Plenary Speakers

Asu Ozdaglar (MIT)

Title: *Learning from Reviews*

Abstract:
Many online platforms present summaries of reviews by previous users. Even though such reviews could be useful, previous users leaving reviews are typically a selected sample of those who have purchased the good in question, and may consequently have a biased assessment. In this paper, we construct a simple model of dynamic Bayesian learning and profit-maximizing behavior of online platforms to investigate whether such review systems can successfully aggregate past information and the incentives of the online platform to choose the relevant features of the review system.
On the consumer side, we assume that each individual cares about the underlying quality of the good in question, but in addition has heterogeneous ex ante and ex post preferences (meaning that she has a different strength of preference for the good in question than other users, and her enjoyment conditional on purchase is also a random variable). After purchasing a good, depending on how much they have enjoyed it, users can decide to leave a positive or a negative review (or leave no review if they do not have strong preferences). New users observe a summary statistic of past reviews (such as fraction of all reviews that are positive or fraction of all users that have left positive review etc.). Our first major result shows that, even though reviews come from a selected sample of users, Bayesian learning ensures that as the number of potential users grows, the assessment of the underlying state converges almost surely to the true quality of the good. More importantly, we provide a tight characterization of the speed of learning (which is a contribution relative to most of the works in this area that focus on whether there is learning or not). Under the assumption that the online platform receives a constant revenue from every user that purchases (because of commissions from sellers or from advertising revenues), we then show that, in any Bayesian equilibrium, the profits of the online platform are a function of the speed of learning of users. Using this result, we study the design of the review system by the online platform.

This is joint work with Daron Acemoglu, Ali Makhdoumi, and Azarakhsh Malekian.

Bio:
Asu Ozdaglar received the B.S. degree in electrical engineering from the Middle East Technical University, Ankara, Turkey, in 1996, and the S.M. and the Ph.D. degrees in electrical engineering and computer science from the Massachusetts Institute of Technology, Cambridge, in 1998 and 2003, respectively.

She is the Joseph F. and Nancy P. Keithley Professor of Electrical Engineering and Computer Science (EECS) Department at the Massachusetts Institute of Technology. She is also the associate head of EECS. Her research expertise includes optimization theory, with emphasis on nonlinear programming and convex analysis, game theory, with applications in communication, social, and economic networks, distributed optimization and control, and network analysis with special emphasis on contagious processes, systemic risk and dynamic control.

Professor Ozdaglar is the recipient of a Microsoft fellowship, the MIT Graduate Student Council Teaching award, the NSF Career award, the 2008 Donald P. Eckman award of the American Automatic Control Council, the Class of 1943 Career Development Chair, the inaugural Steven and Renee Innovation Fellowship, and the 2014 Spira teaching award. She served on the Board of Governors of the Control System Society in 2010 and was an associate editor for IEEE Transactions on Automatic Control. She is currently the area co-editor for a new area for the journal Operations Research, entitled "Games, Information and Networks. She is the co-author of the book entitled “Convex Analysis and Optimization” (Athena Scientific, 2003).

Talk Details:

Monday, July 10th, 2017
9:00am - 10:00am
White Auditorium, 2nd Floor

Yuliy Sannikov (Stanford)

Title: *Dynamic Contracts*

Abstract:
Dynamic incentive problems occupy an important place in economics. They appear in many fields. In macroeconomics incentives pose constraints that lead to inequality. In corporate finance, incentives problems lead financial frictions and impose limits on optimal capital allocation. This talk will address the problem of dynamic incentives through the lens of a continuous-time principal agent model. The agent puts in effort, which is observable only imperfectly, and the principal wishes to design the best contract to motivate the agent. The analysis of this problem involves double dynamic optimization: the principal designs the optimal dynamic contract recognizing that the agent will optimize with respect to an effort strategy given the contract. The principal's optimization problem has to use an endogenous state space, with variables sufficient to summarize the agent's incentives. The optimal contract exhibits inefficiencies: the agent has to face risk, inefficient termination may occur with positive probability, and various distortions may need to be imposed to control the agent's value of private information.

Bio:
Yuliy Sannikov is a theorist who has developed new methods for analyzing continuous time dynamic games using stochastic calculus methods. His work has not only broken new ground in methodology, it has had a substantial influence on applied theory. He has significantly altered the toolbox available for studying dynamic games, and as a result of his contributions, new areas of economic inquiry have become tractable for rigorous theoretical analysis. The areas of application include the design of securities, contract theory, macroeconomics with financial frictions, market microstructure, and collusion. Sannikov's work is impressive. It is elegant, powerful, and it paves the way for further analysis on lots of problems. The early successes highlighted how even simple and well-studied models could yield new insight. His most recent work has tackled more complex models in finance and macroeconomics. Previous models abstracted from crucial economic forces in the name of tractability, but Sannikov's methods allow models to include the most important forces and thus deliver results that are much more relevant. He is one of the few theorists in many years to have introduced a truly novel tool that changed the way theory is done.

Talk Details:

Tuesday, July 11th, 2017
9:00am - 10:00am
White Auditorium, 2nd Floor

R. Srikant (UIUC)

Title: *Approximate Graph Matching on Random Graphs*

Abstract:
We consider an abstraction of the network deanonymization problem, where the goal is to infer the node identities in an anonymized graph by observing the node identities and the topology of a correlated graph. More precisely, the goal is to label the nodes of the anonymized graph so that the adjacency matrix of the anonymized graph matches the adjacency matrix of the correlated graph as closely as possible. We will review prior results on this problem for the case of Erdős–Rényi graphs, and present some new results for stochastic block models. Joint work with Joseph Lubars.

Bio:
R. Srikant is the Fredric G. and Elizabeth H. Nearing Endowed Professor of Electrical and Computer Engineering and a Professor in the Coordinated Science Lab, both at the University of Illinois at Urbana-Champaign. His research interests include communication networks, machine learning, and applied probability. He served as the Editor-in-Chief of the IEEE/ACM Transactions on Networking from 2013-2017. He is the winner of several Best Paper awards, and is a recipient of the IEEE INFOCOM Achievement Award.

Talk Details:

Wednesday, July 12th, 2017
12:15pm - 1:15pm
White Auditorium, 2nd Floor

Marcel Neuts Lecture:

Colm O'Cinneide (QS Investors)

Title: *Phase-type Distributions and Invariant Polytopes.*

Abstract:
A phase-type distribution is the distribution of a hitting time in a finite-state Markov chain. Phase-type distributions are a fundamental building block of matrix-analytic methods, the framework for stochastic modeling that Marcel Neuts pioneered. An invariant polytope is defined as a bounded convex set with a finite number of extreme points that is mapped to itself by a given linear transformation. Phase-type distributions and invariant polytopes are closely linked, and the simple geometry of the latter gives insights into the former. Exploring this link leads to a characterization of all phase-type distributions and to insights into how the properties of a particular phase-type distribution may place restrictions on a Markov chain representation of that distribution. Marcel introduced phase-type distributions into stochastic modeling over four decades ago, and yet there are some interesting questions still awaiting answers.

Bio:
Colm O'Cinneide has worked in the investment management industry since he joined Deutsche Asset Management in 2000. He was a partner at QS Investors when it spun off in 2010, and currently is head of portfolio construction. He held faculty positions at the University of Arkansas and Purdue University from 1983 to 2000. He is currently an adjunct professor in the mathematical finance program at Columbia University, where he co-teaches a course on portfolio management. From 2009 to 2010 he served as the president of the Society of Quantitative Analysts. He holds a PhD in Statistics from the University of Kentucky.

Talk Details:

Tuesday, July 11th, 2017
12:15pm - 1:15pm
White Auditorium, 2nd Floor

Tutorials

Bert Zwart (CWI)

Title: *Selected topics on the interface of power systems and applied probability*

Abstract:
The green revolution is irreversible, and various academic studies and policies point towards a society where all energy, or at least the vast majority of all electricity, will be generated by renewable energy sources in 2050.
My vision is that applied probabilists can contribute to such a society: many outstanding problems require the development of fundamental rather than incremental research, and uncertainty will play a key role.
After giving a short introduction into the physics, economics and control of the power grid, I will give an overview of several recent research results and opportunitiies that are of interest to applied probabilists. A tentative overview is as follows:
- the physics and economics of the power grid
- traffic models for solar and wind power
- performance and control of storage devices using renewable input
- interacting energy markets and nodal pricing
- stochastic control for demand response
- electrical vehicle charging
- reliability

No background knowledge on power systems is assumed.

Bio:
Bert Zwart is a researcher at CWI, where he leads the Stochastics group. He also holds secondary positions at Eindhoven University of Technology (Professor), and Georgia Tech (Adjunct Professor). Previously, he held a Coca-Cola Chair at the H. Milton Stewart School of Industrial and Systems Engineering at the Georgia Institute of Technology. His research is in applied probability and stochastic operations research, inspired by problems in computer, communication, energy and service networks. Zwart is the 2008 recipient of the Erlang prize for outstanding contributions to applied probability by a researcher not older than 35 years old, an IBM faculty award, VENI, VIDI and VICI awards from the Dutch Science Foundation (NWO), the 2015 Van Dantzig award, and numerous best papers awards. He has co-authored more than 100 refereed publications, has been area editor of Stochastic Models for the journal *Operations Research* since 2009, and serves on several additional journal boards and TPCs.

Talk Details:

Wednesday, July 12th, 2017
10:30pm - 12:00pm
Room L - 130

Jose Blanchet (Columbia)

Title: *Optimal Transport Methods in Stochastic Operations Research and Statistics*

Abstract:
In this tutorial we will review recent results at the intersection of optimal transport, stochastic OR and statistics. After reviewing basic notions of optimal transport costs and Wasserstein distances, we will discuss distributionally robust performance analysis and optimization results. For example, we will show how the theory of diffusion approximations can be harnessed in this setting to provide model-robust sample path estimates for general stochastic systems. In addition, using the same mathematical principles, we will show how many machine learning algorithms such as dart-Lasso, regularized logistic regression, group Lasso, adaptive Lasso, support vector machines, among many others admit distrivutionally robust representations based on optimal transport costs. Finally, we also introduce model statistical methodology which can be used to optimally choose the uncertainty size in distrivutionally robust formulations.*

*This tutorial is based on work with Yang Kang and Karthyek Murray.

Bio:
Jose Blanchet is a faculty member in the departments of IEOR and Statistics at Columbia University. Jose holds a Ph.D. in Management Science and Engineering from Stanford University. Prior to joining Columbia he was a faculty member in the Statistics Department at Harvard University. Jose is a recipient of the 2009 Best Publication Award given by the INFORMS Applied Probability Society and of the 2010 Erlang Prize. He also received a PECASE award given by NSF in 2010. He worked as an analyst in Prote go Financial Advisors, a leading investment bank in Mexico. He has research interests in applied probability and Monte Carlo methods. He serves in the editorial board of Advances in Applied Probability, Journal of Applied Probability, Mathematics of Operations Research, QUESTA, Stochastic Models, and Stochastic Systems.

Talk Details:

Monday, July 10th, 2017
1:30pm - 3:00pm
Room L - 130

ABSTRACTS

Monday, 10:30 - 12:00, Room: L-130

Session: Statistical Learning

Chair: Sayan Mukherjee

Title: The (problematic) dynamics of kernel methods on large data and what to do about it

Presenter: Mikhail Belkin

Co-authors:

Abstract: Computational resources put significant limits on the type of algorithms available to machine learning for large data. To maintain feasibility, most /second order methods have to be replaced with approximations or first order gradient descent-type algorithms. While kernel methods show state-of-the-art result on smaller data, they tend to underperfrom neural networks on large datasets. In this talk I will discuss how at least part of that performance gap seems to come from the dynamics of gradient descent constrains the power of kernel methods (with smooth kernels) and how these limitations can be addressed.

Title: MCMC in large n and high dimensional problems: case studies and general principles

Presenter: James Johndrow

Co-authors: Aaron Smith (Ottawa), Natesh Pillai (Harvard), David Dunson (Duke), Paulo Orenstein (Stanford)

Abstract: As a set of procedures, Bayes seemingly has a lot to offer in modern applications. Hierarchical modeling, shrinkage, automatic interval estimates, multiplicity control, and the ability to test complex hypotheses are often touted as advantages of the Bayesian paradigm. However, even if one wants to adopt a Bayesian approach to inference, computation often makes it impractical to do so. MCMC is costly, and alternatives provide good approximations only in limited settings. In this talk, I illustrate some reasons why MCMC commonly scales poorly in big n or high dimensional (large p) settings. Through case studies and a bit of theory, I elucidate some failure modes of commonly used MCMC algorithms for regression models. In each case, I propose an alternative algorithm that scales better with n and/or p. These examples support the hypothesis that careful application of the existing MCMC toolbox combined with thorough understanding of the problem is often sufficient to obtain an algorithm with good scaling properties without resorting to exotic computational tools.

Title: Variational analysis of inference from dynamical systems

Presenter: Kevin McGoff

Co-authors: Andrew Nobel, Sayan Mukherjee

Abstract: A topological dynamical system consists of a continuous self-map $T: X \to X$ of a compact state space X. This talk considers the fitting of a parametrized family of topological dynamical systems to an observed stochastic process. The main results address the convergence of both frequentist and Bayesian inference procedures as the number of observations tends to infinity. In joint work with Andrew Nobel, we establish a general convergence theorem for minimum risk estimators and ergodic observations. Furthermore, in joint work with Sayan Mukherjee and Andrew Nobel, we use the thermodynamic formalism to establish a general convergence theorem for Gibbs posterior distributions.

Title: Generalized Probabilistic Bisection for Stochastic Root-Finding

Presenter: Mike Ludkovski

Co-authors: Sergio Rodriguez

Abstract: We consider numerical schemes for root finding of noisy responses through generalizing the Probabilistic Bisection Algorithm (PBA) to the case where the distribution of the noisy oracle is location-dependent and unknown. To do so, we employ batch querying to construct a knowledge state that is updated in a Bayesian-like fashion. We also propose several sequential sampling policies, based both on Information Gain criteria, and on Quantile Sampling using the knowledge state. Randomized policies are also investigated. Extensive numerical benchmarks are presented to illustrate the new algorithms in both synthetic and application-driven (specifically Simulation-based Optimal Stopping) contexts.

Monday, 10:30 - 12:00, Room: L-120

Session: Applied Probability in Inventory and Service Systems

Chair: Marty Reiman and Qiong Wang

Title: Queues with redundancy: Is waiting in multiple lines fair?

Presenter: Leela Nageswaran

Co-authors: Alan Scheller-Wolf (Carnegie Mellon University)

Abstract: In a service system, a redundant customer is one who may join multiple queues simultaneously and is 'served' when any one of her copies completes service; systems with redundant customers range from supermarkets with multiple checkout lines to multiple listing for organ transplants. We study the performance of two queues serving two classes of customers, one of which is redundant. We allow for two variants of redundancy, differentiated by when the redundant copies are deleted from the system: when one

of the copies completes the required service, or when one of the copies enters service. By analyzing different policies that a non-redundant customer may use to join a queue when faced with different levels of system information, our model provides fundamental insights on optimal queue-joining policies, fairness, and the value of information in such systems. Specifically, we show that joining the shortest queue (JSQ) does not necessarily minimize an arriving non-redundant customer's delay if the entire system state information (including which customers are redundant) is available, but that JSQ is optimal if only the queue lengths are observable. We also show that non-redundant customers forming independent Poisson streams to each queue prefer that the other class be redundant as opposed to utilizing JSQ if the queues are symmetric (i.e., in this case redundancy is fair); however, this may not hold if the queues have different loads.

Title: Dynamic Recommendation at Checkout under Inventory Constraint

Presenter: Will Ma

Co-authors: Xi Chen (NYU), David Simchi-Levi (MIT), Linwei Xin (UIUC)

Abstract: This work is motivated by a new checkout recommendation system at Walmart's online grocery, which offers a customer an assortment of up to 8 items that can be added to an existing order, at potentially discounted prices. We formalize this as an online assortment planning problem under limited inventory, with customer types defined by the items initially selected in the order. Multiple item prices, combined with customer withdrawal when their initially selected items stock out, pose additional challenges for the development of an effective online policy. We overcome these challenges by introducing the notion of an inventory protection level in expectation, and present an online assortment recommendation algorithm which performs well even when the arrival sequence is chosen adversarially. We further conduct numerical experiments which compare the performance of our algorithm with several existing benchmarks.

Title: Managing the Callback Option under Arrival Rate Uncertainty

Presenter: Xiaoshan Peng

Co-authors: Baris Ata (Chicago)

Abstract: We study how to manage the callback option effectively to mitigate congestion due to temporary surges in the arrivals to a call center. The call arrival process can be an arbitrary point process, allowing uncertainty and temporary surges in the arrival rate, provided that the system is stable. However, particular attention will be paid to the Poisson process with the Cox-Ingresoll-Ross (CIR) process as its stochastic intensity both in our model development and numerical results because of its practical importance, although our theoretical results hold for any arbitrary point process. When a customer arrives, the call center manager reviews the system state and decides whether to keep him in the online queue or to offer the callback option. For each customer in the online queue, she incurs a waiting cost of h per time unit. Similarly, whenever she routes a customer to an offline queue (for a callback later), she incurs a one-time penalty of p. Initially, we allow complete foresight policies that look into the entire future. We first study the case where all customers are willing to accept a callback offer. A simple lookahead policy that looks into the future for the next p/h time units is pathwise optimal among the complete foresight policies. Next, we

consider the setting where some customers may reject the callback offer. We show that a modified lookahead policy that looks into the future arrivals and service completion times for the next p/h time units and uses the current number of customers in the system who previously rejected a callback offer (but does not look into the accept/reject decisions of future customers) is pathwise optimal among the complete foresight policies. Building on the insights gleaned from the optimal lookahead policies, we also propose a non-anticipating policy, referred to as the line policy, to decide when to offer the callback option. Lastly, we conduct a simulation study using a dataset from a US bank call center which shows that the line policy has excellent performance.

Title: Asymptotic Optimality of a Control Policy for Managing Assemble-to-Order Inventory Systems

Presenter: Haohua Wan

Co-authors: Martin I. Reiman(Columbia), Qiong Wang(UIUC)

Abstract: We develop an inventory control policy for minimizing the long-run average expected cost of assemble-to-order inventory systems with a general Bill of Materials, and deterministic, but not necessarily identical, lead times. Our replenishment policy deviates from the conventional constant base stock policies to accommodate non-identical lead times. Our component allocation policy serves different demands with different levels of priority. We show that our policy is asymptotically optimal on the diffusion scale, as the lead times grow large, by proving that the expected inventory cost converges to the optimal objective value of a multi-stage stochastic program (SP), which sets a lower bound on the cost. In order to allow our control policy to imitate the optimal solution of the SP, we require that the latter solution is globally Lipschitz continuous with respect to demand inputs. This requirement can be satisfied by transforming the SP into an equivalent linear program (LP) and showing that the latter LP is Lipschitz continuous with respect to the right-hand side of constraints. While global Lipschitz continuity may not hold for general infinite dimensional LPs, we prove it is the case with our ATO inventory control problem, as our model leads to a special structure of the LP constraints.

Monday, 10:30 - 12:00, Room: 2110

Session: Large Deviations #1

Chair: Bert Zwart

Title: Sample-Path Large Deviations for Heavy-Tails: the Principle of Multiple Big Jumps

Presenter: Chang-Han Rhee

Co-authors: Jose Blanchet (Stanford), Bert Zwart (CWI)

Abstract: Many rare events in man-made networks exhibit heavy-tailed phenomena: for example, file sizes and delays in communication networks, financial losses, and magnitudes of systemic events such as the size of a blackout in a power grid. While the theory of large deviations has been wildly successful in providing systematic tools for understanding rare events in light-tailed settings, the theory developed in the heavy-tailed setting has been mostly restricted to model-specific results or results pertaining to events that are caused by a single big jump. In this talk, we present our recent result that goes beyond such restrictions and establish sample-path large deviations for a very general class of rare events associated with heavy-tailed random walks and Levy processes. We will illustrate the implications of our results in the analysis of rare events that arise in mathematical finance, actuarial science, and queueing theory.

Title: Sample path large deviations for random walks and Lvy processes with Weibullian tails

Presenter: Mihail Bazhba

Co-authors: Chang-Han Rhee (CWI), Bert Zwart (CWI), Jose Blanchet (Columbia)

Abstract: Sample path large deviations with light tailed incrementsare obtained due to the existence of the moment generatingfunction. In this paper, we consider sample path large deviationsfor a Lvy process with heavy tailed semiexponentialincrements. In contrast with the light tailed case, proving a LDP requires the development of a different framework, usingan appropriate representation, different normalization, theuse of topological properties, and a concentration inequality. Our result yields a LDP in a suitable function space, improving a result of Gantert (1998).

