posterior simulator using different prior distributions especially for the implicit parameters as well as the order of the model.

Xiaoye Ma (University of Minnesota)

A Hybrid Bayesian Hierarchical Model Combining Cohort and Case-control Studies for Meta-analysis of Diagnostic Tests: Accounting for Disease Prevalence and Partial Verification Bias

Bivariate random effects models have been recommended to jointly model the sensitivities and specificities in meta-analysis of diagnostic accuracy studies accounting for between-study heterogeneity. Because the severity and definition of disease may differ from study to study due to the design and the selected population, the sensitivities and specificities of a diagnostic test may be correlated with the disease prevalence. To account for the potential dependence, trivariate random effects models had been proposed. However, the proposed approach can only include cohort studies, which contain the information for estimating study-specific disease prevalence. In addition, some diagnostic accuracy studies only select a subset of samples based on the test results to be verified by the reference test. It is known that ignoring unverified subjects can lead to partial verification bias in the estimation of prevalence, sensitivities and specificities in a single study. However, the impact of this bias on the meta-analysis of diagnostic tests has not been investigated. As many diagnostic accuracy studies use case-control designs, we propose a novel hybrid Bayesian hierarchical model combining cohort and case-control studies to account for prevalence and to correct partial verification bias at the same time. We investigate the performance of the proposed methods through a set of simulation studies, and a case study on assessing the diagnostic accuracy of gadolinium-enhanced magnetic resonance imaging in detecting lymph node metastases.

Juan Carlos Martínez-Ovando (Banco de Mexico)

Consistent Reading of Local Business Cycles

Trend-cycle components of business cycles for the US and other economies are nowadays computed using dynamic factor models (Stock and Watson, 2001). In practice, business cycles for sectorial or regional components of output are computed independently, which may produce inconsistencies between local and aggregate readings of the business cycles. Crone and Clayton-Matthews (2005) proposed a heuristic procedure to compute consistent business cycles for the US states. In this paper, we develop a nested procedure for computing consistent local and aggregate business cycles to study output dynamics. The model is specified as a hierarchical dynamic factor model, which aims at controlling and isolating idiosyncratic changes in the local business cycles, and synchronizing local cycles with changes in the aggregate business cycle. We illustrate the usefulness of our model using the Philadelphia Feds data for the US states. The business cycles estimated with our model result useful, as well, to monitor with more accuracy the dynamics of the aggregated business cycle, and its regions, when they are incorporated in a recently developed generalized diffusion index.

Kenichiro McAlinn (Duke University)

Fully Parallel Particle Learning for GPGPUs and Other Parallel Devices

We developed a novel parallel algorithm for particle filtering (and learning) which is specifically designed for GPUs (graphics processing units) or similar parallel computing devices. In our new algorithm, a full cycle of particle filtering (computing the value of the likelihood for each particle, constructing the cumulative distribution function (CDF) for resampling, resampling the particles with the CDF, and propagating new particles for the next cycle) can be executed in a massively parallel manner. One of the advantages of our algorithm is that every single numerical computation or memory access related to the particle filtering is executed solely inside the GPU, and no data transfer between the GPUs device memory and the CPUs host memory occurs unless for further processing, so that it can circumvent the limited memory bandwidth between the GPU and the CPU. To demonstrate the advantage of our parallel algorithm, we conducted a Monte Carlo experiment in which we applied the parallel algorithm as well as conventional sequential algorithms for estimation of a simple state space model via particle learning, and compared them in terms of execution time. The results showed that the parallel algorithm was far superior to the sequential algorithm.

Kei Miyazaki (Kansai University)

A Research of Brand Switching Behaviors using Latent Class Dynamic Multinomial Probit Models with Random Effects

In recent studies that analyze scanner panel data, hierarchical Bayes modeling with dynamic structures and random effects, introduced for modeling consumers heterogeneity, is often used. Also in this study, dynamic structures are assumed among latent utilities of brands and brand switching behaviors are researched by estimating and interpreting autoregressive coefficients among these latent utilities. However, introducing random effects only could lead to an erroneous conclusion during real data analysis in cases where subpopulations of a substantial size exist. Although finite mixture or latent class models have been used in cases where the existence

of several subpopulations is supposed, in the presence of substantial within-class heterogeneity in coefficients, the finite mixture solution often requires an excessive number of latent classes or subpopulations to represent the heterogeneity in the data adequately, leading to over parameterization and many, relatively small latent classes. In particular, in the proposed model where dynamic structures are assumed among latent utilities, this phenomenon means explaining all the consumers heterogeneity that relates to brand switching behaviors by using only latent classes, which makes it impossible to estimate brand switching behaviors accurately because of an excessive number of latent classes. The proposed model overcomes these limitations because it is a hybrid version of a hierarchical Bayes model with dynamic structures in which both latent classes and random effects are assumed. The proposed method has been applied to an IRI marketing data set with noteworthy results.