Title: Robust Extreme Event Analysis

Presenter: Henry Lam

Co-authors: Clementine Mottet (Boston University), Xinyu Zhang (University of Michigan)

Abstract: One recurrent issue in extreme event estimation is the limited data size in the tail region of a distribution. Conventional approaches such as extreme value theory, though mathematically justified, may encounter model misspecification issuesdue to difficulties in the simultaneous control of bias and variance. In this talk, I will present an alternate approach to compute tail quantities of interest based on optimization formulations posited over probability distributions. This approach attempts to mitigate the model misspecification issue via nonparametrically specified shape or moment constraints. I will present some structural results and numerical illustrations.

Title: How 'heavy-tailed' are the neighbours?: The role of probabilistic distance based neighborhoods in quantifying model risk

Presenter: Karthyek Murthy

Co-authors: Jose Blanchet (Stanford)

Abstract: Typical studies in distributional robustness involve computing worst-case bounds for the quantity of interest (such as expected risk, probability of default, etc.) regardless of the probability distribution used, as long as the probability distribution lies within a prescribed tolerance (measured in terms of a probabilistic divergence like KL divergence) from a suitable baseline model. With this practice of computing worst-case bounds over probabilistic distance based neighborhoods gaining popularity, we go beyond the standard choice of KL divergence to study the role of putative model uncertainty in the context of estimation of tail probabilities or quantiles. In particular, we precisely characterise 'how heavy the tails of neighboring distributions can be?'. This study seeks to understand the qualitative properties of probabilistic distance based neighborhoods in order to guide the selection of model ambiguity regions.

Monday, 10:30 - 12:00, Room: 2120

Session: Financial Risk

Chair: Sandeep Juneja - Financial Risk Session

Title: Persistence and Procyclicality in Margin Requirements

Presenter: Paul Glasserman

Co-authors: Qi Wu (Chinese University of Hong Kong)

Abstract: Margin requirements for derivative contracts serve as a buffer against the transmission of losses through the financial system by protecting one party to a contract against default by the other party. However, if margin levels are proportional to volatility, then a spike in volatility leads to potentially destabilizing margin calls in times of market stress. Risk-sensitive margin requirements are thus procyclical in the sense that they amplify shocks. We use a GARCHmodel of volatility and a combination of theoretical and empirical results to analyze how much higher margin levels need to be to avoid procyclicality while reducing counterparty credit risk. Our analysis compares the tail decay of conditional and unconditional loss distributions to compare stable and risk-sensitive margin requirements. Greater persistence and burstiness involatility leads to a slower decay in the tail of the unconditional distribution and a higher buffer needed to avoid procyclicality. The tail decay drives other measures of procyclicality as well. Our analysis points to important features of price time series that should inform 'anti-procyclicality' measures but are missing from current rules.

Title: Dynamic Portfolio Credit Risk: Calibration, Modelling and Analysis

Presenter: Sandeep Juneja

Co-authors: Anand Deo (TIFR), Aakash Kalyani

Abstract: We consider the problem of measuring credit risk, particularly portfolio credit risk, as it evolves over time. Our analysis relies on viewing the problem in discrete time in an asymptotic regime where the defaults become rarer asymptotically. We observe that this view leads to considerable simplification in typical calibration techniques that are known to be computationally demanding. In particular, we arrive at approximate and intuitively appealing closed form solution of the underlying parameters in a popular regime and observe that they perform quite well in theory as well as practice, thus obviating the need for cumbersome computational overheads. This closed form expression provides a great deal of insight into factors that influence the underlying parameters. Our key result is that beyond a point adding data from more firms for calibration provides little additional accuracy as the systemic risk cannot be averaged away by further increasing the number of firms. This risk can be reduced by increasing the time periods for which the data is available and we arrive at precise rates at which the mean square error of our estimators vanishes to zero. Further, in our asymptotic regime, we conduct large deviations analysis of large losses where the model explicitly allows contagion effect, and other dependencies.

Title: Resolution of Policy Uncertainty and Sudden Declines in Volatility

Presenter: Dacheng Xiu

Co-authors: Dante Amengual

Abstract: We introduce downward volatility jumps into a general non-affine modeling framework of the term structure of variance. With variance swaps and S&P 500 returns, we find that downward volatility jumps are associated with a resolution of policy uncertainty, mostly through statements from FOMC meetings and speeches of the Fed chairman. We also find that such jumps are priced with positive risk premia, which reflect the price of the 'put protection' offered by the Fed. Ignoring them may lead to an incorrect interpretation of such tail events. Moreover, variance risk premia tend to be insignificant or even positive at the inception of crises. On the modeling side, we explore the structural differences and relative goodness-of-fits of factor specifications, and find that the log-volatility model with two Ornstein-Uhlenbeck factors and double-sided jumps is superior in capturing volatility dynamics and pricing variance swaps, compared to the affine model prevalent in the literature or non-affine specifications without downward jumps.

Title: Stochastic Gradient Descent in Continuous Time

Presenter: Justin Sirignano

Co-authors: Konstantinos Spiliopoulos (Boston University)

Abstract: We consider stochastic gradient descent for continuous-time models. Traditional approaches for the statistical estimation of continuous-time models, such as batch optimization, can be impractical for large datasets where observations occur over a long period of time. Stochastic gradient descent provides a computationally efficient method for such statistical learning problems. The stochastic gradient descent algorithm performs an online parameter update in continuous time, with the parameter updates satisfying a stochastic differential equation. We prove that the stochastic gradient descent algorithm converges. The convergence proof leverages ergodicity by using a type of Poisson PDE to help describe the evolution of

the parameters for large times. Numerical analysis of the stochastic gradient descent algorithm is presented for several applications.

Monday, 10:30 - 12:00, Room: 2130

Session: State-Dependent Queueing Models

Chair: Ivo Adan

Title: Call center model with heterogeneous reneging customers

Presenter: Vidyadhar Kulkarni

Co-authors: Ivo Adan (TUE), Brett Hathaway (UNC-Chapel Hill)

Abstract: We consider a queueing model of a call center with reneging. There are two types of customer: patient and impatient. The patient customers renege at a slower rate than the impatient customers. We study the queueing time process in such a system and derive various performance measures in steady state.

Title: Time-dependent analysis of a multi-server priority system

Presenter: Jori Selen

Co-authors: Brian Fralix (Clemson University)

Abstract: We analyze the time-dependent behavior of an $M/M/c$ priority queue having two customer classes, class-dependent service rates, and preemptive priority between classes. More particularly, we develop a method that determines the Laplace transforms of the transition functions when the system is initially empty. The Laplace transforms corresponding to states with at least c high-priority customers are expressed explicitly in terms of the Laplace transforms corresponding to states with at most $c - 1$ high-priority customers. We then show how to compute the remaining Laplace transforms recursively, by making use of a variant of Ramaswami's formula from the theory of $M/G/1$-type Markov processes. While the primary focus of our work is on deriving Laplace transforms of transition functions, analogous results can be derived for the stationary distribution: these results seem to yield the most explicit expressions known to date.

Title: A rate balance principle and its application to state-dependent queueing models

Presenter: Binyamin Oz

Co-authors: Moshe Haviv (The Hebrew University of Jerusalem), Ivo Adan (TU Eindhoven)

Abstract: We introduce a rate balance principle for general (not necessarily Markovian) stochastic processes. One immediate implication of this principle is the following. For a processes with birth and death like transitions, for any state i, the rate of two consecutive transitions from i_1 to i+1, coincides with the corresponding rate from i+1 to i_1. We show that this observation is very useful in the analysis of state-dependent queueing models such as the Mn/Gn/1. We also demonstrate the use of this principle for models with non-birth and death like transitions. We consider two such models which are variations of the Mn/Gn/1 model. The first model is with batch arrivals and the second is with repeated server vacations.

Title: A shortest queue problem with jockeying

Presenter: David Perry

Co-authors: Ivo Adan, Rachel Ravid

Abstract: We introduce a Markov queueing system with Poisson arrivals, exponential services and jockeying between two of parallel and equivalent servers. An arriving customer admits to the shortest line (when the lines are equal the customer admits to any line with probability 1/2). Every transition, of only the last customer in line, from the longer line to the shorter line is accompanied by a certain fixed cost. Thus, a transition from the longer queue to the shorter queue occurs whenever the difference between the lines reaches a certain discrete threshold. In this study we focus on the stochastic analysis of the number of transitions of an arbitrary customer. A joint work with David Perry and Ivo Adan

Monday, 10:30 - 12:00, Room: 2410

Session: Healthcare Operations and Applied Probability

Chair: Yuan Zhong

Title: Using future information to reduce waiting times in the Emergency Department via diversion

Presenter: Kuang Xu

Co-authors: Carri Chan (Columbia)

Abstract: The development of predictive models in healthcare settings has been growing; one such area is the prediction of patient arrivals to the Emergency Department (ED). The general premise behind these works is that such models may be used to help manage an ED which consistently faces high congestion. In this work, we propose a class of proactive policies which utilizes future information of potential patient arrivals to effectively manage admissions into an ED while reducing waiting times for patients who are eventually treated. Instead of the standard strategy of waiting for queues to build before diverting patients, the proposed policy utilizes the predictions to identify when congestion is going to increase and proactively diverts patients before things get 'too bad'. We demonstrate that the proposed policy provides delay

improvements over standard policies used in practice. We also consider the impact of errors in the information provided by the predictive models and find that even with noisy predictions, our proposed policies can still outperform (achieving shorter delays while serving the same number of patients) standard diversion policies. If the quality of the predictive model is insufficient, then it is better to ignore the future information and simply rely on real-time, current information for the basis of decision making. Using simulation, we find that our proposed policy can reduce delays by up to 15%.

Title: Staffing and Scheduling of Operating Rooms via Robust Online Bin-packing

Presenter: Chaithanya Bandi

Co-authors: Diwakar Gupta (UT Austin)

Abstract: We consider two problems faced by an Operating-Room (OR) manager: (1) how to book surgery scheduling requests that arrive one by one, and (2) how many ORs to plan to staff on a regular basis. The former decisions are made continuously as booking requests arrive, whereas the latter decision is revisited periodically, e.g. quarterly or annually and it determines the cost of regular staff plus overtime. Past work on these problems have either assumed full advance knowledge of the case-length distributions of all cases that need to be scheduled on a given day, or absolutely no knowledge of future arrivals. Both assumptions are not realistic. Historical data are usually available, which may be partially informative about future case uncertainty. We use a robust optimization approach that leverages available information and does not over fit the model of future uncertainty to historical data. In particular, we show that algorithms belonging to the class of interval classification algorithms achieve the best {\em robust competitive ratio}, and develop a tractable approach to calculate the optimal parameters of our proposed algorithm. We implement and test our algorithm on data from a hospital. We also demonstrate how our approach can be extended to capture surgeon preferences and other real-life constraints.

Title: A Queueing Model for Internal Wards

Presenter: Ohad Perry

Co-authors: Jing Dong (Northwestern)

Abstract: We propose a queueing model that takes into account the most salient features of queues associated with large internal wards, including the need for a physician's approval for discharging patients, and subsequent discharge delays. We characterize the maximum long-run workload that the IW can handle, and employ a deterministic fluid approximation for the non-stationary patient-flow dynamics. The fluid model is shown to possess a unique periodic equilibrium, which is guaranteed to be approached as time increases, so that long-run performance analysis can be carried out by simply considering that equilibrium. Consequently, evaluating the effects of policy changes on system's performance, and optimizing long-run operating costs, are facilitated considerably.

Monday, 10:30 - 12:00, Room: 2420

Session: Simulation for Statistical Optimization Problems

Chair: Enlu Zhou

Title: Estimating the Probability that a Function Observed with Noise is Convex

Presenter: Shane G. Henderson

Co-authors: Nanjing Jian (Cornell)

Abstract: Consider a real-valued function that can only be observed with noise at a finite set of design points within a Euclidean space. We wish to determine whether there exists a convex function that goes through the true function values at the design points. We develop an asymptotically consistent Bayesian sequential sampling procedure that estimates the posterior probability that the function is convex given samples. In each iteration, the posterior probability is estimated using Monte Carlo simulation. We offer three variance reduction methods -- change of measure, acceptance/rejection, and conditional Monte Carlo. In numerical experiments, the conditional Monte Carlo method works best for low dimensional functions, and the acceptance rejection method works better for high dimensional functions.

Title: Learning-based robust optimization for data integration in stochastic optimization

Presenter: Zhiyuan Huang

Co-authors: Henry Lam (University of Michigan), Jeff Hong (City University of Hong Kong)

Abstract: We propose a statistical framework to integrate data into stochastic optimization. The framework is based on learning a prediction set using geometric shapes that are tractable in a robust optimization reformulation, and a validation step to achieve statistical guarantees on feasibility. We compare the proposed approach with previous sampling-based approaches in relation to the dimension of the problem.

Title: Using Simulation To Improve Statistical Power In Switchback Experiments At Uber

Presenter: Peter Frazier

Co-authors:

Abstract: We consider A/B testing of systemic changes with time-varying effects, such as changes to the algorithm used to dispatch cars at Uber. Testing such changes is made difficult by correlations in outcomes across dispatches, and by seasonal and random autocorrelated variation in riders' demand for trips. One standard A/B testing method is a switchback experiment, which applies the treatment and control on alternating days over two weeks. We show how to leverage simulation-based predictions of a change's effects to improve this method's statistical power, and make it robust to missing data.

Title: Subsampled Newton Methods for Stochastic Optimization

Presenter: RaghuÊBollapragada

Co-authors: Richard Byrd, Jorge Nocedal

Abstract: We study the solution of stochastic optimization problems in which approximations to the gradient and Hessian are obtained through subsampling. We show how to coordinate the accuracy in the gradient and Hessian to yield a superlinear rate of convergence in expectation. We also consider inexact Newton methods and investigate what is the most effective linear solver in terms of computational complexity.

Monday, 10:30 - 12:00, Room: 2430

Session: Decision Marking in Healthcare

Chair: Naveed Chehrazi

Title: Data Uncertainty in Humanitarian Aid Research

Presenter: Petra Robinson

Co-authors: Wanda Eugene (University of Florida)

Abstract: In recent years, there has been a continued climb in the demand for humanitarian aid relief as man-made disasters, health epidemics, and natural disasters impacting the lives of approximately 250 million people each year, are trending upwards. Historically, humanitarian aid relief research has been conducted from a social science perspective. These recent years have witnessed a spike in operations research-related investigations into humanitarian aid, yet many gaps still remain. There are several challenges to optimizing humanitarian aid, and chief among them is data uncertainty. Practitioners are often forced to make aid delivery decisions based on incomplete or unknown data. Here we will present a synopsis of the research in humanitarian aid relief optimization, discuss the challenges, highlighting recent efforts to address data uncertainty, and introduce future research opportunities. We conclude by summarizing the contributions our paper makes to optimization, the humanitarian aid community as well as those receiving aid now and in the future, with the goal of enriching discussions of how we can further benefit populations around the world.

Title: Dynamics of Drug Resistance: Optimal Control of an Infectious Disease

Presenter: Naveed Chehrazi

Co-authors: Lauren Cipriano (Western University), Eva Enns (University of Minnesota)

Abstract: We examine the problem of a social planner charged with developing an optimal policy for treating infected individuals within a fixed-size, closed population using a single existing drug. We assume that individuals eventually recover naturally through their own immune response. Drug treatment expedites recovery, reducing symptom burden and productivity losses, but only in individuals infected with the drug-susceptible strain of the disease. When the drug is used to treat the infected population, selection pressure increases the fraction of patients infected with the drug-resistant strain of the disease, reducing the drug "quality" (the fraction of infections that are drug-susceptible). The state space of this control problem is described by the size of the infected population and drug quality. The social planner's objective is to minimize the total discounted economic cost of the disease when drug quality is irrecoverable and the planning horizon is infinite. We show that the optimal prescription policy is a bang-bang policy with a single switching time. The action/inaction regions can be described by a single boundary that is strictly increasing when viewed as a function of drug quality. We also obtain the social planner's optimal value function in a semi-closed form and show that it is increasing and concave with respect to the size of the infected population and decreasing and concave with respect to drug quality. The optimal value function and/or its derivatives are neither C1 nor Lipschitz continuous.

Title: Patient Type Bayes-Adaptive Treatment Plans

Presenter: M. Reza Skandari

Co-authors: Steven Shechter (University of British Columbia)

Abstract: Patient heterogeneity in disease progression is prevalent in many settings. Treatment decisions that explicitly consider this heterogeneity can lower the societal cost of care and improve outcomes by providing the right care for the right patient at the right time. In this paper, we analyze the problem of designing ongoing treatment plans for a population with heterogeneity in disease progression and response to medical interventions. We create a model that learns the patient type by monitoring the patient health over time and updates a patient's treatment plan according to the gathered information. We formulate the problem as a multidimensional state-space, partially observable Markov decision process and provide structural properties of the value function, as well as the optimal policy. As a case study, we consider the optimal timing of vascular access surgery for patients with progressive chronic kidney disease, and establish policies that consider a patient's rate of disease progression in addition to the kidney health state. We provide further policy insights that sharpen existing guidelines.

Monday, 1:30 - 3:00, Room: 2120

Session: Financial Engineering

Chair: Justin Sirignano

Title: Deep Learning for Limit Order Books

Presenter: Justin Sirignano

Co-authors:

Abstract: Deep learning is used to model price movements in the limit order book. The deep neural network architecture is specifically designed to take advantage of the limit order book's structure. Several models are compared on a very large dataset. Numerical analysis of the out-of-sample performance is presented.

Title: Optimal Kernel Estimation of Spot Volatility of Stochastic Differential Equations

Presenter: Jose E. Figueroa-Lopez

Co-authors: C. Li (Purdue)

Abstract: The selections of the bandwidth and kernel function of a kernel estimator are of great importance in practice. In the context of spot volatility estimation, most of the proposed methods are either heuristic or just formally stated without any feasible implementation. In this work, an efficient method of bandwidth and kernel selection is proposed, under some mild conditions on the volatility, which not only cover classical Brownian motion driven dynamics but also some processes driven by long-memory fractional Brownian motions. Under such a unifying framework, we characterize the leading order terms of the mean squared error. Central limit theorems for the estimation error are also obtained. As a byproduct, an approximated optimal bandwidth is derived in closed form, which allows us to develop a feasible plug-in type bandwidth selection procedure, for which, as a sub-problem, we propose a new estimator of the volatility of volatility. The optimal selection of the kernel function is also considered. For volatilities driven by Brownian Motion, the optimal kernel is an exponential function, which is also shown to have desirable computational properties. Simulation studies further confirm the good performance of the proposed methods.

Title: A Dynamic Network Model of Interbank Lending

Presenter: Agostino Capponi

Co-authors: David Yao (Columbia University) and Xu Sun (Columbia University)

Abstract: We develop a dynamic model of interbank borrowing and lending activities in which banks are organized into clusters, and adjust their monetary reserve levels so as to meet prescribed capital requirements. Each bank has its own initial monetary reserve level and faces idiosyncratic risks characterized by an independent Brownian motion; whereas system wide, the banks form a hierarchical structure of clusters. We model the interbank transactional dynamics through a set of interacting measure-valued processes. Each individual process describes the intra-cluster borrowing/lending activities, and the interactions among the processes capture the inter-cluster Þnancial transactions. We establish the weak limit of the interacting measure-valued processes as the number of banks in the system grows large. We then use the limiting results to develop asymptotic approximations on two proposed macro-measures, the liquidity stress index and the concentration index, both capturing the dynamics of systemic risk. Numerical

examples are used to illustrate the applications of the asymptotics and related sensitivity analysis with respect to various indicators of Þnancial activity.

Monday, 1:30 - 3:00, Room: 2130

Session: Queueing

Chair: Yoav Kerner

Title: Studying an offloading policy for multi-resource Cloud services under Kelly's Regime

Presenter: Guilherme Thompson

Co-authors:

Abstract: We use a local state-aware policy in order to improve the performance of data centres in a distributed Cloud Computing system offering multi-resource services. Using Kelly's Scaling, we define threshold parameters of an offloading policy to enable cooperation between data centres. Thresholds are chosen to anticipate sufficiently in advance potential shortages of any resource in any data centre. Congestion-maker clients with the largest demand of an overly demanded resource are systematically forwarded to another data centre when the threshold level is reached.We express the performance of the system in terms of the invariant distribution of a Random Walk in a Markovian Enviroment. Using an approach similar to the Wiener-Hopf factorization, we obtain an explicit expression for the probability of a customer demand being redirected from a given data centre to another. Based on this result, we derive optimal threshold parameters, improving the performance of the distributed Cloud Computing system in such a way that it approaches the efficiency of a centralised system.

Title: Mean-field limits for multi-hop random-access networks

Presenter: Peter van de Ven

Co-authors: Fabio Cecchi (Eindhoven University of Technology)

Abstract: The proliferation of mesh and sensor networks ensures that random-access algorithms are increasingly used in multi-hop settings. Here packets may be forwarded along multiple intermediate nodes, and buffers occasionally empty, temporarily preventing nodes from competing for the medium. While saturated models provide a useful first-order approximation for such networks, they fail to fully capture and explain the complex behavior seen in multi-hop random-access networks. Partial results are available for certain small multi-hop networks, but the complex interactions between node activity and buffer contents makes a detailed analysis for general multi-hop networks unlikely. Motivated by the emergence of massive wireless mesh networks, and the upcoming device-to-device mode in 5G, we consider the mean-field limit of these multi-hop networks for general interference graphs. We show that the resulting ODE provides a

remarkably tractable and accurate description of these networks, even when the number of nodes is small. We discuss some examples, illustrating the strength of this approach, and the mathematical challenges that arise.

Title: Strategic behavior in an observable retrial queue

Presenter: Yoav Kerner

Co-authors: Ricky Roet-Green (Rochester)

Abstract: We consider a Markovian service system at which customer who finds the server busy upon arrival is sent to an orbit. While waiting in the orbit, each customer should decide when to inspect the server's availability. The decision should taking into account the number of customers in the orbit and their strategies. We show that the symmetric Nash equilibrium profile is mixed and obtain the sequence of continuous distributions on the positive half real line that possess the symmetric Nash equilibrium. We also obtain the performance measures (e.g., distributions of waiting time in the orbit and the number of customers in the orbit) of the system.

Title: Instantaneous Control of Brownian Motion with a Positive Lead Time

Presenter: Zhen XU

Co-authors: Jiheng Zhang (HKUST), Rachel Zhang (HKUST)

Abstract: Consider a storage system where the content is driven by a Brownian motion absent control. At any time, one may increase or decrease the content at a cost proportional to the amount of adjustment. A decrease of the content takes effect immediately, while an increase is realized after a fixed lead time l. Holding costs are incurred continuously over time and are a convex function of the content. The objective is to find a control policy that minimizes the expected present value of the total costs. Due to the positive lead time for upward adjustments, one needs to keep track of all the outstanding upward adjustments as well as the actual content at time t as there may also be downward adjustments during [t, t + l), i.e., the state of the system is a function on [0,l]. To the best of our knowledge, this is the first paper to study instantaneous control of stochastic systems in such a functional setting. We first extend the concept of L_(L-natural)-convexity to function spaces and establish the L_-convexity of the optimal cost function. We then derive various properties of the cost function and identify the structure of the optimal policy as a state-dependent two-sided reflection mapping making the minimum amount of adjustment necessary to keep the system states within a certain region.

Monday, 1:30 - 3:00, Room: 2410

Session: Healthcare Applications

Chair: Baris Ata and Cem Randa

Title: Steady-state Diffusion Approximations for Discrete-time Queue in Hospital Inpatient Flow Management

Presenter: Pengyi Shi

Co-authors: Jiekun Feng (Cornell)

Abstract: We analyze a discrete-time queue that is motivated from studying hospital inpatient flow management, where the customer count process in this queue captures the hospital midnight inpatient census. The stationary distribution of the customer count has no explicit form and is difficult to compute in certain parameter regimes. Using the Stein's method framework, we identify a continuous random variable to approximate the steady-state customer count. This continuous random variable corresponds to a diffusion process with state-dependent diffusion coefficients. We characterize the error bounds of the approximation under a variety of system load conditions and identify the critical role that the service rate plays in the convergence rate of the error bounds.

Title: An Empirical Analysis of the Effect of Kidney Allocation Policies on Patient Behavior

Presenter: A.Cem Randa

Co-authors: Baris Ata (University of Chicago)

Abstract: Organ Procurement and Transplantation Network (OPTN) allocates deceased donor kidneys to the patients based on an additive point system. The patients accumulate points for waiting time and other factors, and organs are offered to patients according to their points. The patients carry no obligation to accept any organ offers. OPTN continuously updates the weight of factors that are affecting contributing to point system, in order to have a more effective allocation policy. However, this effort excludes the patient behavior. We will develop an empirical model to capture the affect of patient behavior in this system and evaluate different counterfactual policies empirically.

Title: Robust Wait Time Estimation in General Resource Allocation Systems: an application to the Kidney Allocation System

Presenter: Chaithanya Bandi

Co-authors: Nikos Trichakis (MIT), Phebe Vayanos (USC)

Abstract: In this paper we study systems that allocate different types of scarce resources to heterogeneous allocatees based on predetermined priority rules, e.g., the U.S. deceased-donor kidney allocation system or the public housing program. We tackle the problem of estimating the wait time of an allocatee who possesses incomplete system information with regard, for example, to his relative priority, other allocatees' preferences, and resource availability. We model such systems as multiclass, multiserver queuing systems that are potentially unstable or in transient regime. We propose a novel robust

optimization solution methodology that builds on the assignment problem. For first-come, first-served systems, our approach yields a mixed-integer programming formulation. For the important case where there is a hierarchy in the resource types, we strengthen our formulation through a drastic variable reduction and also propose a highly scalable heuristic, involving only the solution of a convex optimization problem (usually a second-order cone problem). We back the heuristic with a tight approximation guarantee that becomes tighter for larger problem sizes. We illustrate the generalizability of our approach by studying systems that operate under different priority rules, such as class priority. We conduct a wide range of numerical studies, demonstrating that our approach outperforms simulation. We showcase how our methodology can be applied to assist patients in the U.S. deceased donor kidney waitlist. We calibrate our model using historical data to estimate patients' wait times based on their kidney quality preferences, blood type, location and rank in the waitlist.

Title: Advance Service Reservations with Heterogeneous Customers

Presenter: Xinshang Wang

Co-authors: Cliff Stein, Van-Anh Truong

Abstract: We study a fundamental model of resource allocation in which a finite number of resources must be assignedin an online manner to a heterogeneous stream of customers. The customers arrive randomly over timeaccording to known stochastic processes. Each customer requires a specific amount of capacity and has aspecific preference for each of the resources, with some resources being feasible for the customer and somenot. The system must find a feasible assignment of each customer to a resource or must reject the customer.The aim is to maximize the total expected capacity utilization of the resources over the horizon. This modelhas application in services, freight transportation, and online advertising. We present online algorithms withbounded competitive ratios relative to an optimal offline algorithm that knows all stochastic information.Our algorithms perform extremely well compared to two common heuristics as demonstrated on a real dataset from a large hospital system in New York City

Monday, 1:30 - 3:00, Room: 2420

Session: Risk-Aware Stochastic Optimization

Chair: Ruiwei Jiang

Title: Ambiguous Risk Constraints with Moment and Unimodality Information

Presenter: Ruiwei Jiang

Co-authors: Bowen Li (Michigan), Johanna L. Mathieu (Michigan)

Abstract: This talk discusses risk constraints based on probabilistic guarantees and conditional value-at-risk, when the probability distribution of the uncertain parameters is ambiguous. In particular, we assume that the distributional information consists of the first two moments of the uncertainty and a generalized notion of unimodality. We provide equivalent reformulations for these risk constraints based on second-order conic sets. We also demonstrate the theoretical results via a computational case study on power system operations.

Title: Asymptotics of Bayesian Risk Formulations for Data-driven Stochastic Optimization

Presenter: Di Wu

Co-authors: Enlu Zhou (Georgia Tech)

Abstract: A large class of stochastic programs involve optimizing an expectation taken with respect to some underlying distribution that is unknown in practice. When data is available, a popular way to deal with such uncertainty is via distributionally robust optimization (DRO), which usually aims to hedge against the worst case over an uncertainty set. However, despite the tractability and performance guarantee of DRO, inappropriate construction of the uncertainty set can sometimes result in over-conservative solutions. To explore the middle ground between completely ignoring the distributional uncertainty and optimizing over the worst case, we consider a Bayesian risk formulation for parametric underlying distributions, which is to optimize a risk measure takenwith respect to the posterior distribution of an unknown distribution parameter. Of our particular interest are four risk measures: mean, mean-variance, value-at-risk, and conditional value-at-risk. We show the consistency of objective functions and optimal solutions, as well as the asymptotic normality of objective functions and optimal values. Moreover, our analysis reveals the hidden intuition of risk formulation: the risk formulation can be approximately viewed as a weighted sum of posterior mean performance and the (squared) half-width of the true performance's confidence interval.

Title: Distributionally Robust Inventory Control when Demand is a Martingale

Presenter: Linwei Xin

Co-authors: David A. Goldberg (Georgia Tech)

Abstract: Independence of random demands across different periods is typically assumed in multi-period inventory models. In this talk, we consider a distributionally robust model in which the sequence of demands must take the form of a martingale with given mean and support. We explicitly compute the optimal policy and value, and shed light on the interplay between the optimal policy and worst-case martingale. We also compare to the analogous setting in which demand is independent across periods.

Title: An Adjustable Uncertainty Set Approach to Address Wind Uncertainty in Power Generation

Presenter: Feng Qiu

Co-authors: C. Wang (Tsinghua), J. Wang (ANL)

Abstract: Uncertainty sets are used in robust optimization for characterizing the ranges of randomness. Most often the uncertainty set is designed for computational convenience, rather than better serving the applications. A unfavorable consequence is that the solution can get unnecessarily conservative. In this work, we attempt to design the uncertainty set by considering the potential costs when uncertain parameters fall out of the uncertainty set. We apply this logic in robust unit commitment (RUC) to address the following two questions: 1) how much the potential operational loss could be if the realization of uncertainty is beyond the prescribed uncertainty set; 2) how large the prescribed uncertainty set should be when it is used for RUC decision making. In this regard, a robust risk-constrained unit commitment (RRUC) formulation is proposed to cope with large-scale volatile and uncertain wind generation. Differing from existing RUC formulations, the wind generation uncertainty set in RRUC is adjustable via choosing diverse levels of operational risk. By optimizing the uncertainty set, RRUC can allocate operational flexibility of power systems over spatial and temporal domains optimally, reducing operational cost in a risk-constrained manner. Moreover, since impact of wind generation realization out of the prescribed uncertainty set on operational risk is taken into account, RRUC outperforms RUC in the case of rare events. Three algorithms based on column and constraint generation (C&CG) are derived to solve the RRUC. As the proposed algorithms are quite general, they can also apply to other RUC models to improve their computational efficiency. Simulations on a modified IEEE 118-bus system demonstrate the effectiveness and efficiency of the proposed methodology.

Monday, 1:30 - 3:00, Room: 2430

Session: Sequential Learning and Exploration

Chair: Dan Russo

Title: Matching while Learning

Presenter: Vijay Kamble

Co-authors: Yash Kanoria (Columbia), Ramesh Johari (Stanford)

Abstract: We consider the problem faced by a service platform that needs to match supply with demand, but also to learn attributes of new arrivals in order to match them better in the future. We introduce a benchmark model with heterogeneous workers and jobs that arrive over time. Job types are known to the platform, but worker types are unknown and must be learned by observing match outcomes. Workers depart after performing a certain number of jobs. The payoff from a match depends on the pair of types and the goal is to maximize the steady-state rate of accumulation of payoff.Our main contribution is a complete characterization of the structure of the optimal policy in the limit that each worker performs many jobs.The platform faces a trade-off for each worker between myopically maximizing payoffs (exploitation) and learning the type of the worker (exploration). This creates a multitude of multi-armed bandit problems, one for each worker, coupled together by the constraint on availability of jobs of different types (capacity constraints). We find that the platform should estimate a shadow price for each job type, and use the

payoffs adjusted by these prices, first, to determine its learning goals and then, for each worker, (i) to balance learning with payoffs during the "exploration phase", and (ii) to myopically match after it has achieved its learning goals during the "exploitation phase."

Title: Reinforcement with fading memories

Presenter: Kuang Xu

Co-authors: Se-Young Yun (Los Alamos National Laboratory)

Abstract: We study the effect of lossy memory on decision making in the context of a sequential action-reward problem. An agent chooses a sequence of actions which generate discrete rewards at different rates. She is allowed to make new choices at rate lambda, while past rewards disappear from her memory at rate mu. We focus on a family of decision heuristics where the agent makes a new choice by randomly selecting an action with a probability approximately proportional to the amount of past rewards associated with each action in her memory. We provide closed-form formulae for the agent's steady-state choice distribution in the regime where the memory span is large, and show that the agent's success critically depends on how quickly she updates her choices relative to the speed of memory decay. If lambda >> mu, the agent almost always chooses the best action, i.e., the one with the highest reward rate. Conversely, if lambda << mu, the agent chooses an action with a probability roughly proportional to its reward rate.

Title: Thompson Sampling for the MNL-Bandit

Presenter: Vashist Avadhanula

Co-authors: Shipra Agrawal, Vineet Goyal, Assaf Zeevi (Columbia University)

Abstract: We consider a sequential subset selection problem under parameter uncertainty, where at each time step, the decision maker selects a subset of cardinality K from N possible items (arms), and observes a (bandit) feedback in the form of the index of one of the items in said subset, or none. Each item in the index set is ascribed a certain value (reward), and the feedback is governed by a Multinomial Logit (MNL) choice model whose parameters are a priori unknown. The objectiveof the decision maker is to maximize the expected cumulative rewards over a finite horizon T, or alternatively, minimize the regret relative to an oracle that knows the MNL parameters. We refer to this as the MNL-Bandit problem. This problem is representative of a larger family of exploration-exploitation problems that involve a combinatorial objective, and arise in several importantapplication domains. We present an approach to adapt Thompson Sampling to this problem and show that it achieves near-optimal regret as well as attractive numerical performance.

Title: Towards a Richer Understanding of Adaptive Sampling in the Moderate-Confidence Regime

Presenter: Kevin Jamieson

Co-authors:

Abstract: We study a structured multi-arm bandit problem in the fixed-confidence pure exploration setting. Using a novel lower bound technique for adaptive sampling, we show that constraints on the means imply a substantial gap between the moderate-confidence sample complexity, and the asymptotic sample complexity as $\delta \to 0$ found in the literature. Moreover, our lower bounds zero-in on the number of times each individual arm needs to be pulled, uncovering new phenomena which are drowned out in the aggregate sample complexity. Our new analysis inspires a simple and near-optimal algorithm for the best-arm and top-k identification, the first practical algorithm of its kind for the latter problem which removes extraneous log factors, and outperforms the state-of-the-art in experiments.

Monday, 3:30 - 5:00, Room: L-130

Session: Applied Probability and Statistical Inference

Chair: Zongming Ma

Title: Prediction under check loss in Gaussian models with unknown covariance

Presenter: Gourab Mukherjee

Co-authors:

Abstract: A host of modern business applications require prediction under asymmetric loss functions. Here, we develop new Empirical Bayes methods that can produce optimal prediction under asymmetric check losses. The check loss function is piecewise linear and penalizes underestimation and overestimation in different ways. Because of the nature of this loss, our inferential target is a pre-chosen quantile of the predictive distribution rather than the mean of the predictive distribution. Prediction here differs in fundamental respects from estimation or prediction under symmetric quadratic loss which is considered in most high-dimensional statistics literature. Under unknown covariance structure, we develop a new method for constructing efficient asymptotic risk estimates for conditionally linear predictors. Our risk estimation method uses resolvent formalism for estimating quadratic forms associated with functionals of the unknown covariance uniformly over the set of hyper-parameters. Thereafter, minimizing the risk estimates we obtain asymptotically optimal Empirical Bayes prediction rule.

Title: Kernel Additive Principal Components

Presenter: Xin Lu Tan

Co-authors:

Abstract: Additive principal components (APCs for short) are a nonlinear generalization of linear principal components. We focus on smallest APCs to describe additive nonlinear constraints that are approximately satisfied by the data. Thus APCs fit data with implicit equations that treat the variables symmetrically, as

opposed to regression analyses which fit data with explicit equations that treat the data asymmetrically by singling out a response variable. We propose a regularized data-analytic procedure for APC estimation using kernel methods. In contrast to existing approaches to APCs that are based on regularization through subspace restriction, kernel methods achieve regularization through shrinkage and therefore grant distinctive flexibility in APC estimation by allowing the use of infinite-dimensional functions spaces for searching APC transformation while retaining computational feasibility. To connect population APCs and finite-sample kernel APCs, we study population kernel APCs and their associated eigenproblems, which eventually lead to the establishment of consistency of the estimated APCs. Lastly, we discuss an iterative algorithm for computing sample kernel APCs.

Title: Inference in Ising Models

Presenter: Bhaswar B. Bhattacharya

Co-authors: Sumit Mukherjee (Columbia)

Abstract: The Ising model, the Sherrington-Kirkpatrick model of spin glasses, and the Hopfield model are all one-parameter exponential families for binary data with quadratic sufficient statistics. These have found applications in spatial modeling, social networks, image processing, neural networks, protein folding among others. Estimating the natural parameter in such models is notoriously difficult by likelihood-based methods due to the appearance of an intractable normalizing constant in the likelihood. One alternative is to use the maximum pseudolikelihood estimator (MPLE) of Besag (1975), which avoids computing the normalizing constant. We will show that the MPLE of the natural parameter is $\sqrt{a_N}$-consistent at a point whenever the log-normalizing function has order a_N in a neighborhood of that point. This gives consistency rates of the MPLE at all parameter values away from criticality, extending results of Chatterjee (2008) where only $\sqrt N$-consistency of the MPLE was shown. As a consequence, we derive sharp phase transitions in the error rate of the MPLE for Ising models on a wide class of graphs. Moreover, using the theory of graph limits (Lovsaz (2012)), we show that consistent estimation is impossible in the high temperature phase for Ising models on a converging sequence of dense graphs.

Title: Eigenvectors of Random Matrices and Applications to Inference

Presenter: Yash Deshpande

Co-authors: James Zou

Abstract: We prove a precise distributional characterization of eigenvectors in the large signal-noise ratio regime via classical techniques in random matrix theory. This generalizes results by Paul (2007), Bai and Yao (2012) and Ding (2015). We discuss applications of the results to regression correction in errors-in-variables models, estimating important features and analysis-of-variance tests.

Monday, 3:30 - 5:00, Room: L-120

Session: Inventory Models

Chair: Yonit Baron

Title: Perishability models with lead time under (S,s)

Presenter: Yonit Barron

Co-authors: Opher Baron (Rotman School of Management, Toronto)

Abstract: We consider cost minimization for an (S,s) continuous-review perishable inventory system with random lead-times, perishability time, and a state-dependent Poisson demand. Based on Queuing and Markov Chain Decomposition, we derive the stationary distributions for the inventory level under both backordering and lost sales assumptions. Numerical results and insight are provided.

Title: Inventory Systems with Lost Sales and Emergency Orders

Presenter: Sapna Isotupa

Co-authors: Michael Haughton (Wilfrid Laurier University)

Abstract: We analyze a continuous review lost sales inventory system with two types of orders (regular and emergency) for a manufacturing system based in US. The regular order is placed with a Canadian supplier. Due to security measures at border checkpoints, there is uncertainty in border crossing times. This results in risks that the order from the Canadian supplier will not arrive before inventory drops to a threshold level. This has led to interest in a dual sourcing system of placing an emergency order with a domestic supplier or with the competition when the regular order is delayed. The total costs for this system are compared to a system without emergency placement of orders. Situations under which this type of dual sourcing is effective are investigated with the help of numerical examples.

Title: Distributionally Robust Newsvendor Problems with Variation Distance}

Presenter: Tito Homem-de-Mello

Co-authors: Hamed Rahimian (Ohio State University), Guzin Bayraksan (Ohio State University)

Abstract: We use distributionally robust optimization (DRO) to model a general class of newsvendor problems where the underlying demand distribution is unknown, and so the goal is to find an order quantity that minimizes the worst-case expected cost among an ambiguity set of distributions. The ambiguity set consists of those distributions that are not far---in the sense of the so-called variation distance---from a nominal distribution. The maximum distance allowed in the ambiguity set (called level of robustness) places the DRO between the "classical" expected value and robust optimization models, which correspond to setting the level of robustness to zero and infinity, respectively. The structure of the newsvendor problem allows us to analyze the problem from multiple perspectives: First, we derive explicit formulas and

properties of the optimal solution as a function of the level of robustness. Moreover, we determine the regions of demand that are critical (in a precise sense) to optimal cost from the viewpoint of a risk-averse decision maker. Finally, we establish quantitative relationships between the distributionally robust model and the corresponding risk-neutral and classical robust optimization models, which include the price of optimism/pessimism, and the nominal/worst-case regrets, among others. Our analyses can help the decision maker better understand the role of demand uncertainty in the problem and can guide him/her to choose an appropriate level of robustness. We illustrate our results with numerical experiments on a variety of newsvendor problems with different characteristics.