Jared Murray (Duke University)

Flexible Bayesian Density Regression without Discrete Mixtures

In many problems researchers are interested in how the entire distribution of a response changes across predictors. Numerous nonparametric Bayes methods have been proposed to address this problem, most often by a) jointly modeling the response and predictors or b) modeling mixture weights and/or atoms as functions of predictors. Joint models are at best inefficient, and hard to justify when predictors are nonrandom. Including covariates in the infinite collection of atoms or weights often leads to models which are difficult to understand, and an explosion of parameters with more than a few covariates. I propose a new method for density regression by expressing the observed response as an unknown smooth function of latent continuous variables which themselves follow a regression model, a generalization of discrete mixtures to the uncountably infinite case. This framework is flexible, parsimonious and in many cases a more realistic model for the process under consideration. The continuous representation also alleviates concerns about mixing over an ultra-high dimensional discrete space in MCMC. Expressing the unknown function as an adaptive linear combination of B splines I obtain computationally efficient posterior inference without discrete approximations. I present examples including hierarchical density models and regression on continuous and categorical covariates, and discuss extensions to more complex settings.

Andriy Norets (University of Illinois)

Credibility of Confidence Sets in Nonstandard Econometric Problems

Confidence intervals are commonly used to describe parameter uncertainty. In nonstandard problems, however, their frequentist coverage property does not guarantee that they do so in a reasonable fashion. For instance, confidence intervals may be empty or extremely short with positive probability, even if they are based on inverting powerful tests. We apply a betting framework to formalize the ”reasonableness” of confidence intervals and use it for two purposes. First, we quantify the degree of unreasonableness of previously suggested confidence intervals. Second, we derive alternative sets that are reasonable by construction. As was already realized in the literature, any bet-proof set must be a superset of a Bayesian credible set relative to some prior. This suggests that attractive bet-proof confidence sets may be obtained by selecting a prior that induces a given type of credible set (such as HPD) to have frequentist coverage. The main theoretical result of this paper is to show that such a prior exists. Previous results for Bayesian sets with frequentist coverage are either for particular families of distributions, for invariant problems, or for asymptotic equivalence of coverage and credibility in LAN models. In contrast, our existence result is entirely generic as it applies to any (discretized) inference problem under very mild regularity conditions. We apply our framework to several nonstandard problems involving a parameter near a boundary, weak instruments, near unit roots, and moment inequalities. We find that most previously suggested confidence intervals are not reasonable, and numerically determine alternative confidence sets that satisfy our criteria.

Alexandra Posekany (Vienna University of Economics and Business)

Applying Standard and Semiparametric Bayesian IV for Health Economic Data

We compare three inferential approaches applied to a very large and challenging data set from health economics. The chosen instrumental variable model aims for determining the causal effect of family size on labour and health outcomes. Contrasting frequentist and Bayesian parametric and semiparametric approaches, we find that the parametric version outperforms the rest regarding computational efficiency and estimation precision. In addition, we present an approach for merging the information of horizontally partitioned data which we apply for the DP sampler's results.

Gregor Semieniuk & Ellis Scharfenaker (The New School for Social Research)

Bayesian Mixture Modeling for Filtering Economic Data with Applications to the Rate of Profit

We extract the “signal” of the profit rate distribution in a U.S. firm level data set using a Bayesian mixture model with latent variables, that track each firm observation's location in the mixture. Using a Laplace distribution as our prior for the signal we are able to model the distribution of the profit rate as a representation of competition among firms in the U.S. economy both pooled and sector-wise. Further, we use a Bayesian change point analysis on the moments of this distribution conditional over the partition of the size of firms to determine the minimum firm size above which entry and exit dynamics no longer skew the distribution. We conclude that the profit rate distribution approximates a statistical equilibrium.