Title: Optimization of time-dependent processing rates in stochastic systems

Presenter: Raik Stolletz

Co-authors: Jannik Vogel (University of Mannheim)

Abstract: A key challenge in production and service systems is to adapt the resource capacity to a time-dependent and uncertain demand. Planned changes of the processing rate based on demand forecasts are one way to cope with this.We investigate the optimal processing rates for stochastic multi-server systems. The objective function considers a reward for finished items, holding cost linear in the work-in-process, and service cost that depend on the current processing rate. We first present results for the optimal processing rate in a stationary $M/M/c$ system. Then, the time-dependent $M(t)/M(t)/c$ system is analyzed. We present a general integrated decision model for the optimization of time-dependent processing rates. In order to approximate the time-dependent performance of the system, a deterministic fluid approach and a stochastic stationary-backlog carryover (SBC) approach are developed. For the deterministic system, analytical solutions are determined. For the SBC-approach, an iterative procedure is presented that adapts the period length to the result of the optimization in order to improve the approximation quality.It is shown that under stationary conditions the optimal solution in a stochastic environment adds a kind of safety capacity to the optimal deterministic solution. For the time-dependent system, insights on the anticipation of future demand changes is found: In deterministic environments, demand changes in the future do not influence the planned processing rates. In stochastic environments, however, decisions depend on future demand changes.

Title: On the impact of treatment protocol restrictions for the un-/under-insured suffering from a chronic disease: A stylized model for compassionate dialysis

Presenter: Olga Bountali

Co-authors: Sila Cetinkaya (SMU), Vishal Ahuja (SMU)

Abstract: Un-/under-insured patients (e.g., undocumented immigrants) with a chronic condition (e.g., End Stage Renal Disease, ESRD) are not immediately eligible for regular treatment (e.g., scheduled dialysis) on a systematic basis, despite the life threatening nature of their disease. County hospitals serving the indigent are subject to district protocols under which such patients are admitted for treatment only if, during a screening process in the Emergency Room, their clinical condition is evaluated as terminally ill. This practice is known as compassionate dialysis in the case of ESRD patients who do not have access to

regular treatment. Motivated by clinical observations at a county hospital, we develop a stylized queueing model representative of the process of compassionate dialysis. We evaluate the impact of protocol restrictions on patient-centric and systems-level metrics related to treatment delays, overcrowding, and costs, and we investigate opportunities for systemic improvement and protocol policy recommendation.

Monday, 3:30 - 5:00, Room: 2110

Session: Large Deviations #2

Chair: Bert Zwart

Title: Malliavin-based Multilevel Monte Carlo estimators for density of max-stable process

Presenter: Zhipeng Liu

Co-authors: Jose Blanchet (Columbia)

Abstract: We introduce a class of unbiased Monte Carlo estimators for multivariatedensities of max-stable fields generated by Gaussian processes. Our estimatorstake advantage of recent results on the exact simulation of max-stable fieldscombined with identities studied in the Malliavin calculus literature andideas developed in the multilevel Monte Carlo literature.

Title: Efficient Rare-Event Simulation for Multiple Jump Events in Regularly Varying Random Walks and Compound Poisson Processes

Presenter: Bohan Chen

Co-authors:

Abstract: We propose a class of strongly efficient rare event simulation estimators for random walks and compound Poisson processes with a regularly varying increment/jump-size distribution in a general large deviations regime. Our estimator is based on an importance sampling strategy that hinges on the heavy-tailed sample path large deviations result recently established in Rhee, Blanchet & Zwart (2016). The new estimators are straightforward to implement and can be used to systematically evaluate the probability of a wide range of rare events with bounded relative error. They are "universal" in the sense that a single importance sampling scheme applies to a very general class of rare events that arise in heavy-tailed systems. In particular, our estimators can deal with rare events that are provoked by multiple big jumps (therefore, beyond the usual principle of single big jump) as well as multidimensional processes such as the buffer content process of a queueing network. We illustrate the versatility of our approach with several applications that arise in the context of mathematical finance, actuarial science, and queueing theory.

Title: Limit distributions related to discretized Brownian motion and Gaussian walks about random times

Presenter: Guido Lagos

Co-authors: Ton Dieker (Columbia)

Abstract: In this talk we study the simulation of barrier-hitting events andextreme events of Brownian motion, when using a discretization on anequidistant time mesh. Specifically, we study the times and positionof the discretized Brownian motion in these events and compare it tothe ones for the "real" Brownian motion. We establish new results onweak convergence of the (normalized) errors of time and position inall these cases, and give explicit analytic expressions for thelimiting distributions. In doing this we derive new results ondiffusion approximations of Gaussian random walks by Brownian motions.More importantly, our results give new insight on the connectionbetween several works in the literature dating back to the 60's wherethe constant $-\zeta(1/2)/\sqrt{2\pi}$ has appeared, where ζ isthe Riemann zeta function.

Title: Modeling Power Laws in Directed Social Networks

Presenter: Phyllis Wan

Co-authors: Tiandong Wang (Cornell), Richard A. Davis (Columbia), Sidney I. Resnick (Cornell)

Abstract: Preferential attachment is an appealing mechanism for modeling the widely observed power-law behavior of the degree distributions in directed social networks. In this talk, we consider fitting a 5-parameter linear preferential model to network data under two data scenarios. In the case where full history of the network formation is available, we derive the maximum likelihood estimators of the parameters and show that they are strongly consistent and asymptotically normal. In the case where only a single-time snapshot of the network is available, we propose an estimation method that combines method of moments with an approximation to the likelihood. The resulting estimators are also strongly consistent and performs well compared to the MLE estimator based on the full history of the network. The usage of this model is explored in a real data example.

Monday, 3:30 - 5:00, Room: 2120

Session: Approximation Methods in Financial Engineering

Chair: Lingfei Li

Title: A General Valuation Framework for SABR and Stochastic Local Volatility Models

Presenter: ZHENYU CUI

Co-authors: Justin Kirkby (Georgia Tech). Duy Nguyen (Marist College)

Abstract: In this paper, we propose a general framework for the valuation of options in stochastic local volatility (SLV) models with a general correlation structure, which includes the Stochastic Alpha Beta Rho (SABR) model as a special case. Standard stochastic volatility models, such as Heston, Hull-White, Scott, Stein-Stein, alpha-Hypergeometric, 3/2, 4/2, mean-reverting, and Jacobi stochastic volatility models, also fall within this general framework. We propose a novel double-layer continuous-time Markov chain (CTMC) approximation respectively for the variance process and the underlying asset price process. The resulting regime-switching continuous-time Markov chain is further reduced to a single CTMC on an enlarged state space. Closed-form matrix expressions for European options are derived. We also propose a recursive risk-neutral valuation technique for pricing discretely monitored path-dependent options, and use it to price Bermudan, and barrier options. In addition, we provide single Laplace transform formulae for arithmetic Asian options as well as occupation time derivatives. Numerical examples demonstrate the accuracy and efficiency of the method using several popular SLV models, and reference prices are provided for SABR, Heston-SABR, quadratic SLV, and the Jacobi model.

Title: Analysis of Markov Chain Approximation for Option Pricing and Hedging Part II: Delta and Gamma

Presenter: Gongqiu Zhang

Co-authors: Lingfei Li (The Chinese University of Hong Kong)

Abstract: We propose a simple but effective method to calculate option delta and gamma using continuous time Markov chain approximation. The delta and gamma are calculated by central difference of option prices obtained from Markov chain approximation. For general non-uniform grids, we show that in diffusion models, surprisingly, delta and gamma have the same convergence order as the option price for various types of non-smooth payoffs. We also propose a non-uniform grid that allows us to remove oscillations in the convergence and restore second order convergence for digital-type payoffs. Extensions to jump processes are discussed through numerical examples.

Title: Analysis of Markov Chain Approximation for Option Pricing and Hedging Part I: Option Price

Presenter: Lingfei Li

Co-authors: Gongqiu Zhang (The Chinese University of Hong Kong)

Abstract: Recently the method of continuous time Markov chain approximation has become popular in option pricing, however sharp convergence rates of the method for various types of payoffs are still not available. We obtain sharp convergence rates for pricing European and barrier options in diffusion models under general non-uniform grids. In particular, we show that convergence is second order for call/put-type payoffs and only first order in general for digital-type payoffs. We propose a non-uniform grid that allows us to remove oscillations in the convergence and restore second order convergence for digital type payoffs. Extensions to jump processes are discussed through numerical examples.

Title: Approximation methods in financial engineering

Presenter: Munchen Zhao

Co-authors:

Abstract: Approximation methods in financial engineering

Monday, 3:30 - 5:00, Room: 2130

Session: New Directions in Queueing Theory

Chair: Jamol Pender

Title: Queues with Delayed Information

Presenter: Jamol Pender

Co-authors:

Abstract: Delay or queue length information has the potential to influence the decision of a customer to join a queue. Therefore, it is imperative for managers of queueing systems to understand how the information that they provide will affect the performance of the system. In this talk, we will analyze two two-dimensional deterministic fluid models that incorporate customer choice behavior based on delayed queue length information. In the first fluid model, customers join each queue according to a Multinomial Logit Choice Model, however, the queue length information the customer receives is delayed by a constant amount of time which we call the delay. We show that oscillations or asynchronous behavior in the queueing model can occur based on the size of the delay. In the second model, customers receive information about the queue length through a moving average of the queue length. Although it has been shown empirically that giving patients moving average information causes oscillations and asynchronous behavior to occur in U.S. hospitals. We will also show that the moving average fluid model can exhibit oscillations and determine its dependence on the moving average window. Thus, our analysis provides new insight on how managers of queueing systems should report queue length information to customers and how delayed information can produce unwanted behavior.

Title: Static Profit Optimal Staffing of Dynamic Erlang-A Queues

Presenter: William A Massey

Co-authors:

Abstract: The Erlang A queue is a Markovian multi-server queueing model with customer abandonment. This system was inspired by call centers and has many applications in the world of healthcare operations. We can obtain a dynamical system to model the Erlang-A queueing process by using the Halfin-Whitt asymptotics of simultaneously scaling up the customer demand along with the service resource supply. The limiting results for Markovian service networks then give a deterministic fluid limit. Assuming time varying customer demand, we can create a new algorithm to find a fixed staffing size that finds profit optimality for the fluid model.

Title: A Constrained Optimization Problem for a Two-Class Queueing Model

Presenter: Mark Lewis

Co-authors: Cory Girard (Cornell University), Linda Green (Columbia), Jingui Xie (University of Science and Technology of China)

Abstract: We discuss dynamic server control in a two-class service system under a constraint on the number of high-priority customers. A class of randomized threshold policies is defined, and is proven to contain an optimal policy in the case without abandonments. The proof of optimality is then used to construct heuristic policies for the case of low-priority abandonments, which we test numerically. The model is motivated by a hospital emergency department (ED) where both urgent and non-urgent patients seek treatment. In this setting, it is crucial to assure that urgent patients are served within a specified amount of time in order to avoid adverse consequences. At the same time it is also important to minimize waiting times for non-urgent patients who may leave the system before being treated if wait times are too long.

Title: A Particle Queuing Inference Engine: Inferring Parking Occupancy from Parking Meter Data

Presenter: Robert C Hampshire

Co-authors: Dan Jordon (University of Michigan)

Abstract: The excessive search for parking, known as cruising, generates pollution and congestion. Cities are looking for policy levers to reduce the damage caused by searching for parking. However, measuring the number of searching cars is difficult and requires sensing technologies. In this paper, we develop an approach that eliminates the need for sensing technology by using parking meter payment transactions to estimate parking occupancy and the number of cars searching for parking. The estimation scheme is based on Particle Markov Chain Monte Carlo. We validate the performance of the Particle Markov Chain Monte Carlo approach using data simulated from a GI/GI/s queue. We show that the approach generates asymptotically unbiased Bayesian estimates of the parking occupancy and underlying model parameters such as arrival rates, average parking time, and the payment compliance rate. Finally, we estimate parking occupancy and cruising using parking meter data from a large scale parking experiment called SFpark. We compare the Particle Markov Chain Monte Carlo parking occupancy estimates against the ground truth measured by parking sensors. This method enables city planners and policy makers to measure the congestion and pollution caused by drivers looking for parking in their cities. The method requires using data that cities already possess, namely historical parking payment transactions.

Monday, 3:30 - 5:00, Room: 2410

Session: Applications in Healthcare and Service Operations

Chair: Pengyi Shi

Title: Inpatient Bed Over ow: An Approximate Dynamic Programming Approach

Presenter: Pengyi Shi

Co-authors: Jim Dai (Cornell University)

Abstract: When a patient waits excessively long in the emergency department before a primary inpatient bed becomes available, hospital managers may decide to overflow her to a non-primary bed though it is undesirable. To aid this overflow decision making, we model hospital inpatient flow as a multi-class, multi-pool parallel-server queueing system and formulate a discrete-time, infinite-horizon average cost Markov decision process (MDP). The MDP incorporates many realistic and important features such as the patient arrival and discharge patterns depending on the times of the day. To overcome the curse-of-dimensionality of this MDP, we resort to the approximate dynamic programming (ADP) technique. A critical part in the algorithm is to choose appropriate basis functions to approximate the relative value function. Using a novel combination of fluid control and single-pool approximation, we develop analytical forms to approximate the relative value functions at the midnight, which then guides the choice of the basis functions for different times of the day. We demonstrate, via numerical experiments in realistic hospital settings, that our proposed ADP algorithm is remarkably effective in finding good overflow policies. These ADP policies can significantly improve various system performance over some commonly used overflow strategies.

Title: The Periodic Little's Law and its Application to Emergency Department Data

Presenter: Xiaopei Zhang

Co-authors: Ward Whitt (Columbia)

Abstract: We have developed a Periodic Little's Law (PLL) in discrete time, which generalizes the sample-path version of Little's law (L= lambda W) due to Stidham (1974). In addition to requiring that limits exist for periodic averages of the arrival rate and waiting times at each time within a periodic cycle, we require that periodic averages exist for the waiting-time distribution at each time within a cycle. Under those conditions, (i) a limit exists for the periodic average number in system at each time within the periodic cycle, (ii) these limits are all periodic functions, and (iii) a relation between the limits is established, which is consistent with the time-varying Little's law. This PLL helps explain the remarkably accurate fit in our comparisons of a simulation of a stochastic model to patient flow data from an Israeli Emergency Department in Whitt and Zhang (2017). We also extend the PLL by developing a central-limit-theorem version paralleling Glynn and

Whitt (1986, 1988), which can assist in statistical analysis, as in Glynn and Whitt (1989) and Kim and Whitt (2013).

Title: Proactive Customer Service

Presenter: Kraig Delana

Co-authors: Nicos Savva (London Business School), Tolga Tezcan (London Business School)

Abstract: We examine the proactive service of customers. This serves to match demand to provider capacity in settings where customers are more flexible than provider capacity. Our research uses queueing theory (diffusion limit approximations) to quantify the performance improvement and economic theory (simultaneous move game) to identify under what conditions customers would willing to be flexible (i.e. adopt the proactive service). We show that all customers benefit by proactive service even if only a fraction of customers are flexible. Despite the substantial reductions in delays, we find customers systematically under adopt proactive service compared to socially optimal due to a positive externality.

Monday, 3:30 - 5:00, Room: 2420

Session: New Developments on the Efficiency and Accuracy of Stochastic Simulation and Optimization

Chair: Jing Dong

Title: Simulation Optimization of Multi-class Queueing Networks via Robust Optimization

Presenter: Chaithanya Bandi

Co-authors:

Abstract: We propose a robust optimization approach to analyze and optimize the expected performance of general multi-class queueing networks arising in data centers. We model uncertainty in the demand at the queueing nodes via polyhedral sets which are inspired from the limit laws of probability. We characterize the uncertainty sets by variability parameters which control the degree of conservatism of the model, and thus the level of probabilistic protection. We then go beyond the traditional robust approach and treat the variability parameters as random variables. This allows us to devise a methodology to approximate and optimize the expected behavior via averaging the worst case values over the possible realizations of the variability parameters. We illustrate our approach by finding optimal policies for fairly complex multi-class queueing networks. Our computations suggest that our methodology (a) generates optimal solutions that match the optimal solutions obtained via stochastic optimization within time that is orders of magnitude better than stochastic optimization, and provides optimal policies that consistently outperform the solutions obtained via the traditional robust optimization approach.

Title: Optimization-based Quantification of Simulation Input Uncertainty via Empirical Likelihood

Presenter: Huajie Qian

Co-authors: Henry Lam (University of Michigan)

Abstract: We study the empirical likelihood approach to construct statistically accurate confidence bounds for stochastic simulation under nonparametric input uncertainty. The approach is based on positing distributionally robust optimization problems with suitably averaged divergence constraints to provide asymptotic coverage guarantees. We present the theory giving rise to the constraints and their calibration, and demonstrate how the approach compares to existing methods such as the Bootstrap and the Delta method in terms of computational effort and stability.

Title: Unbiased Monte Carlo Computations for Optimization and Functions of Expectations

Presenter: Yanan Pei

Co-authors: Jose Blanchet (Columbia), Peter Glynn (Stanford)

Abstract: We present general principles for the design and analysis of unbiased Monte Carlo estimators for quantities such as g (E[X]), where E[X] denotes the expectation of a random variable X, and g() is a given deterministic function. Our estimators posses finite work-normalized variance under mild regularity conditions. We apply our estimators to various settings of interest, such as optimal value estimation in the context of Sample Average Approximations, and unbiased steady-state simulation of regenerative processes. Other applications include unbiased estimators for distribution quantiles, particle filters and conditional expectations.

Title: Computational Efficiency for Sub-canonical Convergence Rate Estimators

Presenter: Zeyu Zheng

Co-authors: Jose Blanchet (Columbia, Stanford), Peter W. Glynn (Stanford)

Abstract: A number of recently developed Monte Carlo algorithms give rise to estimators having either infinite variance per observation or having infinite mean for the computer time per observation. Such algorithms exhibit sub-canonical convergence rates, in the sense that they converge at slower than the 'square root' canonical rate typical of most Monte Carlo algorithms. In this talk, we extend the efficiency framework developed by Glynn and Whitt for estimators exhibiting canonical convergence rates to this sub-canonical setting. This theory also permits us to develop a new class of confidence interval procedures for such sub-canonical estimators. We then apply our theory to debiased multi-level Monte Carlo algorithms, thereby deriving exact convergence rates and asymptotically valid confidence interval procedures for several SDE (stochastic differential equations) simulation schemes. This work is joint with Jose Blanchet and Peter Glynn.

Monday, 3:30 - 5:00, Room: 2430

Session: Dynamic Learning and Decision-Making

Chair: Mohsen Bayati

Title: Discontinuous demand functions: estimation and pricing

Presenter: arnoud den boer

Co-authors: Bora Keskin (Duke)

Abstract: We consider a dynamic pricing problem with an unknown and discontinuous demand function. There is a seller who dynamically sets the price of a product over a multi-period time horizon. The expected demand for the product is a piecewise continuous and parametric function of the charged price, allowing for possibly multiple discontinuity points. The seller initially knows neither the locations of the discontinuity points nor the parameters of the demand function, but can infer them by observing stochastic demand realizations over time. We measure the seller's performance by the revenue loss relative to a clairvoyant who knows the underlying demand function with certainty. We first demonstrate that ignoring demand discontinuities in dynamic pricing can be extremely costly. Then, we construct a dynamic estimation-and-pricing policy that accounts for demand discontinuities, derive the convergence rates of discontinuity-andparameter-estimation errors under this policy, and prove that it achieves near-optimal revenue performance. We also extend our analysis to the cases of time-varying demand discontinuities and inventory constraints.

Title: Exploiting the Natural Exploration in Online Decision-Making

Presenter: Khashayar Khosravi

Co-authors: Hamsa Bastani (Stanford), Mohsen Bayati (Stanford)

Abstract: Growing availability of data has enabled practitioners to tailor decisions at the individual-level. This involves learning a model of decision outcomes conditional on individual-specific covariates (contexts). Recently, contextual bandits have been introduced as a framework to study these online and sequential decision making problems. This literature predominantly focuses on algorithms that balance an exploration-exploitation tradeoff, since greedy policies that exploit current estimates without any exploration may be sub-optimal in general. However, exploration-free greedy policies are desirable in many practical settings where experimentation may be prohibitively costly or unethical (e.g. clinical trials).In this talk we show that, for a general class of context distributions, the greedy policy benefits from a natural exploration obtained from the varying contexts and becomes asymptotically optimal under some assumptions on problem parameters. Motivated by these results, we introduce Greedy-First, a new algorithm that uses only observed contexts and rewards to determine whether to follow a greedy policy or to explore. We prove that this algorithm is asymptotically optimal without any additional assumptions. Through simulations we

demonstrate that Greedy-First successfully reduces experimentation and outperforms existing (exploration-based) algorithms.

Title: Simple Bayesian algorithms for identifying the best arm in a multi-armed bandit

Presenter: Daniel Russo

Co-authors:

Abstract: This talk considers the optimal adaptive allocation of measurement effort for identifying the best among a finite set of options or designs. An experimenter sequentially chooses designs to measure and observes noisy signals of their quality with the goal of confidently identifying the best design after a small number of measurements. Just as the multi-armed bandit problem crystallizes the tradeoff between exploration and exploitation, this "pure exploration" variant crystallizes the challenge of rapidly gathering information before committing to a final decision.I will propose several simple Bayesian algorithms for allocating measurement effort, and by characterizing fundamental asymptotic limits on the performance of any algorithm, formalize a sense in which these seemingly naive algorithms are the best possible. I will also present numerical experiments exhibiting performance surpassing competing approaches.

Tuesday, 10:30 - 12:00, Room: 2110

Session: Appointment Scheduling

Chair: Rahul Jain

Title: Steady-state analyses of the appointment scheduling problem

Presenter: Alex Kuiper

Co-authors: Michel Mandjes (University of Amsterdam), Ruben Brokkelkamp (University of Amsterdam)

Abstract: A prevalent operations management problem in healthcare concerns the generation of appointment schedules. The goal of an appointment schedule is to strike an appropriate balance between the interests of client and healthcare provider. This problem is nontrivial due to randomness in both the service times and the workload. In this talk we consider the appointment scheduling problem in steady state. Optimal steady-state solutions provide insight in the appointment scheduling problem in two ways. Firstly, it provides an upper bound on the corresponding optimal transient solution. Secondly, it can be used to determine the interarrival times of an equidistant schedule. We try to find the dependency of the steady-state solution on the coefficient of variation. The goal is to arrive at a simple equation that describes this relation. We attack the problem from three different angles, using an analytical, an empirical, and a heavy-traffic approach. Finally, we show how uncertainty in the workload (e.g., presence of no-shows and walk-ins) can be incorporated.

Title: Performance of the smallest-variance-first rule in appointment sequencing

Presenter: Madelon de Kemp

Co-authors: Michel Mandjes(University of Amsterdam), Neil Olver(VU Amsterdam)

Abstract: In appointment scheduling problems, the deterministic arrival times of patients in a single-server queue should be determined. This should be done such that there is both little idle time and little waiting time. Part of this problem is the sequencing problem, in which the order in which different patients arrive should be determined.The so-called smallest-variance-first rule, which sequences patients in order of increasing variance of their service durations, is often found to perform better than other heuristics. Under some simplifying assumptions, we will consider the performance of the smallest-variance-first rule. By comparing an upper bound on the expected waiting times under this rule with a lower bound valid for any sequence, we will able to bound the difference in performance between the smallest-variance-first rule and the optimal sequence.We will also consider the limit as the number of patients tends to infinity, when we have a finite number of groups, and the fraction of patients within each group is kept constant. We will find a limiting objective function, for which the smallest-variance-first rule is optimal.

Title: Efficient Procedures for Appointment Scheduling

Presenter: Michel Mandjes

Co-authors: Alex Kuiper (Univ. of Amsterdam), Jeroen de Mast (Univ. of Amsterdam)

Abstract: A prevalent operations management problem in healthcare concerns the generation of appointment schedules that effectively deal with variation in service times and other uncertainties. We present a powerful, yet easily implemented approach that minimizes an objective function incorporating the healthcare provider's idle times and the patients' waiting times. The procedure offers fast and accurate evaluation of schedules by approximating the service-time distribution by its phase-type counterpart. We first consider the situation with a relatively large number of patients having stochastically identical service times (stationary schedules). We give accurate closed-form approximations that either exploit the distributional form of specific phase- type distributions or explicit heavy-traffic results. We then focus on the situation with a limited number of patients, for which we develop an approach for generating optimal schedules including relevant phenomena such as no-shows and overtime. Our webtool allows healthcare providers to generate appointment schedules that significantly outperform existing approaches.

Title: Non-indexability of the Stochastic Appointment Scheduling Problem

Presenter: Rahul Jain

Co-authors: Mehdi Jafarnia (USC)

Abstract: Consider a sequence of N jobs to be scheduled, each lasting a duration Xi with CDF Fi. Job durations are independent but not identically distributed. The jobs could be, for example, patients to be seen by a physician in his office, or surgeries to be performed in an operating room. The objective is to

determine optimal sequence and appointment times such that expectation of sum of squared idle time and squared delay is minimized. We first consider the sequence to be fixed, and formulate an optimization program whose solution yields optimal appointment times. We show convexity of the objective and existence of a unique solution to the optimization program. Thus, optimal appointment times when the sequence is given can be easily determined. We then consider the more general problem where the optimal sequence or order of appointments also needs to be determined. It is conjectured that the optimal sequence is given by a 'Least Variance First' (LVF) policy. This is known to be optimal for N=2. We show that this need not be true for N>2. Furthermore, we show that this problem is not indexable, i.e., there exists no index (a map from a random variable to the reals) that yields the optimal sequence. We also provide upper and lower bounds on the optimal expected cost. It turns out that minimizing the upper bound yields the LVF policy. So it can be a good heuristic for some problems. Similar results can also be obtained for an L1 objective.

Tuesday, 10:30 - 12:00, Room: 2110

Session: Stochastic Systems

Chair: Guodong Pang

Title: Ergodic control of multiclass multi-pool networks in the Halfin-Whitt regime

Presenter: Guodong Pang

Co-authors: Ari Arapostathis (UT Austin)

Abstract: We study the scheduling and routing control of Markovian multiclass multi-pool networks under the long-run average (ergodic) cost criteria in the Halfin-Whitt regime. Two formulations are considered: (i) both queueing and idleness costs are minimized, and (ii) the queueing cost is minimized while a constraint is imposed upon the idleness of all server pools. We develop a new framework to study the associated ergodic diffusion controls and characterize the optimal solutions via the HJB equations and prove the asymptotic optimality. We have studied the recurrence properties of the controlled diffusions, by developing a leaf elimination algorithm to obtain an explicit expression for the drift. We have discovered a family of state-dependent Markov balanced saturation policies that stabilize the controlled diffusion-scaled state processes (either geometrically or subgeoemetrically stable). We also consider a class of bounded-queue bounded-state (BQBS) stable networks, in which any moment of the state is bounded by that of the queue only (for both the limiting diffusion and diffusion-scaled processes). For this class of networks, we study the ergodic control problem in which the queueing cost is minimized while a fair allocation constraint on the idleness among server pools is imposed.

Title: Sensitivity Analysis of Reflected Diffusions

Presenter: Kavita Ramanan

Co-authors: David Lipshutz (Brown University)

Abstract: Reflected diffusions arise in a variety of disciplines, including as approximations of queueing and chemical reaction networks, interacting particle systems, Leontief systems in economics, and in the study of Atlas models in math finance. We establish pathwise differentiability of a large class of obliquely reflected diffusions in convex polyhedral domains, and show how they can be used to construct estimators for expectations of functionals of reflected diffusions with respect to all their defining parameters, including the initial conditions, drift and covariance coefficients, and directions of reflection.

Title: A queueing system with on-demand servers: local stability of fluid limits

Presenter: Alexander Stolyar

Co-authors: Lam Nguyen (Lehigh University)

Abstract: A random flow of customers is served by servers (or agents) invited on-demand. Each invited agent arrives into the system after a random time and may leave the system with a fixed probability after each service completion. In addition, customers and/or agents may be impatient. We consider a feedback scheme, which controls the number of pending agent invitations, depending only on the agent and customer queue lengths and their changes. The basic objective is to minimize both customer and agent waiting times. We consider the (large-scale) system fluid limits, and study their stability at the desired equilibrium point. Using the machinery of switched linear systems and common quadratic Lyapunov functions, we derive a variety of sufficient local stability conditions. For our model, we conjecture that local stability is in fact sufficient for global stability of fluid limits; the validity of this conjecture is supported by numerical and simulation experiments. When local stability conditions do hold, simulations show good overall performance of the scheme.

Title: Diffusion Approximations for Load Balancing Mechanisms in Cloud Storage Systems

Presenter: Eric Friedlander

Co-authors: Amarjit Budhiraja (University of North Carolina at Chapel Hill)

Abstract: Analysis of large-scale communication networks (e.g. ad hoc wireless networks, cloud computing systems, server networks etc.) is of great practical interest. The massive size of such networks frequently makes direct analysis intractable. Asymptotic approximations using hydrodynamic and diffusion scaling limits provide useful methods for approaching such problems. In this talk, we explore an application of these approximation techniques to a model for load balancing in a large, cloud-based, storage system. In these types of systems, files are often coded across several servers to improve reliability and retrieval speed. We consider a network of n servers storing a set of files using the Maximum Distance Separable (MDS) code where file requests are routed using the Batch Sampling routing scheme (cf. Li, Ramamoorthy, Srikant (2016)). Specifically, each file is stored in equally sized pieces across L servers such that any k pieces can reconstruct the original file. When a request for a file is received, the dispatcher routes the job

into the k-shortest queues among the L for which the corresponding servers contain the file being requested. We establish a law of large numbers and a central limit theorem as the system becomes large (i.e. $n\to\infty$). Since the queues have infinite buffers, the state descriptors are infinite dimensional. In particular, the diffusion limit is described in terms of a Hilbert space valued stochastic differential equation driven by a cylindrical Brownian motion. The well studied Power-of-d routing scheme, also known as the supermarket model, is a special case of the model considered here. This is joint work with Amarjit Budhiraja.

Tuesday, 10:30 - 12:00, Room: 2130

Session: Advances in Financial Engineering

Chair: Agostino Capponi

Title: Intraday market making with overnight inventory costs

Presenter: Agostino Capponi

Co-authors: Tobias Adrian, Erik Vogt, Hongzhong Zhang

Abstract: We model a market making HFT that seeks to end the day flat to avoid overnight inventory costs. Although these costs only apply at the end of the day, they impact intraday price dynamics and generate a negative price-inventory relationship. The sensitivity of prices to inventory levels intensifies with time, strengthening price impact and widening bid-ask spreads. These predictions are consistent with U.S. Treasury data. A comparative statics analysis reveals that while inventory costs harm price stability, this effect is attenuated by higher trading activity. A welfare analysis shows that these costs have the greatest negative impact in inactive markets.

Title: A Stochastic Game Approach to Human-Machine Interaction Systems

Presenter: Matt Stern

Co-authors: Agostino Capponi (Columbia)

Abstract: Autonomous systems can substantially enhance human effectiveness in complex environments by handling routine or cognitively challenging operations. It is crucial, however, that the human provides incentives to the autonomous robots so to make its objectives compatible with those of the human. This guarantees that the robot's actions contribute to maximize the human's utility. This process is complicated by the fact that the robot is unable to observe the risk preferences of the human, and the latter may also change his attitude toward risk over time upon observing the effect of joint actions on the system's state. We propose a novel framework based on Partially Observable Stochastic Games (POSGs); both the human and the robot take actions so as to maximize their respective utilities, and adaptively learn the risk

preferences of the human. Our framework is broad enough to capture common characteristics of smart cities, such as autonomous car driving. We show how the theory of POSG models can be embedded into our framework, and used to model incentives. Those incentives induce the robot to act so as to ensure the highest degree of human's satisfaction.

Title: Multi-Product Production Planning: How a Mean-Variance Hedging Strategy Improves the Risk-Return Tradeoff

Presenter: Liao Wang

Co-authors: David D. Yao (Columbia University)

Abstract: We study production planning in a multi-product setting, in which demand for each product depends on multiple financial assets (commodities, market indices, etc). In addition to the production quantity decision at the beginning of the planning horizon, there is also a real-time hedging decision throughout the horizon; and we optimize both decisions jointly. With a mean-variance problem formulation, we derive the optimal hedging strategy, given the production quantities, and provide an explicit objective function by which the production quantities can be solved as a static optimization problem. This way, we are able to give a complete characterization of the mean-variance efficient frontier, and quantify the contribution of the hedging strategy by the variance reduction it achieves.

Title: Risk Sensitive Asset Management and Cascading Defaults

Presenter: John Birge

Co-authors: Lijun Bo (University of Science and Technology of China), Agostino Capponi (Columbia University)

Abstract: We consider an optimal risk-sensitive portfolio allocation problem accounting for the possibility of cascading defaults. Default events have an impact on the distress state of the surviving stocks in the portfolio. We study the recursive system of {non-Lipschitz} quasi-linear parabolic HJB-PDEs associated with the value function of the control problem in the different default states of the economy. We show the existence of a classical solution to this system via super-sub solution techniques and give an explicit characterization ofthe optimal feedback strategy in terms of the value function. {We prove a verification theorem establishing the uniqueness of the solution.} A numerical analysis indicates that the investor accounts for contagion effects when making investment decisions, reduces his risk exposure as he becomes more sensitive to risk, and that his strategy depends non-monotonically on the aggregate risk level.

Tuesday, 10:30 - 12:00, Room: 2410

Session: Applications of Queueing Theory: Communication Systems, Road Traffic

Chair: Murtuza Ali Abidini

Title: Waiting time and heavy-traffic analysis of M^X/G/1 type queuing models with dependent service durations

Presenter: Abhishek

Co-authors: Rudesindo Nez Queija (University of Amsterdam), Marko Boon (Eindhoven University of Technology)

Abstract: We consider a single-server queue with N types of services, in which, customers arrive according to a batch Poisson process. The service times of customers have general distribution functions and are correlated. The correlations are due to the different service types, which form a Markov chain that itself depends on the sequence of service lengths. In addition, the first customer in a busy period has a different service time distribution than regular customers served in the busy period. Our work is motivated by an application of the model to road traffic, where a stream of vehicles on a minor road merges with a stream on a main road at an unsignalized intersection such that the merging times of two subsequent vehicles on the minor road are dependent.Based on the results from a previous paper on the steady-state distribution of the queue length, we derive the waiting time and sojourn time distributions. The waiting times and sojourn times of customers depend on their positions in the batches, as well as on the type of service of the first customers in their batches. We also determine the heavy-traffic distribution of the scaled stationary queue length.

Title: Perturbation analysis for random time-limited polling models

Presenter: Mayank

Co-authors: Onno Boxma (Eindhoven university of technology), Stella Kapodistria (Eindhoven university of technology), Rudesindo Nunez Queija (University of Amsterdam)

Abstract: In this talk, we illustrate how perturbation analysis can be used in order to analyze the joint queue length distribution of a single server polling model, with a special service discipline. In particular, we assume a variation of the randomly timed gated service discipline: if a queue becomes empty, the server does not switch to the other queue, but only does so when an exponential timer expires. There are two advantages of this model. It enables to: i) keep the frequency of switching at a predetermined level (thus controlling the total cost, if there is a switching cost), ii) balance the time that the server spends in each queue (since, contrary to exhaustive or gated service disciplines, this discipline does not depend on the number of customers present in the various queues). This polling model violates the branching property, thus a direct analytic derivation of the joint queue length distribution turns out to be very difficult. To overcome this issue, we explore the use of (parametric) perturbation. In doing so, we have two options for the choice of the perturbed parameters: either to perturb the service and arrival rates, or to perturb the

residing time (i.e., the random time the server spends in each queue). In this talk, we discuss both cases and illustrate the advantages and disadvantages of each choice. Furthermore, we demonstrate how to compute the joint queue length distribution.

Title: Heavy traffic analysis of a polling model with retrials and glue periods

Presenter: Murtuza Ali Abidini

Co-authors: Jan-Pieter Dorsman (University of Amsterdam), Jacques Resing (Eindhoven University of Technology)

Abstract: We present a heavy traffic analysis of a single-server polling model, with the special feature of retrials. Just before the server arrives at a station there is some deterministic glue period. Customers (both new arrivals and retrials) arriving at the station during this glue period will be served during the visit of the server. Customers arriving in any other period leave immediately and will retry after an exponentially distributed time. Our main focus is on the scaled heavy traffic queue length analysis, both at embedded time points (beginnings of glue periods, visit periods and switch periods) and at arbitrary time points. We also use the analysis to approximate the mean number of customers in the system under different loads.The idea of looking at models with retrials and glue periods is motivated by the study of optical networks. Here, the retrials represent the fiber delay loops and the glue periods are introduced to model the feature of slowing down of packets using a higher refractive index in a small part of the loop. Not restricting ourselves to optical networks, one can also interpret a glue period as a reservation period, i.e., a period in which customers can make a reservation at a station for service in the subsequent visit period of that station. In our model, the reservation period immediately precedes the visit period, and could be seen as the last part of a switchover period.

Tuesday, 10:30 - 12:00, Room: 2420

Session: Matching Models

Chair: Rene Caldentey

Title: On the Optimal Design of a Bipartite Matching System

Presenter: Rene Caldentey

Co-authors: Philipp Afeche (U. Toronto), Varun Gupta (U. Chicago)

Abstract: We explore the optimal design of matching topologies for a multi-class multi-server system. Each customer class has a specific preference over server types and a joining rate that is based on its expected 'quality' of the matching and its expected delay in the system. We investigate the performance of the

system from the perspective of a central planner who imposes a fairness constraints and in addition decides the set of feasible customer-server pairs.

Title: Spatial Pricing in Ride-Sharing Networks

Presenter: Ozan Candogan

Co-authors: Kostas Bimpikis (Stanford), Daniela Saban (Stanford)

Abstract: We explore spatial price discrimination in the context of a ride-sharing platform that serves a network of locations. Riders are heterogeneous in terms of their destination preferences and their willingness-to-pay for receiving service. Drivers decide whether, when, and where to provide service so as to maximize their expected earnings, given the platform's prices. We establish that profits and consumer surplus are maximized when the demand pattern is 'balanced' across the network's locations. In addition, we show that they both increase monotonically with the balancedness of the demand pattern (as formalized by its structural properties). Furthermore, if the demand pattern is not balanced, the platform can benefit substantially from pricing trips differently depending on the location they originate from. Finally, we consider other pricing and compensation schemes that are commonly used in practice and explore their performance for the platform.

Title: A Better Model for Job Redundancy: Decoupling Server Slowdown and Job Size

Presenter: Kristen Gardner

Co-authors: Mor Harchol-Balter (Carnegie Mellon), Alan Scheller-Wolf (Carnegie Mellon)

Abstract: Recent computer systems research has proposed using redundant requests---creating multiple copies of the same job and waiting for the first copy to complete service---to reduce latency. In the past few years, queueing theorists have begun to study redundancy, first via approximations, and, more recently, via exact analysis. Unfortunately, for analytical tractability, most existing theoretical analysis has assumed a model in which the replicas of a job each experience independent runtimes (service times) at different servers. This model is unrealistic and has led to theoretical results which can be at odds with computer systems implementation results. We introduce a much more realistic model of redundancy. Our model allows us to decouple the inherent job size (X) from the server-side slowdown (S), where we track both S and X for each job. Analysis within the S&X model is, of course, much more difficult. Nevertheless, we design a policy, Redundant-to-Idle-Queue (RIQ) which is both analytically tractable within the S&X model and has provably excellent performance.

Title: Dynamic Matching in School Choice: Efficient Seat Reallocation After Late Cancellations

Presenter: Irene Lo

Co-authors: Itai Feigenbaum (CUNY), Yash Kanoria (Columbia), Jay Sethuraman (Columbia)

Abstract: In many centralized school admission systems, a significant fraction of allocated seats are later vacated, often due to students obtaining better outside options. We consider the problem of reassigning these seats in a fair and efficient manner while also minimizing the movement of students between schools. Centralized admissions are typically conducted using the Deferred Acceptance (DA) algorithm, with a lottery used to break ties caused by indifferences in school priorities. We introduce the Permuted Lottery Deferred Acceptance (PLDA) mechanisms, which reassign vacated seats using a second round of Deferred Acceptance with a lottery given by a suitable permutation of the first round lottery numbers. We show that a mechanism based on a simple reversal of the first round lottery order performs the best among all PLDA mechanisms. We also characterize PLDA mechanisms as the class of truthful mechanisms satisfying some natural efficiency and fairness properties. Empirical investigations based on data from the NYC high school admissions system support our theoretical findings.