Yan Shang (Duke University)

Nonparametric Bayesian Analysis in International Air Cargo Logistics

Global trade has grown considerably in recent decades and has fuelled the evolution of logistics and supply chain management, many companies now have overseas facilities and supply chain partners. As one of the speedies transportation means, air transport is delivering high quality products at competitive prices to consumers worldwide. However, the reliability of air transport is a real problem. In this paper, we are using a one year air logistics data provided by one the world biggest international logistic company to investigate the root causes of frequent transport delays and disruptions. On one hand, not being able to be loaded onto the scheduled flight at the origin airport or connecting airport is the most severe problem, which happens in 15% of the total shipments. Missing flight could cause days of delays, which depends on the gap between the scheduled flight and the actual flight the cargo eventually takes. As a result, the delay has a complex multimodal structure. On the other hand, airlines' overselling cargo capacity and unreliable operations are the main causes of cargo missing flight. In the data, there are usually 2-3 different airlines providing services on each origin-desination route at very different service levels. We would like to find out the worse and best airlines on each route and overall. Given the interesting feature of the data and the analysis goal, we are seeking a more realistic ANOVA-like dependent nonparametric bayesian process to analyze the delays. So far, we are comparing dependent dirichlet process with dependent probit stick-breaking process. After this, we are going to develop some novel models to better fit the reality, and this should be finished before the presentation in December.

Shinichiro Shirota (Duke University)

Cholesky Realized Stochastic Volatility Model: Flexible Simultaneous Modeling of High-dimensional Covariance Matrices

Cholesky stochastic volatility model enable us to model the high dimensional covariance matrix in the financial time series. This model is based on the cholesky decomposition of covariance matrix, and updating the each diagonal and off diagonal components in the univariate context. We include the cholesky decomposed components of the realized covariance as the information of true covariance matrix. We show the estimation procedures, results and the improved volatility forecasting performance with the minimum variance portfolio comparison.

Matthew Simpson (Iowa State University)

Ancillarity-Sufficiency or not; Interweaving to Improve MCMC Estimation of the Local Level DLM

In dynamic linear models (DLMs), MCMC sampling can often be very slow for estimating the posterior density — especially for longer time series. In particular, in some regions of the parameter space the standard data augmentation algorithm can mix very slowly. Recently ancillarity-sufficiency interweaving has been introduced as a method to take advantage of alternate parameterizations in multilevel models in order to improve the mixing and convergence properties of the chain. Focusing on the local level DLM, we explore alternate parameterizations and various interweaving algorithms through simulation in order to improve mixing. We conclude by explaining what our results may mean for MCMC in a more general DLM.

Sanvesh Srivastava (Duke University)

An Asynchronous Scalable Distributed EM Algorithm For Massive Data: The DEM Algorithm

Massive data with complex latent structures have become common independent of discipline. The computer architectures to store these data are also rapidly evolving. Classical iterative statistical algorithms, such as Expectation-Maximization (EM), for fitting models with latent structures are practically infeasible for these data due to two main reasons: massive size of the data and the large number of parameters required to model the complex dependencies in the data. These two limitations are relaxed by the Distributed Iterative Statistical Computing (DISC) framework presented in this work for implementing iterative statistical algorithms by taking advantage of widely available computing power, such as cluster of computers. Using EM as a concrete example of an iterative algorithm, DISC extends and scales it for massive data as DISC-EM (DEM). We analyze the convergence properties of the sequence of parameter estimates generated by DEM and show that DEM retains the attractive properties of EM: monotone ascent of the log likelihood at each iteration and stability of iterations. DEM can also be easily implemented in cluster and grid computing environments using R package disc and existing EM implementations. To illustrate its application, we use DEM for estimating the effect of movie genres on their ratings in a movie ratings data.

Aixin Tan (University of Iowa)

Estimates and Standard Errors for Ratios of Normalizing Constants from Multiple Markov Chains

In the classical biased sampling problem, we have k densities $\pi_1(\cdot), \ldots, \pi_k(\cdot)$, each known up to a normalizing constant, i.e. for $i = 1, \ldots, k$, $\pi_i(\cdot) = v_i(\cdot)/m_i$ where v_i is a known function and m_i is an unknown constant. For each i, we have an iid sample from π_i, and the problem is to estimate the ratios m_i/m_s for all i, s. This problem arises frequently in several situations in both frequentist and Bayesian inference. An estimate of the ratios was developed and studied by Vardi and his co-workers over two decades ago, and since then there has been much subsequent work on this problem from many different perspectives. In spite of this, there are no rigorous results on how to estimate the standard error of the estimate. We present a class of estimates of the ratios of normalizing constants that are appropriate for the case where the samples from the π_i's are not iid sequences, but are Markov chains. We also develop an approach based on regenerative simulation for obtaining standard errors for the estimates of ratios of normalizing constants. These standard error estimates are valid for both the iid case and the Markov chain case. As an example, we show that the above result is useful in forming importance sampling estimators that are based on multiple Markov chains. We illustrate the method with an application to the Bayesian variable selection problem in linear regression, for which importance sampling enables an empirical Bayes approach to variable selection.