Tuesday, 10:30 - 12:00, Room: 2430

Session: Queueing Control Models and Applications

Chair: Amy Ward

Title: On the capacity of information processing systems

Presenter: Kuang Xu

Co-authors: Laurent Massoulie (Microsoft Research-Inria)

Abstract: We propose and analyze a family of information processing systems, where a finite set of experts or servers are employed to extract information about a stream of incoming jobs. Each job is associated with a hidden label drawn from some prior distribution. An inspection by an expert produces a noisy outcome that depends both on the job's hidden label and the type of the expert, and occupies the expert for a finite time duration. A decision maker's task is to dynamically assign inspections so that the resulting outcomes can be used to accurately recover the labels of all jobs, while keeping the system stable. Among our chief motivations are applications in crowd-sourcing, diagnostics, and experiment designs, where one wishes to efficiently learn the nature of a large number of items, using a finite pool of computational resources or human agents. We focus on the capacity of such an information processing system. Given a level of accuracy guarantee, we ask how many experts are needed in order to stabilize the system, and through what inspection architecture. Our main result provides an adaptive inspection policy that is asymptotically optimal in the following sense: the ratio between the required number of experts under our policy and the theoretical optimal converges to one, as the probability of error in label recovery tends to zero.

Title: Criticality and Adaptivity in Enzymatic Networks

Presenter: Ruth Williams

Co-authors: Paul J. Steiner (UCSD), Jeff Hasty (UCSD) and Lev S. Tsimring (UCSD)

Abstract: The contrast between stochasticity of biochemical networks and regularity of cellular behavior suggests that biological networks generate robust behavior from noisy constituents. Identifying the mechanisms that confer this ability on biological networks is essential to understanding cells. Here we use stochastic queueing models to investigate one potential mechanism. We show that queueing for a limited shared resource in enzymatic networks can produce strong and long-ranged correlations between molecular species when these systems are poised near a critical state where the substrate input flux is equal to the maximum processing capacity. We then consider enzymatic networks with adaptation, where the limiting resource is produced in proportion to the demand for it. In this setting, we show that strong correlations are robustly produced across a broad range of system parameters. This adaptive queueing motif suggests a natural control mechanism for producing strong correlations in biological systems.

Title: Asymptotically Optimal Policies for Many Server Queues with Reneging

Presenter: Amber Puha

Co-authors: Amy Ward (University of Southern California)

Abstract: This work in progress aims to determine asymptotically optimal policies for many server queues with general reneging distributions. We model multiclass many server queues with reneging using the framework developed for single-class queues by Kapsi, Kang, and Ramaman in a series of three papers and extended to multiclass many server queues by Atar, Kaspi, and Shimkin. These four papers identify fluid models, prove fluid limit theorems, and discussion stationary behavior for these many server queue models. The multiclass paper ultimately considers general interarrival and service time distributions and proves asymptotic optimality in the fluid limit of static priority in the case of exponential reneging distributions. They also derive the well-known c-mu-theta rule for prioritizing the classes in this setting.We continue along these lines, but incorporate non-exponential reneging distributions. For reneging distributions with bounded, nonincreasing hazard rates, we find that static priority is not necessarily asymptotically optimal. We identify a new class of policies, which we are calling Random Buffer Selection and prove that these are asymptotically optimal in this setting. We further identify the fluid approximation to the limiting cost as the optimal value of a certain optimization problem.

Title: Scheduling Impatient Customers

Presenter: Chenguang Allen Wu

Co-authors: Achal Bassamboo (Northwestern), Ohad Perry (Northwestern)

Abstract: The valuation for service of an arriving customer often depends on his individual service requirement; in this work we consider a queueing model in which these two random variables are stochastically dependent. Specifically, customers are price and delay sensitive, and decide whether to queue for service based on their service valuations, waiting cost and the price of service. Employing a

general dependence order, we show that the provider's optimal revenue decreases with the strength of the dependence. Moreover, considering the valuation and service requirement to be independent when they are in fact dependent can lead to substantial revenue losses.

Tuesday, 10:30 - 12:00, Room: L-130

Session: Theoretical Advances for Estimation in High-Dimensional Data

Chair: Christian B. Hansen

Title: A lava attack on the recovery of sums of dense and sparse signals

Presenter: Yuan Liao

Co-authors: Victor Chernozhukov (MIT), Christian Hansen (UChicago)

Abstract: Common high-dimensional methods for prediction rely on having either a sparse signal model, a model in which most parameters are zero and there are a small number of non-zero parameters that are large in magnitude, or a dense signal model, a model with no large parameters and very many small non-zero parameters. We consider a generalization of these two basic models, termed here a "sparse+dense" model, in which the signal is given by the sum of a sparse signal and a dense signal. Such a structure poses problems for traditional sparse estimators, such as the lasso, and for traditional dense estimation methods, such as ridge estimation. We propose a new penalization-based method, called lava, which is computationally efficient. With suitable choices of penalty parameters, the proposed method strictly dominates both lasso and ridge. We derive analytic expressions for the finite-sample risk function of the lava estimator in the Gaussian sequence model. We also provide an deviation bound for the prediction risk in the Gaussian regression model with fixed design. In both cases, we provide Stein's unbiased estimator for lava's prediction risk. A simulation example compares the performance of lava to lasso, ridge, and elastic net in a regression example using feasible, data-dependent penalty parameters and illustrates lava's improved performance relative to these benchmarks.

Title: Efficient Policy Learning

Presenter: Stefan Wager

Co-authors: Susan Athey (Stanford)

Abstract: There has been considerable interest across several fields in methods that reduce the problem of learning good treatment assignment policies to the problem of accurate policy evaluation. Given a class of candidate policies, these methods first effectively evaluate each policy individually, and then learn a policy by optimizing the estimated value function; such approaches are guaranteed to be risk-consistent whenever the policy value estimates are uniformly consistent. However, despite the wealth of proposed

methods, the literature remains largely silent on questions of statistical efficiency: there are only limited results characterizing which policy evaluation strategies lead to better learned policies than others, or what the optimal policy evaluation strategies are. In this paper, we build on classical results in semiparametric efficiency theory to develop quasi-optimal methods for policy learning; in particular, we propose a class of policy value estimators that, when optimized, yield regret bounds for the learned policy that scale with the semiparametric efficient variance for policy evaluation. On a practical level, our result suggests new methods for policy learning motivated by semiparametric efficiency theory.

Title: Discretizing Unobserved Heterogeneity

Presenter: Elena Manresa/Stephane Bonhomme

Co-authors: Stephane Bonhomme (University of Chicago) and Thibaut Lamadon (University of Chicago)

Abstract: We study panel data estimators based on a discretization of unobserved heterogeneity, when individual heterogeneity is not necessarily discrete in the population. We focus on two-step grouped-fixed effects estimators, where individuals are classified into groups based on moments using kmeans clustering in a first step, and the model is estimated allowing for group-specific heterogeneity in a second step. Discrete estimation is used as an approximation to the heterogeneity, and we analyze its properties as the number of groups grows with the sample size. We show bias reduction methods can improve the performance of discrete estimators. When allowing for time-varying unobserved heterogeneity, discrete estimators enjoy fast rates of convergence provided the underlying dimension of heterogeneity is not too large. We study two applications: a structural dynamic discrete choice model of migration, and a model of wages with worker and firm heterogeneity. These applications to settings with continuous heterogeneity suggest computational and statistical advantages of two-step grouped fixed-effects methods.

Tuesday, 10:30 - 12:00, Room: L-120

Session: Applications of Sequential Decisions

Chair: Hamed Valizaadeh Haghi

Title: Dynamic Inventory Control With Stockout Substitution And Demand Learning

Presenter: Beryl Chen

Co-authors: Xiuli Chao (University of Michigan, Ann Arbor)

Abstract: Stock-out substitution is the phenomenon that if the primary choice of a customer is out of stock, besides leaving the market immediately, the customer may also substitute for other products. In this paper, we study a data-driven inventory management problem and infer the customer substitution behavior from historical sales data.

Title: Sterrett procedure for the generalized group testing problem

Presenter: Yaakov Malinovsky

Co-authors:

Abstract: Group testing is a useful method that has broad applications in medicine, engineering, and even in airport security control. Consider a finite population of N items, where item i has a probability p_i to be defective. The goal is to identify all items by means of group testing. This is the generalized group testing problem. The optimum procedure, with respect to the expected total number of tests, is unknown even in case when all pi are equal. Hwang (1975) proved that an ordered partition (with respect to p_i) is the optimal for the Dorfman procedure (procedure D), and obtained an optimum solution (i.e., found an optimal partition) by dynamic programming. In this work, we investigate the Sterrett procedure (procedure S). We provide close form expression for the expected total number of tests, which allows us to find the optimum arrangement of the items in the particular group. We also show that anordered partition is not optimal for the procedure S or even for a slightly modified Dorfman procedure (procedure D_). This discovery implies that finding an optimal procedure S is a hard computational problem. However, by using an optimal ordered partition for all procedures, we show that procedure D_is uniformly better than procedure D, and based on numerical comparisons, procedure S is uniformly and significantly better than procedures D and D_.

Title: Real-Time Dynamic Pricing for Revenue Management with Reusable Resources and Deterministic Service Time Requirements

Presenter: Yanzhe (Murray) Lei

Co-authors: Stefanus Jasin (University of Michigan)

Abstract: We consider the setting of a firm that manages a finite amount of resources to serve price-sensitive customers who arrive randomly over time according to a specified non-stationary rate. Each customer requires a service that consumes one unit of resource for a deterministic amount of time. The resource is reusable as it can be immediately used to serve a new customer upon the completion of the previous service. The firm's objective is to set the price dynamically to maximize its expected total revenues. This is a fundamental problem faced by many firms in many industries. We model this problem using an optimal stochastic control formulation and develop two heuristic controls based on the solution of the deterministic relaxation of the original stochastic problem. The first heuristic control is static since the corresponding price sequence is determined before the selling horizon starts; the second heuristic control is dynamic, it uses the first heuristic control as its baseline control and adaptively adjusts the price based on previous demand realizations. We show that both heuristic controls are asymptotically optimal in the regime with large demand and supply. We further generalize both of the heuristic controls to the setting with multiple service types requiring different service times and advance reservation.

Title: Computational Inference and Probability Functions for Adaptive Power Networks

Presenter: Hamed Valizadeh Haghi

Co-authors:

Abstract: Distributed integration of renewable energy and demand response adds complexity to the control and optimization of smart power grid. Forecasts are essential due to the existence of stochastic variations and uncertainty. Smart grid operations must take account of, and in fact benefit from the spatio-temporal dependence among distributed energy resources. This is particularly important considering the potential buffering effects of energy storage devices such as batteries, heating/cooling systems and electric vehicles. This research develops prescriptive data analytics which enable electric utilities to become more proactive in decision-making by adjusting their strategies in real-time based on predictive views. In particular, spatio-temporal modeling in view of non-Gaussian and high-dimensional data sets will be addressed. Also, stochastic forecasting and real-time control of wind and solar power will be demonstrated.

Tuesday, 1:30 - 3:00, Room: 2110

Session: Stochastic Modelling of the Bitcoin Blockchain

Chair: Peter Taylor

Title: Selfish Mining and the Bitcoin Blockchain

Presenter: Peter Taylor

Co-authors: Johannes Gobel (University of Hamburg), Paul Keeler (Weierstrass Institute), Tony Krzesinski (University of Stellenbosch)

Abstract: In the context of the 'selfish-mine' strategy proposed by Eyal andSirer, we study the effect of propagation delay on the evolution of theBitcoin blockchain. Using a simplified Markov model that tracks the contrastingstates of belief about the blockchain of a small pool of miners andthe 'rest of the community' we establish that the use of block-hidingstrategies, such as selfish-mine, causes the rate of production oforphan blocks to increase.

Title: A Bitcoin-inspired infinite-server model with interacting customers

Presenter: Maria Frolkova

Co-authors: Michel Mandjes

Abstract: Motivated by certain synchronization processes in the Bitcoin network, we develop an infinite-server model where, unusually for these types of models, customers do interact. Among the closed-form characteristics that we derive for this model is the busy period distribution which, counterintuitively, does not depend on the arrival rate. We explain this using the equivalence of two service disciplines, which we also use to derive the model's stationary distribution. Another type of result we obtain is asymptotic: a fluid limit

in the presence of service delays. Remarkably, the fluid limit is itself a stochastic process (of growth collapse type). In addition, our methodology here provides a non-Markovian counterpart to the fluid limit analysis of window-based congestion control schemes by Dumas et al. (2002) and Guillemin et al. (2004).

Title: The Bitcoin block arrival process

Presenter: Rhys Bowden

Co-authors: Peter Taylor (University of Melbourne), Tony Krzesinski (Stellenbosch University), Paul Keeler (Weierstrass Institute)

Abstract: Bitcoin is the new "cash of the Internet". A system based on a peer-to-peer network with no central control, Bitcoin nevertheless maintains a single distributed global ledger called the blockchain. As part of this process, miners and their computers run independent Bernoulli tests at a huge rate (around 3*10^16 per second), each with the same very low probability of success. Each time a test succeeds a new block is added to the end of the blockchain. This means block arrivals should be very well approximated by a Poisson process. The total computer power dedicated to mining is constantly changing leading to a inhomogeneous Poisson process where the underlying rate is known only through the number of arriving blocks. Further, the probability of success is changed every 2016 blocks in an attempt to keep the block arrival rate constant. This, coupled with other features like propagation delay of blocks and only having one realisation makes modelling and fitting the block arrival process a challenge. We show that simple models of the system are insufficient to reproduce important phenomena, but we can form a sequence of increasingly accurate approximations.

Title: Increased block size and Bitcoin blockchain dynamics

Presenter: Anthony Krzesinski

Co-authors: J. Goebel, Department of Informatics, University of Hamburg, Hamburg, Germany

Abstract: This talk investigates whether larger block sizes can achieve highertransaction processing rates in Bitcoin. We first present an overviewof the Bitcoin protocol and we discuss various proposals to increase thebitcoin transaction processing rate. Simulation experiments indicatethat larger block sizes will not provide VISA-class transactionthroughput rates. We next present a simulation analysis of Bitcoin-NextGeneration where blocks (macroblocks) stripped of transactions propagaterapidly through the peer-to-peer network. Once a macroblock is mined,only the miner of the macroblock is entitled to broadcast transactionswhich are gathered into small microblocks which are broadcast on averageevery 10 seconds. Initial simulation experiments show that Bitcoin-NGcan sustain substantially larger transaction rates than Bitcoin classic.

Tuesday, 1:30 - 3:00, Room: 2110

Session: Applications of Mean Field Games

Chair: Baris Ata and Nasser Barjestesh

Title: Ride-Hailing Networks with Strategic Drivers: The Impact of Platform Control Capabilities on Performance

Presenter: Philipp Afche

Co-authors: Zhe Liu (Columbia), Costis Maglaras (Columbia)

Abstract: This work is motivated by the emergence of ride-hailing platforms such as Uber, Lyft and Gett that match demand (passengers) with service capacity (drivers) over a geographically dispersed network. This matching problem is complicated by two challenges. (i) There are significant demand imbalances in the network. (ii) Drivers are self-interested and behave strategically in deciding whether to join, and if so, how to reposition (route) themselves when not transporting passengers.To address these challenges we study the value of two operational controls, demand-side admission control and supply-side repositioning control, on the performance of a revenue-maximizing ride-hailing platform. Considering a fluid model of a two-location network in a game-theoretic framework, we characterize the system equilibrium under three operating regimes, ranging from minimal control to centralized admission and repositioning control. These results contribute novel insights on the interplay between the platform's admission control and the drivers' strategic routing decisions.We also quantify the impact of control capabilities on the platform revenue, the capacity and the per-driver profits. The value of control is largest at moderate utilization and increases with demand imbalances.

Title: Dynamic Pricing in Ridesharing Platforms

Presenter: Ramesh Johari

Co-authors: S. Banerjee (Cornell), C. Riquelme (Stanford)

Abstract: We study optimal pricing strategies for ride-sharing platforms, such as Lyft, Sidecar, and Uber. Analysis of pricing in such settings is complex: On one hand these platforms are two-sided -- this requires economic models that capture the incentives of both drivers and passengers. On the other hand, these platforms support high temporal-resolution for data collection and pricing -- this requires stochastic models that capture the dynamics of drivers and passengers in the system.In this paper we build a queueing-theoretic economic model to study optimal platform pricing. In particular, we focus our attention on the value of dynamic pricing: where prices can react to instantaneous imbalances between available supply and incoming demand. We find two main results: We first show that performance (throughput and revenue) under any dynamic pricing strategy cannot exceed that under the optimal static pricing policy (i.e., one which is agnostic of stochastic fluctuations in the system load). This result belies the prevalence of dynamic pricing in practice. Our second result explains the apparent paradox: we show that dynamic pricing is much more robust to fluctuations in system parameters compared to static pricing. Thus dynamic pricing does not

necessarily yield higher performance than static pricing -- however, it lets platforms realize the benefits of optimal static pricing, even with imperfect knowledge of system parameters.

Title: Pricing of Rides in Ride-sharing Platforms Based on Pickup Location

Presenter: Nasser Barjesteh

Co-authors: Baris Ata (University of Chicago), Sunil Kumar (Johns Hopkins University)

Abstract: We study the pricing of rides in ride-sharing platforms based on their pickup location. The platform operates a two-sided market. We investigate the effect of geography, distribution of demand across the city, and the distribution of the destinations of the customers on the prices. We assume the system is in steady-state and look for the mean-field equilibrium of the system. Assuming the platform sets the prices and collects a fixed percentage of the profits, we characterize the pricing scheme that maximizes the profit of the platform.

Title: Optimal Signaling Mechanisms in Unobservable Queues

Presenter: Krishnamurthy Iyer

Co-authors: David Lingenbrink (Cornell)

Abstract: We consider the problem of optimal information sharing in the context of a service system. In particular, we consider an unobservable single server queue offering a service at a fixed price to a Poisson arrival of delay-sensitive customers. The service provider can observe the queue, and may share information about the state of the queue with each arriving customer. The customers are Bayesian and strategic, and incorporate any information provided by the service provider into their beliefs about the queue size before making the decision whether to join the queue or leave without obtaining service. We pose the following question: which signaling mechanism should the service provider adopt to maximize her revenue? We establish that, in general, the optimal signaling mechanism requires the service provider to strategically conceal information from the customers to incentivize them to join. In particular, under mild technical conditions, we show that a signaling mechanism with binary signals and a threshold structure is optimal. Furthermore, for the case of linear waiting costs, we obtain analytical expressions for the thresholds of the optimal signaling mechanism. Finally, we compare the revenue of the optimal signaling mechanism to that of the optimal state-dependent pricing, and observe that setting an optimal single price and using the optimal signaling mechanism can achieve the revenue of the latter. Our work contributes to the literature on Bayesian persuasion in stochastic settings, and provides many interesting directions for extensions.

Tuesday, 1:30 - 3:00, Room: 2130

Session: Applications of Hawkes Processes, Fractional Brownian Motion, and Levy Processes in Financial Engineering

Chair: Matt Lorig

Title: Limit theorems for Markovian Hawkes processes with a large initial intensity

Presenter: Xuefeng Gao

Co-authors: Lingjiong Zhu (Florida State University)

Abstract: Hawkes process is a class of simple point processes that is self-exciting and has clustering effect. The intensity of this point process depends on its entire past history. It has wide applications in finance, social networks, and many other fields. In this paper, we study the linear Hawkes process with an exponential kernel in the asymptotic regime where the initial intensity of the Hawkes process is large. We derive limit theorems for this asymptotic regime as well as the regime when both the initial intensity and the time are large. The limit theorems could be useful for approximating the transient behavior of Hawkes processes.

Title: Short-Time Asymptotics for Options on Leveraged ETFs under Exponential Levy Models with Local Volatility

Presenter: Ruoting Gong

Co-authors: Jose E. Figueroa-Lopez (Washington University at St. Louis), Matthew Lorig (University of Washington)

Abstract: In this talk, we consider the small-time asymptotics of options on a Leveraged Exchange-Traded Fund (LETF) when the underlying Exchange Traded Fund (ETF) exhibits both local volatility and Levy jumps of either finite or infinite activity. We show that leverage modifies the drift, volatility, jump intensity, and jump distribution of an LETF in addition to inducing the possibility of default, even when the underlying ETF price remains strictly positive. Our main results are closed-form expressions for the leading order terms of off-the-money European call and put LETF option prices, near expiration, with explicit error bounds. These results show that the price of an out-of-the-money European call on a LETF with positive (negative) leverage is asymptotically equivalent, in short-time, to the price of an out-of-the-money European call (put) on the underlying ETF, but with modified spot and strike prices. Similar relationships hold for other off-the-money European options. These observations, in turn, suggest a method to hedge off-the-money LETF options near expiration using options on the underlying ETF. Finally, we derive a second-order expansion for the implied volatility of an off-the-money LETF option and show both analytically and numerically how this is affected by leverage.

Title: On small time asymptotics for rough differential equations driven by fractional Brownian motions

Presenter: Cheng Ouyang

Co-authors: Fabrice Baudoin (U Conn), Xuejing Zhang

Abstract: In this talk, we survey some results on the small time asymptotics of the density function of the solutions to such SDEs, such as Varadhan asymptotics and full expansion of the density function.

Title: Hedging in Fractional Stochastic Volatility Models

Presenter: Alexandra Chronopoulou

Co-authors:

Abstract: Long memory stochastic volatility (LMSV) models have been used to explain the persistence of volatility in the market, while rough stochastic volatility (RSV) models have been shown to reproduce statistical properties of high frequency financial data. In these two classes of models, the volatility process is often described by a fractional Ornstein-Uhlenbeck process with Hurst index H, where H>1/2 for LMSV models and H<1/2 for RSV models. The goal of this talk is to discuss hedging in the fractional stochastic volatility framework, where following the approach by Renault and Touzi, we study the variation of the strike to the underlying asset and we determine when the option is underhedged, overhedged or perfectly hedged. We also discuss filtering methods for the volatility and we apply our results to pricing and hedging options written on the S&P 500.

Tuesday, 1:30 - 3:00, Room: 2410

Session: Statistics and Queues

Chair: Alex Goldenshlunger

Title: A multiclass M/M/1 queueing problem with model uncertainty

Presenter: Asaf Cohen

Co-authors:

Abstract: We consider a multiclass M/M/1 queueing problem under heavy-traffic with model uncertainty. Namely, it is assumed that the decision maker is ambiguous about the rates of arrivals to the system and the rates of service and acts to optimize an overall cost that accounts for this uncertainty. We present a stochastic differential game that governs the heavy-traffic limit behavior. Moreover, we provide an asymptotic optimal control policy that uses a c-mu priority rule derived from the game. Finally, the asymptotic optimal policy and cost are shown to depend on the ambiguity parameters through the unique solutions to some free-boundary problems associated with the game.

Title: Statistical inference for the M/G/infinity queue

Presenter: Alexander Goldenshluger

Co-authors: None

Abstract: The subject of this talk is the problem of estimating the service time distribution in the M/G/infinity queue. We will discuss three different observation schemes with incomplete data on the queue: observations of arrivals and departures without identification of customers,observations of the superposed arrival-departure point process and observations of the queue-length (number-of-busy-servers) process. In these settings we derive some probabilistic results on the processes involved and construct estimators of the service time distribution with provable accuracy guarantees. The problems of estimating the service time expectation and the arrival rate are discussed as well. We will present also some results on comparison of different estimators.

Title: model selection in data-driven queueing optimization problems

Presenter: arnoud den boer

Co-authors: dirk sierag (cwi)

Abstract: In optimization problems, simple mathematical models that discard important factors may sometimes be preferred to more realistic models. This may occur if the parameters of the simple model are easier to estimate than the parameters of the complex model, or if the optimization problem corresponding to the simple model can be solved exactly whereas the optimization problem corresponding to the 'realistic model' is intractable. This trade-off between three sources of errors (modelling, estimation, and optimization errors) is encountered in many stochastic optimization problems. The question we address is: how can one determine if it is better to use a simplified model, rather than a more realistic model? In other words: given a data set and a particular optimization problem, how do we know whether the model-misspecification error induced by a simple model is dominated by estimation and optimization errors corresponding to more realistic models?In this research we propose a generic decision-based model selection method that determines when simplicity is preferred to realism. We explain the theoretical framework of our method, and illustrate the potential performance improvement in queueing optimization problems.

Title: Some Inference Problems in Queues

Presenter: Assaf Zeevi

Co-authors:

Abstract: In this talk we survey some estimation problems that arise in service systems, illustrated in the context of simple stylized queueing models. The problems will primarily focus on inferring system operational statistics from partial observations, namely, when information pertaining to the desired statistics is only available in indirect, incomplete, or corrupted form. In particular, we will discuss how various levels of observability impact estimation accuracy and more broadly how partial information impacts resource allocation decisions.

Tuesday, 1:30 - 3:00, Room: 2420

Session: Queues in the Sharing Economy

Chair: Rouba Ibrahim

Title: Flexible Workers or Full-Time Employees? On Staffing Systems with a Blended Workforce

Presenter: Rouba Ibrahim

Co-authors: Jing Dong (Northwestern)

Abstract: The rise of the blended workforce in the gig economy is prompting companies to reevaluate their staffing models. We study the optimal staffing of service systems hiring a blend of independent contractors and full-time employees, and characterize the trade-offs between supply uncertainty, the quality of service, and operating costs.

Title: Optimizing Shared-Vehicle Systems

Presenter: Sid Banerjee

Co-authors: Daniel Freund (Cornell), Thodoris Lykouris (Cornell)

Abstract: Pricing in shared vehicle systems is challenging due to complex network externalities: altering prices in any location affects future supply throughout the system within very short timescales. Such externalities are well captured by steady-state Markov chain models. However, optimizing over these models is computationally difficult as the resulting problems are high-dimensional and non-convex. We develop a framework for designing pricing policies in such systems, based on a novel convex relaxation which we term elevated flow relaxation, coupled with a new infinite-projection and pullback technique for proving approximation bounds. Our approach gives the first efficient algorithms with rigorous approximation guarantees and also extends beyond pricing to other demand-supply balancing controls used in shared vehicle systems. For each of these, we obtain efficient algorithms with the first finite-system approximation guarantees and recover recently-discovered asymptotic optimality results.

Title: Dynamic Matching for Real-time Ridesharing

Presenter: Erhun Ozkan

Co-authors: Amy R. Ward (University of Southern California)

Abstract: In a ridesharing system such as Uber or Lyft, arriving customers must be matched with available drivers. These decisions affect the overall number of customers matched, because they impact whether or not future available drivers will be close to the locations of arriving customers. A common policy used in practice is the closest driver (CD) policy that offers an arriving customer the closest driver. This is an

attractive policy because no parameter information is required. However, we expect that a parameter-based policy can achieve better performance.We propose to base the matching decisions on the solution to a continuous linear program (CLP) that accounts for (i) the differing arrival rates of drivers and customers in different areas of the city, (ii) how long customers are willing to wait for driver pick-up, and (iii) the time-varying nature of all the aforementioned parameters. We prove asymptotic optimality of a CLP-based policy in a large market regime. However, solving the CLP is difficult, thus we also propose matching policies based on a linear program (LP). We prove asymptotic optimality of an LP-based policy in a large market regime in which drivers are fully utilized.

Title: Empty-car routing in ridesharing systems

Presenter: Anton Braverman

Co-authors: J.G. Dai (Cornell), Liu Xin (Arizona State University), Lei Ying (Arizona State University)

Abstract: We consider a closed queueing network that models the flow of drivers in a ridesharing system such as Lyft or Uber. Each time a driver drops off a passenger at their destination, a routing decision needs to be made. Should the driver stay and wait for the next customer at their current location, or should they drive empty to another part of town to try their luck there?The way this decision is made greatly affects the supply of drivers across a city, and can even cause extreme driver shortages in certain regions. We analyze the fluid model corresponding to our network to develop a centralized routing policy for drivers.

Tuesday, 1:30 - 3:00, Room: 2430

Session: Optimization and Applied Probability

Chair: Harsha Honnappa

Title: Weak convergence approach to the multi-arm bandit problem

Presenter: David Goldberg

Co-authors: Yilun Chen (Georgia Tech), Timur Tankayev (Georgia Tech)

Abstract: For the Bayesian multi-arm bandit problem, the optimal policy is known to be the so-called Gittins index policy. However, this policy is essentially a "black box", and it is apriori unclear how such a policy behaves, i.e. what an observer would actually see if a decision-maker implemented this policy. Using the framework of weak convergence, we shed light onto this question and the behavior of several associated stochastic processes. Time permitting, we will also discuss some related results for Thompson sampling.

Title: Theory and algorithms for robust countable state Markov decision processes

Presenter: Saumya Sinha

Co-authors: Archis Ghate (University of Washington)

Abstract: A majority of existing theoretical results and convergence analyses of solution algorithms for robust Markov decision processes are limited to the finite-state case. In this talk, we will present new results and convergence analyses for the countable state case both with bounded and unbounded rewards.

Title: Optimal Traffic Schedules

Presenter: Harsha Honnappa

Co-authors: Mor Armony (NYU), Rami Atar (Technion)

Abstract: We consider the problem of optimally scheduling a finite, but large, number ofcustomers over a finite time horizon at a single server FIFO queue, in the presence of 'no-shows'. The stochastic optimization problem is unlike a dynamic optimization problem, since the optimal schedule must be chosen before service commences. This complicates the optimization problem significantly, and we consider fluid and diffusion approximations to the stochastic optimization problem. The approximations are developed in a large population limiting regime where the number of customers scales to infinity and the appointment duration scales to zero. We show that in the fluid scale, the heavy-traffic condition is obtained as a result of optimization. We also identity an asymptotically optimal sequence of fluid-scaled schedules that achieve the value of the fluid optimization problem. The fluid-optimal solution indicates that the stochastic optimization problem could be approximated by an equivalent Brownian optimization problem. We prove that when the time horizon is large, the value of the Brownian optimization problem is achieved by a stationary reflected Brownian motion. We also identify a sequence of diffusion-scaled schedules that achieve the value of the Brownian optimization problem.

Title: Uniformly bounded regret for the multi-secretary problem

Presenter: Alessandro Arlotto

Co-authors: Itai Gurvich (Cornell University)

Abstract: In the classic formulation of the secretary problem, n positive numbers are sequentially presented to a decision maker who decides when to stop and pick the current element. The goal is to maximize probability of choosing the largest of the n items. In the k-choice (multi-secretary) variant the decision maker is allowed to choose k elements, and the goal is to maximize the sum of the chosen elements. Assuming that items are independent, identically distributed, and drawn from a known distribution with finite support, we prove that the best regret --- the gap between the optimal online policy and the offline (sorting) upper bound where all n values are made visible before any selection is made --- is uniformly bounded in k and n.

Tuesday, 1:30 - 3:00, Room: L-130

Session: Recent Advances in High-Dimensional Statistics

Chair: Wenguang Sun

Title: Concentration inequalities for empirical processes of linear time series

Presenter: Wei Biao Wu

Co-authors: Likai Chen (University of Chicago)

Abstract: The talk concerns suprema of empirical processes for linear time series indexed by functional classes. We derive a Gaussian approximation and an upper bound for the tail probability of the suprema under conditions on the size of the function class, the sample size, temporal dependence and the moment conditions of the underlying time series. Due to the dependence and heavy-tailedness, our tail probability bound is substantially different from those classical exponential bounds obtained under the independence assumption in that it involves an extra polynomial decaying term. We allow both short- and long-range dependent processes. For empirical processes indexed by half intervals, our tail probability inequality is sharp up to a multiplicative constant.

Title: A General Framework for Large-Scale Two-Sample Inference

Presenter: Wenguang Sun

Co-authors: T. Tony Cai, Weinan Wang, Yin Xia

Abstract: Two-sample multiple testing has a wide range of applications. The conventional practice is to first reduce the original observations to a vector of p-values and then choose a cutoff to adjust for multiplicity. However, the data reduction step could cause significant loss of information and thus lead to suboptimal testing procedures. In this paper, we introduce a new framework for two-sample multiple testing by constructing primary and auxiliary variables from the original observations and incorporating both in the inference procedure to improve the power. A data-driven multiple testing procedure is developed by employing a covariate-assisted ranking and screening (CARS) approach that optimally combines the information from both the primary and auxiliary variables.The proposed procedure is shown to be asymptotic valid with proper control of the false discovery rate (FDR). Numerical results confirm the effectiveness of CARS in FDR control and show that it achieves substantial power gain over existing methods.

Title: Random projection ensemble classification

Presenter: Timothy I Cannings

Co-authors: Richard J. Samworth (University of Cambridge)

Abstract: We introduce a general method for high-dimensional classification, based on careful combination of the results of applying an arbitrary base classifier to random projections of the feature vectors into a lower-dimensional space. In one special case presented here, the random projections are divided into non-overlapping blocks, and within each block we select the projection yielding the smallest estimate of the test error. Our random projection ensemble classifier then aggregates the results of applying the base classifier on the selected projections, with a data-driven voting threshold to determine the final assignment. Our theoretical results elucidate the effect on performance of increasing the number of projections. Moreover, under a boundary condition implied by the sufficient dimension reduction assumption, we control the test excess risk of the random projection ensemble classifier. A simulation comparison with several other popular high-dimensional classifiers reveals its excellent finite-sample performance.

Title: Cross: Efficient Low-rank Tensor Completion

Presenter: Anru Zhang

Co-authors:

Abstract: The completion of tensors, or high-order arrays, attracts significant attention in recent research. Current literature on tensor completion primarily focuses on recovery from a set of uniformly randomly measured entries, and the required number of measurements to achieve recovery is not guaranteed to be optimal. In addition, the implementation of some previous methods are NP-hard. In this article, we propose a framework for low-rank tensor completion via a novel tensor measurement scheme we name Cross. The proposed procedure is efficient and easy to implement. In particular, we show that a third order tensor of Tucker rank-(r_1, r_2, r_3) in p_1-by-p_2-by-p_3 dimensional space can be recovered from as few as $r_1r_2r_3 + r_1(p_1-r_1) + r_2(p_2-r_2) + r_3(p_3-r_3)$ noiseless measurements, which matches the sample complexity lower-bound. In the case of noisy measurements, we also develop a theoretical upper bound and the matching minimax lower bound for recovery error over certain classes of low-rank tensors for the proposed procedure. The results can be further extended to fourth or higher-order tensors. Simulation studies show that the method performs well under a variety of settings. Finally, the procedure is illustrated through a real dataset in neuroimaging.

Tuesday, 1:30 - 3:00, Room: L-120

Session: Learning in Markov Decision Processes

Chair: Michael Katehakis and Flora Spieksma

Title: Optimizing Breast Cancer Diagnostic Decisions to Reduce Overdiagnosis

Presenter: Oguzhan Alagoz

Co-authors: Sait Tunc (University of Chicago), Elizabeth Burnside (University of Wisconsin-Madison)

Abstract: Mammography has been the most commonly used technique for early diagnosis of breast cancer. However, mammography has several negative effects such as false positives and overdiagnosis. Overdiagnosis, defined as diagnosing a cancer that would otherwise not cause symptoms or death in a patient's lifetime, has been a growing concern for breast cancer screening. The major reason for overdiagnosis is due to the difficulty in predicting breast cancer subtypes and their potential on future outcomes. Overdiagnosis rates, estimated to be around 10 to 40%, may be reduced if indolent breast cancer subtypes can be identied and followed with noninvasive imaging rather than biopsy and treatment. However, there are no validated/established guidelines for radiologists to decide when to choose noninvasive imaging options. In this study, we develop a large-scale finite-horizon Markov decision process (MDP) with over 4 million states to optimize the post-mammography diagnostic decisions. Our MDP is large since it explicitly represents various breast cancer subtypes unlike previous studies optimizing diagnostic decisions after mammography. To reduce the computational burden, we develop and prove the optimality of a novel algorithm that relies on upper bounds on the optimal decision thresholds. We project the high-dimensional MDP onto two lower-dimensional MDPs to obtain feasible and tight upper bounds for the optimal decision thresholds. We then use real data from two private mammography databases to conduct numerical experiments and solve our MDP optimally. We find that the use of a large-scale MDP model and our novel solution algorithm can save up to 9% of overdiagnosis among biopsied women. We also observe that the reduction in overdiagnosis rates achieved by our model increases with age.

Title: When to observe a Markov process

Presenter: Aditya Mahajan

Co-authors:

Abstract: We consider a model in which an observer may either incur a cost to observethe state of a Markov process or use past observations to estimate the stateand incur a cost for inaccurate estimation. The objective is to choose astrategy that minimize the expected total observation and estimation cost.Such a system can be modelled as a partially observable Markov decisionprocess (POMDP). We analyze the reachable set of the POMDP and show that itcan be viewed as a countable state Markov decision process (MDP). Weidentify conditions under which the optimal observation strategy has athreshold structure.

Title: Optimal Data Driven Policies for MDPs

Presenter: Michael N. Katehakis

Co-authors: Wesley Cowan (Rutgers)

Abstract: We first give a brief survey of the state of the art of the area of computing optimal data driven (adaptive) policies for MDPs with unknown rewards and or transition probabilities. Then, we present certain simple algorithms for adaptively optimizing the average reward in an unknown irreducible MDP. The first

algorithm uses estimates for the MDP and and chooses actions by maximizing an inflation of the estimated right hand side of the average reward optimality equations. The second is based on estimating the optimal rates at which actions should be taken. For the first we show that the total expected reward obtained by this algorithm up to time n is within O(ln n) of the reward of the optimal policy, and in fact it achieves asymptotically minimal regret. Various computational challenges and simplifications will be discussed.

Title: Primal-Dual Policy Learning

Presenter: Mengdi Wang

Co-authors:

Abstract: We consider the online estimation of the optimal policy estimation of Markov decision processes. We propose a Stochastic Primal-Dual (SPD) method which exploits the inherent minimax duality of Bellman equations. SPD updates a few coordinates of the value and policy estimates as state transitions are sampled. We show that the SPD has superior space and computational complexity and it finds with high probability an optimal policy using near optimal samples and iterations.

Tuesday, 3:30 - 5:00, Room: 2110

Session: Control of Service Systems

Chair: Yao Yu

Title: An Invitation Control Policy for Proactive Service System: Balancing Efficiency, Value and Service Level

Presenter: Galit Yom-Tov

Co-authors: Yueming Xie (Technion), Liron Yedid-Zion (Technion)

Abstract: Proactive service systems permit controllable arrival rate managed by the service provider, which is different from classic service systems. Conceptually, some (or all) of the customers are invited to the system, so as to allow for a better control over operational indicators and profitability. Such proactive service system is used, for example, to model online chat service system, or for planning preventive care strategies for health care service providers.Through an empirical study of a proactive chat service system, the validity of customer ranking information is elaborated for optimizing invitation control. It is also shown that service level measures can be formulated in terms of penalty for abandonment and cost of waiting. Hence, a infinite-time-horizon multiclass multiserver queueing system has been built with impatient customer. We find asymptotic optimal policy using a fluid approximation solving a linear programming problem that maximizes revenue. The asymptotic optimal invitation policy we developed, invites customers by their r_ ranking in decreasing order until there are no idle servers. Meanwhile, an equivalent threshold

policy is proposed, such policy is easy to implement in practice. Numerical simulations are performed to demonstrate the strength of the policy and identify its limitations. We show that the fluid policy has a good performance but is also crude. In order to refine the fluid policy, we analyzed a fluid approximation of the system under more flexible threshold policy. The equilibrium is found to be strongly depends on system parameters. In particular, it depends on the threshold value. It is also shown that the equilibrium is globally asymptotically stable via trajectory and Lyapunov analysis. Furthermore, in order to propose an invitation policy for proactive service system that balances revenue and service level, the probability of implementing admission control is approximated, and several approximations of performance metrics are calculated. Simulations are performed to examine the performance of these approximations.

Title: Dynamic Pricing In One-way Car Sharing Networks: A Distributional Fluid Approximation Approach

Presenter: Ling Zhang

Co-authors: Yunan Liu (North Carolina State Univeristy), Liu Yang (National University of Singapore), Shuangchi He (National University of Singapore)

Abstract: Balancing supply and demand across different areas is a critical issue in one-way car sharing networks. We study dynamic pricing in order to maximize the profit of a car sharing network. Since the stochastic network model is analytically intractable, we propose a fluid approximation to represent the supply and demand of vehicles. In contrast to conventional transportation fluid models that assume deterministic processing times, general rental time distributions are built into our fluid model. Moreover, our model allows for time-varying demand rates and rental time distributions. Under this formulation, dynamic pricing is reduced to a quadratic optimization problem that is efficiently solvable.