Gracia Toubia-Stucky (College of Coastal Georgia)

A Sequential Bayesian Cumulative Conformance Count Approach to Deterioration Detection in High Yield Processes

The motivation behind this research which looks at the control of high yield processes is: 1) The stringent quality requirement of highly technological products, 2) the importance of monitoring rare events with adverse consequences and quickly responding to process shift/deterioration, 3) the reduction in the environmental consequences of waste and 4) the fact that the process can nowadays be totally and quickly inspected. The Cumulative conformance count (CCC) control chart is a powerful alternative to the traditional p-control chart, particularly in monitoring high yield processes with extremely low proportions of nonconformance. However, a prevalent limitation of the CCC control chart is its inability to detect small process deterioration. A sequential Bayesian CCC approach capable of detecting small process deterioration is proposed. The new approach outperforms the traditional CCC chart in that it does not require a large sample of initial observations of the process, which may be difficult, if not impossible to obtain in practice. Moreover, the approach is self-starting, and thus may be used in short production runs. A Bayesian updating procedure is developed, which allows for the determination of initial control limits based on only three initial observations or some prior knowledge about the proportion of nonconformance of the process. Values of proportions of nonconformance, ranging from 0.1 to 0.00001, are tested to demonstrate the deterioration detection capability of the new approach in conjunction with the proposed deterioration detection rules.

Tao Wang (University of North Carolina, Chapel Hill)

Empirical Analysis of Sequential Trade Models for Market Microstructure

Market microstructure concerns how different trading mechanisms affect asset price formation. It generalizes the classical asset pricing theory under frictionless perfect market conditions in various directions. Most market microstructure models focus on two important aspects: (a) asymmetric information shared by different market participants (informed traders, market makers, liquidity traders, et al.) (b) transaction costs reflected in bid-ask spreads. The complexity of those models presents significant challenges to empirical studies in such a research area. In this work, we consider some extensions of the seminal sequential trade model in Glosten and Milgrom (Journal of Financial Economics, 1985) and perform Bayesian MCMC inference based on the TAQ (trade and quote) database in Wharton Research Data Services. Issues in both (a) and (b) are addressed in our study. In particular, the latent process of fundamental asset value is modeled with GARCH volatilities; the observed and predicted bid-ask price sequences are related by incorporating parameters for pricing errors and for informed traders impact.

Hao Wang (University of South Carolina)

Scaling It Up: Stochastic Search Structure Learning in Graphical Models

Gaussian concentration graph models and Gaussian covariance graph models are two classes of graphical models useful for uncovering latent dependence structures among multivariate variables. In the Bayesian literature, structure learning of graphs is often achieved by priors over the space of positive definite matrices with fixed zeros, but these methods present daunting computational burdens in large problems. Motivated by the superior computational efficiency of continuous shrinkage priors for regression analysis, we propose a new framework for structure learning that is based on continuous spike and slab priors and uses latent variables to identify graphs. We discuss model specification, computation, and inference for both covariance graph models and concentration graph models. The new approach produces reliable estimates of graphs and efficiently handles problems of hundreds of variables.

Jiangyong Yin (Ohio State University)

Portfolio Optimization Using Constrained Hierarchical Bayes Models

It is well-known that traditional mean-variance optimal portfolio delivers rather erratic and unsatisfactory out-of-sample performance because of estimation errors. Regularized solutions, such as "no-shortsale" constrained and norm-constrained portfolios, can usually achieve much higher *ex post* Sharpe ratio. Bayesian methods have also been shown to be superior to traditional plug-in estimator but the outperformance is generally limited. By inducing priors directly on optimal portfolio weights and imposing constraints a priori in our Hierarchical Bayes models, we can construct portfolios that are well diversified with great out-of-sample performance. Our method is tested on a number of Fama-French industry portfolios against naïve diversification strategy, no-shortsale minimal variance portfolio as well as an existing Bayesian model based on economic theory.

Yong Zeng (University of Missouri at Kansas City)

Real-time Bayesian Stochastic Volatility Estimation via Filtering Equation for a Partially-Observed Heston Model

This presentation first briefly reviews a general partially-observed framework of Markov processes with marked point process observations recently proposed for ultra-high frequency data, and the Bayes Estimation via Filtering Equation (BEFE). In recent years, Graphics Processing Units (GPUs) evolved from rendering graphics (linear algebra-like computations) for electronic games and video applications to becoming low-cost and green supercomputing units. With harnessing the newly available GPU high performance computing power in mind and targeting a Heston stochastic volatility model, we develop a new easily-parallelized, uniformly consistent, recursive algorithm via BEFE for propagating and updating the joint posterior distributions. We show that the recursive algorithm is well suited for GPU parallel computing. We present simulation and empirical results obtained from supercomputers to demonstrate that the recursive algorithm works. Real time tracking and feeding stochastic volatility is made possible.

Zoey Yi Zhao (Duke University)

Dynamic Dependence Networks: Multiregression Dynamic Models for Financial Time Series Analysis and Portfolio Decisions

We discuss novel developments in Bayesian analytics for the class of multiregression dynamic models. Our core focus is on increasingly high-dimensional multivariate time series; we aim to enable analysis that can be decomposed into sets of univariate models processed in parallel yet coupled for forecasting and decisions. Our innovations involves sparse conditional dependence networks, idiosyncracies in time-varying auto-regressive lag structures, and flexibility of discounting methods for stochastic volatilities. Conditional dependence networks provide an inherently parallelizable framework under which we develop sequential updating, on-line forecasting and decisions, broadening the current theory and methods of Bayesian dynamic models. Innovations in parallelization resolve issues of computational tractability for increasingly high-dimensional data. Dynamic conditional dependence networks require model structure search and uncertainty analysis; we address this via a novel application of shotgun stochastic search (SSS) concepts. The examples show that this can succeed in effectively exploring spaces of high posterior probability over network structure, while also identifying practically superior predictive and lucrative models in financial application. These examples, drawn from studies of international exchange rate forecasting and predictive portfolio analysis, provide substantial insights into the significant improvements and practical benefits of the new modelling and analysis framework.

Bayes250 Day ◇ Invited Talks

Tuesday 17th December 2013

Stephen Stigler (University of Chicago)

A Posterior Estimate of Thomas Bayes, Conditional on Recently Discovered Evidence

Thomas Bayes's essay was read to the Royal Society of London in December 1763, 250 years ago. For many years it was for the most part ignored, yet it has achieved iconic status in the present era. I shall present new historical evidence regarding three issues. (1) Evidence that indicates the commonly received title for the essay is wrong and provides support for a story of how the essay came to be written and what its initial goal was. (2) Some new evidence regarding the known best (but sometimes disputed) picture of Bayes. (3) A brief statement will be presented from one of the twentieth century's founding Bayesians regarding Ronald Fisher and objective Bayesian inference.

Stephen Fienberg (Carnegie Mellon University)

Two Hundred Years after Bayes: The Neo-Bayesian Revival

The "neo-Bayesian revival" had many deep intellectual roots, especially in the work of Ramsey, de Finetti, Turing, and Jeffreys. But those who brought it to life in the 1950s added elements that led to a small but vigorous intellectual community that helped to form the foundation for the broader Bayesian movement and the ultimate creation of ISBA. In this presentation I focus on some key 1950s contributions, e.g., by I.J. Good, Dennis Lindley, L. Jimmie Savage, and Howard Raiffa and Robert Schlaifer, and the consequences of an amazing confluence of statisticians at the University of Chicago during 1954-1955.

Adrian Smith (University of London)

The Past 50 years of Bayes: A UK and European Perspective

In an interactive interview format, we will review and reflect on the development of Bayesian statistical thinking and influence over the past 50 years. In particular we will provide a UK and European perspective on the philosophical and practical debates and developments that took place over that period and the contributions and influence of key individuals.

Christopher Sims (Princeton University)

Limits to Probabilistic Inference

People preprocess data through accounting schemes, deseasonalization, and time-aggregation. Data are run through pattern-finding algorithms with little attention to probability modeling. Manageable probability models are used even though it is recognized that an unwieldy model would fit better. Real world decision problems are discussed as if the state space were discrete and small. Is this just because not enough people have had a good course in Bayesian inference in high school? Or are there recognizable limits to the applicability of probability-based inference and decision making? We compromise on these issues all the time. This talk tries to stimulate thought on how to keep the compromises principled.

Michael Jordan (University of California, Berkeley)

A Short History of Topic Models

The past two decades have seen major growth in statistical approaches to the analysis of collections of text documents and the modeling of underlying linguistic phenomena. Compared to other applied domains, where Bayesian and non-Bayesian methods are often present in equal measure, here the majority of the work has been Bayesian. I will provide a short history of a particular slice through this line of work—that known as "topic modeling"—emphasizing the whys and wherefores of the Bayesian success stories in this domain, and providing some glimpses into the future.