Title: Dimensioning of the fixed-cycle traffic-light queue

Presenter: Marko Boon

Co-authors: Marko Boon, Augustus Janssen, Johan van Leeuwaarden (Eindhoven University of Technology)

Abstract: The Quality-and-Efficiency Driven (QED) regime has proven its tremendous value in a wide range of application areas. We apply the QED principle of matching capacity with demand to urban road traffic. We develop a capacity sizing rule for determining fixed-cycle traffic signal settings, ensuring that the specified traffic flow will operate in the QED regime. As a consequence, we can let the system load tend to one, while retaining a strictly positive probability that arriving vehicles experience no delay. Moreover, we show how to optimize traffic signal settings such that ALL flows at the intersection exhibit this behavior.

Title: Optimal Controls to Remote Queues

Presenter: Yao Yu

Co-authors: Shuangchi He (National University of Singapore) and Yunan Liu (North Carolina State University)

Abstract: We develop an efficient routing policy for remote queueing systems, in which each arrival, after being routed to one of the several dedicated queues, will experience a pre-arrival delay. Motivated by service systems in which system state (e.g., queue length and waiting time) is available for routing decisions, we intend to use pre-arrival delays to model commute times of arrivals, such as patients' transportation times before arriving at clinics and data packets' transmission times to web servers. In order to minimize the delay, we propose a new state-dependent probabilistic routing policy.

Tuesday, 3:30 - 5:00, Room: 2110

Session: Scaling Limits for Multi-Server Queues

Chair: Michel Mandjes

Title: Large Deviations for Queues in Cells

Presenter: Justin Dean

Co-authors: Ayalvadi Ganesh (Bristol), Edward Crane (Heilbronn Institute, Bristol)

Abstract: Biochemical processes within cells of the transcription of DNA into RNA, and the translation of RNA into protein, can be described using stochastic models. Questions of biological interest pertain to fluctuations in protein molecule numbers within cells, and their temporal dynamics. In this talk, we present models of transcription and translation as infinite server queues. Motivated by these models, we derive a large deviation principle for the empirical measure of a Cox process on a Polish space. We use this to obtain sample path LDPs for molecule numbers within cells.

Title: Weak convergence of modulated Erlang models

Presenter: H. M. Jansen

Co-authors: M. Mandjes (University of Amsterdam), K. De Turck (CentraleSuplec), S. Wittevrongel (Ghent University)

Abstract: We consider Erlang A, B, and C models that are modulated by a continuous-time Markov chain (called the background process), meaning that the arrival rate and server speed depend on the state of the background process. This process may be interpreted as an independently evolving random environment to which the queue reacts. Often, this has a detrimental effect on the performance of a queue, although it may have a positive impact on performance as well. This has been shown recently for the Erlang B (loss) model in steady state. In this talk, we focus on the transient behavior of modulated Erlang models under a diffusion scaling. In particular, we would like to know how the background process influences the limiting behavior of the queue. To this end, we introduce a QED-type scaling together with a sublinear, linear, or superlinear speedup of the time scale of the background process. Under each of these scalings we derive a

diffusion approximation for the number of jobs in the system. The form of the diffusion approximation shows the strong influence of the background process on the limiting results.

Title: The Method of Chaining for Many Server Queues

Presenter: Guodong Pang

Co-authors: Yuhang Zhou (Penn State University)

Abstract: The method of chaining, originating from Kolmogorov, has been a very powerful tool to obtain probability and moment bounds for stochastic processes. We explore the application of the method of chaining in non-Markovian many-server queues with a general arrival process, and with either (i) general time-varying service times (e.g., arrival dependent services), or (ii) dependent service times. In these models, we study two-parameter stochastic processes that can be used to describe the system dynamics, in particular, X(t,y) representing the number of jobs in the system at time t that have received an amount of service less than or equal to y (or that have a residual amount of service strictly greater than y). We prove functional central limit theorems for these two-parameter processes. The method of chaining provides important maximal probability and moment bounds on the two-parameter processes, which are key to prove their weak convergence.

Title: The Exact Analysis of the Markov-modulated Erlang Loss System

Presenter: Peter Taylor

Co-authors: Michel Mandjes (University of Amsterdam), Koen de Turck (Universit'e Paris Saclay)

Abstract: We present a closed-form expression for the stationary distribution of the Markov-modulated Erlang loss queue. This, in particular, provides us with an explicit formula for the probability that the queue is full, which can be regarded as the Markov-modulated counterpart of the well-known Erlang loss formula. Central to this result is the proof of the non-singularity of a certain matrix, which has an interesting physical interpretation.

Tuesday, 3:30 - 5:00, Room: 2130

Session: Recursive Estimation & Control of Stochastic Processes

Chair: Thomas A. Weber

Title: Monte Carlo simulation of aggregate insurance claims using reproducibility

Presenter: Ad Ridder

Co-authors: Shaul Bar Lev (University of Haifa)

Abstract: In this paper we consider the problem of computing tail probabilities of the distribution of a random sum of positive random variables. We assume that the individual variables follow a reproducible natural exponential family (NEF) distribution, and that the random number has a NEF counting distribution with a cubic variance function. This specific modelling is supported by data of the aggregated claim distribution of an insurance company. Large tail probabilities are important as they reflect the risk of large losses, however, analytic or numerical expressions are not available. We propose several simulation algorithms which are based on an asymptotic analysis of the distribution of the counting variable and on the reproducibility property of the claim distribution.

Title: First Passage Time Estimation of Diffusion Processes and Financial Application

Presenter: Francois Watier

Co-authors: Imene Allab (Universite du Quebec a Montreal)

Abstract: We present a Monte Carlo-based algorithm for estimating the FPT density of a one-dimensional time-homogeneous SDE through a time-dependent frontier. Brownian bridges are considered, as well as localized Daniels curve approximations, to obtain tractable estimations of upcrossing probabilities between successive points of a simulated path of the process. Finally we will apply our technique to a portfolio management problem.

Title: Risk sensitive portfolio optimization in a jump diffusion model with regimes

Presenter: Anindya Goswami

Co-authors: Milan Kumar Das (IISER Pune), Nimit Rana (York University)

Abstract: This article studies a portfolio optimization problem, where the market consisting of several stocks is modeled by a multi-dimensional jump diffusion process with age-dependent semi-Markov modulated coefficients. We study risk sensitive portfolio optimization on the finite time horizon. We study the problem by using a probabilistic approach to establish the existence and uniqueness of the classical solution to the corresponding Hamilton-Jacobi-Bellman (HJB) equation. We also implement a numerical scheme to investigate the behavior of solutions for different values of the initial portfolio wealth, the maturity and the risk of aversion parameter.

Title: Dynamic Credit-Collections Optimization

Presenter: Thomas A. Weber

Co-authors: Naveed Chehrazi (UT Austin), Peter Glynn (Stanford)

Abstract: Based on a dynamic model of the stochastic repayment behavior exhibited by delinquent credit-card accounts in the form of a self-exciting point process, a bank can control the arrival intensity of

repayments using costly account-treatment actions. A semi-analytic solution to the corresponding stochastic optimal control problem is obtained using a recursive approach. For a linear cost of treatment interventions, the optimal policy in the two-dimensional (intensity,balance)-space is described by the frontier of a convex action region. The unique optimal policy significantly reduces a bank's loss given default and concentrates the collection effort onto the best possible interventions at the best possible times, so as to minimize the sum of the expected discounted outstanding balance and the discounted cost of the collection effort, thus maximizing the net value of any given delinquent credit-card account.

Tuesday, 3:30 - 5:00, Room: 2410

Session: Queues in Random Environment

Chair: Wei You

Title: On the M/M/1 queue in a Markovian environment

Presenter: Brian Fralix

Co-authors: Jason Joyner (Wingate University)

Abstract: We present a new representation for the stationary distribution of an M/M/1 queue, whose arrival and service rates are both governed by an external Markovian environment. Implementation issues, as well as possible extensions to multiserver models, will also be discussed.

Title: Infinite server queues with shot noise arrival intensities

Presenter: David Koops

Co-authors: M. Mandjes (Univ of Amsterdam), O. Boxma (TU Eindhoven)

Abstract: We study infinite-server queues in which the arrival process is a Cox process (or doubly stochastic Poisson process), of which the arrival rate is given by a shot-noise process. A shot-noise rate emerges naturally in cases where the arrival rate tends to exhibit sudden increases (or: shots) at random epochs, after which the rate is inclined to revert to lower values. Exponential decay of the shot noise is assumed, so that the queueing systems are amenable to transient analysis. In addition, heavy-traffic asymptotics can be derived under a certain scaling, and the analysis can be extended to a network setting. A closely related arrival model, the (self-exciting) Hawkes process, will be considered in a queueing setting as well.

Title: Queueing systems in a random environment: asymptotic analysis and MOL staffing

Presenter: Mariska Heemskerk

Co-authors: Michel Mandjes (UvA), Johan van Leeuwaarden (TU/e), Julia Kuhn (UvA), Britt Mathijsen (TU/e)

Abstract: In this talk we consider queueing systems in a random environment. As the latter entails that no communication between customers and servers occurs, the uncertain arrival stream is usually modeled under the assumption of Poisson arrivals. It's a general finding that the often used (nonhomogenous) Poisson process in some cases fails to capture the dynamics of an arrival process: not only do some systems face arrival streams that are almost deterministic (mean >> variance), there are others, which are of special interest to us, that have to deal with highly variable arrivals (mean << variance) - they face overdispersion. The model we propose here is a comprehensive yet simple model for overdispersed arrival processes that can additionally handle nonhomogeneity and dependency between arrival rates of subsequent time slots. The random environment is modeled by a mixed Poisson process, where the random parameter takes a new value every time slot of fixed size.We are interested in the effect of such an arrival process on the performance of an infinite-server system. As it turns out, in a rapidly changing random environment (i.e., slot size is small relative to system size) the overdispersion of the arrival process hardly affects system behavior, whereas in a slowly changing random environment it is fundamentally different; this applies to both the central limit and the large deviations regime. Having studied these effects, we apply our results via MOL staffing for the corresponding finite-server counterpart. The resulting staffing procedure stabilizes system performance rather well.

Title: Time-Varying Robust Queueing

Presenter: Wei You

Co-authors: Ward Whitt (Columbia)

Abstract: We develop a time-varying robust-queueing (TVRQ) algorithm for the continuous-time workload in a single-server queue with a time-varying arrival-rate function. We apply this TVRQ to develop approximations for (i) the time-varying expected workload in models with a general time-varying arrival-rate function and (ii) for the periodic steady-state expected workload in models with a periodic arrival-rate function. We apply simulation to examine the performance of periodic TVRQ (PRQ). We find that PRQ predicts the timing of peak congestion remarkably well. We show that the PRQ converges to a proper limit in appropriate long-cycle and heavy-traffic regimes, and coincides with long-cycle fluid limits and heavy-traffic diffusion limits for long cycles.

Tuesday, 3:30 - 5:00, Room: 2420

Session: Efficient Marketplace Design

Chair: Yash Kanoria

Title: Communication requirements and Informative signaling in matching markets

Presenter: Yash Kanoria

Co-authors: Itai Ashlagi (Stanford), Mark Braverman (Princeton), Peng Shi (Microsoft Research, USC)

Abstract: We study how much communication is needed to find a stable matching in a two-sided matching market with private preferences. The Gale-Shapley deferred acceptance (DA) algorithm requires communication effort per agent that can grow (almost) linearly in the size of the market, and this amount of communication is also necessary to find a stable matching in worst case. We show that in a two-sided market with workers and firms, where workers' preferences are arbitrary and private and firms' preferences follow an additively separable latent utility model, a stable matching can be found with much less communication effort. Our efficient communication protocol modifies workers proposing DA, via firms signaling workers they privately like, while also broadcasting qualification requirements to discourage other workers who have no realistic chance of being hired. In the special case of tiered random markets, simultaneous and decentralized two-sided signaling suffices. Our protocols suggest that to reduce market congestion, each agent should reach out to her favorites among those who are likely to consider her, while waiting for her dream matches to approach her.

Title: Analysis of large scale closed networks with reservationn

Presenter: Christine Fricker

Co-authors: Cdric Bourdais (Ecole Polytechnique)

Abstract: Car-sharing systems can be modeled by closed networks with reservation. In these systems, a user arrives at a station, picks up a car, uses it for a while and returns it to another station. He can book the car and the parking place at the arrival. Thus contrary to previous models on bike sharing systems, the state of a station is multi-dimensional.Mean-field results allow to analyse the limit of an homogeneous model as the number of stations is large. Even if the state at the different stations is not Markov, its limit process can be obtained. We prove that there is a unique equilibrium point of the dynamical system. Then the influence of the different parameters on the system behaviour is discussed, especially the fleet size. It differs from a similar bike-sharing system, because of reservation, when the traffic is high.

Title: Online Resource Allocation under Partially Learnable Demand

Presenter: Dawsen Hwang

Co-authors: Vahideh Manshadi

Abstract: We study a basic online resource allocation problem, known as the single-leg revenue management with 2 fare classes, where stochastic information about demand is unknown a priori, and it can only be partially learned. In our demand model, an adversary determines a sequence of customers to be revealed to the online algorithms. However, a random subset of customers does not follow this prescribed order, and instead, arrives at uniformly random times. The presence of such customers enables

us to partially learn the future demand. We use this to design online algorithms (adaptive and non-adaptive) with competitive ratios significantly higher than that of algorithms designed for adversarial customer arrival model.In the two extreme cases that all or none of the customers follow the adversary, we recover known performance guarantees. For the regime in between, we show that our algorithms achieve competitive ratios better than what can be achieved by algorithms designed for the extreme cases. We also show that using an adaptive algorithm is particularly beneficial when the initial resource capacity is of the same order of as time horizon (maximum number of customers). Our work bridges the gap between adversarial and stochastic arrival models, and it highlights the value of learning in online resource allocation.

Title: Segmenting Two-Sided Markets via Directed Discovery

Presenter: Sid Banerjee

Co-authors: Kostas Kollias (Google), Sreenivas Gollapudi (Google), Kamesh Munagala (Duke)

Abstract: A common feature of many online marketplaces is that the platform has full control over search and discovery, but prices are determined by the buyers and sellers. Motivated by this, we study the algorithmic aspects of market segmentation via directed discovery in two-sided markets with endogenous prices. We consider a model where an online platform knows each buyer/seller's characteristics, and associated demand/supply elasticities. Moreover, the platform can use discovery mechanisms (search/recommendation/etc.) to control which buyers/sellers are visible to each other. This leads to a segmentation of the market into pools, following which buyers and sellers endogenously determine market-clearing transaction prices within each pool. The aim of the platform is to maximize the resulting volume of transactions/welfare in the market. We develop efficient algorithms for this setting, with provable guarantees under a variety of assumptions on the demand and supply functions.

Tuesday, 3:30 - 5:00, Room: 2430

Session: Control of Queues

Chair: Hayriye Ayhan and Doug Down

Title: Optimal slot reservation with application to disaster relief

Presenter: Michael Veatch

Co-authors:

Abstract: In a multiclass, multiserver loss system with arrival control, it can be optimal to reject arrivals from a lower priority class, reserving servers for higher priority customers. Motivated by the application, we consider deterministic arrivals, using an approximate Markov representation at these deterministic times. The form of the optimal policy is investigated and compared numerically to a fixed reservation time, where

a slot is reserved for a future higher priority arrival due within a certain time. The model is applied to scheduling an airlift into a small airport after a disaster, using data from the Haiti earthquake. Deterministic arrivals model the fact that flights are scheduled in advance.

Title: Asymptotic Performance of Energy-Aware Multiserver Queueing Systems with Setup Times

Presenter: Douglas Down

Co-authors: Vincent Maccio (McMaster)

Abstract: We study an M/M/c queue where each server can be turned on, with an exponentially distributed setup time, or turned off instantaneously. The control problem of interest involves a trade off between energy costs and performance. Due to the complexity of the model analysis, authors often examine a specific policy. Moreover, different authors examine different policies under different cost functions. This in turn causes difficulties when making statements or drawing conclusions regarding competing policies. Therefore, we analyze this well established model under the asymptotic regime where the number of servers approaches infinity, while the load remains fixed, and show that not only are many of the policies in the literature equivalent under this regime, but are also optimal under any cost function which is non-decreasing in the expected energy cost and response time.

Title: A Markov Decision Process Approach for Optimal Control of Complex Authentication Systems

Presenter: Daniel F. Silva

Co-authors: Bo Zhang (IBM Research), Hayriye Ayhan (Georgia Tech)

Abstract: We consider an authentication system that receives requests from different types of users, that must be assigned to one of several authentication methods. We model the system as a multi-class network of parallel, multi-server queues. We assume that each request has a known probability of coming from an impostor, which depends on its class. A central controller must assign an authentication method to each request, considering the request's class, the state of the system and the characteristics of each available method. Each method may have different capacity, service rate, level of security (represented by two error probabilities), holding cost and operating cost. We consider three objectives: minimizing average error probability, holding cost and operating cost. We use constrained and unconstrained Markov decision processes to characterize the structure of policies that effectively balance the three objectives.

Title: Semi-online Scheduling for Queues with Revealed Workloads and Degreading Server

Presenter: Jin Xu

Co-authors: H. Tran (TAMU), N. Gautam (TAMU) and S. Bukkapatnam (TAMU)

Abstract: Motivated by applications in smart custom manufacturing, we consider a queueing system with a single machine where arrivals occur arbitrarily. Upon arrival at the machine but before start of processing, the workload of jobs are revealed. The jobs can be processed by the machine at different speeds. Faster

speeds while beneficial also results in the tool getting used up faster. It takes a significant time to replace a tool. Under such a setting, our objective is to obtain a control policy to determine both the processing speed as well as a maintenance plan for tool replacement. We formulate an optimization problem and develop strategies that we show are asymptotically optimal. We present numerical examples to illustrate our results.

Tuesday, 3:30 - 5:00, Room: L-130

Session: Learning with Algebraic and Combinatorial Structures

Chair: Philippe Rigollet

Title: Optimal rates of estimation for the multi-reference alignment problem

Presenter: Jonathan Weed

Co-authors: Afonso S. Bandeira (NYU), Philippe Rigollet (MIT)

Abstract: How should one estimate a signal, given only access to noisy versions of the signal corrupted by unknown circular shifts? This simple problem has surprisingly broad applications, in fields from structural biology to aircraft radar imaging. We describe how this model can be viewed as a multivariate Gaussian mixture model whose centers belong to an orbit of a group of orthogonal transformations. This enables us to derive matching lower and upper bounds for the optimal rate of statistical estimation for the underlying signal. These bounds show a striking dependence on the signal-to-noise ratio of the problem.

Title: Exact recovery in the Ising blockmodel

Presenter: Quentin Berthet

Co-authors: Philippe Rigollet (MIT), Piyush Srivastava (Caltech)

Abstract: We consider the problem associated with recovering the block structure of an Ising model given independent observations on the binary hypercube. This new model, called the Ising blockmodel, is a perturbation of the mean field approximation of the Ising model known as the Curie-Weiss model: the sites are partitioned into two blocks of equal size and the interaction between those of the same block is stronger than across blocks, to account for more order within each block. We study probabilistic, statistical and computational aspects of this model in the high-dimensional case when the number of sites may be much larger than the sample size.

Title: Testing Network Structure Using Relations Between Small Subgraphs Probabilities

Presenter: Chao Gao

Co-authors: John Lafferty (UChicago)

Abstract: We study the problem of testing for structure in networks usingrelations between the observed frequencies of small subgraphs.Starting with the simple test statistic $$T_1=(\text{edge frequency})^3 - \text{triangle frequency},$$ we prove a central limittheorem for T_1 under an Erd\H{o}s-R\'{e}nyi null model, andanalyze the power of the test statistic under a general class ofalternative models. In particular, when the alternative is a k-community stochastic block model, with k unknown, the power ofthe test approaches one. Moreover, the signal-to-noise ratiorequired is strictly weaker than that required for communitydetection. We also study the relation with other statistics overthree-node subgraphs, and analyze the error under two naturalalgorithms for sampling small subgraphs. Together, our results showhow global structural characteristics of networks can be inferred fromlocal subgraph frequencies, without requiring the global communitystructure to be explicitly estimated.

Title: On the Computational Invariance of Community Detection to Distribution

Presenter: Guy Bresler

Co-authors: Wasim Huleihel (MIT)

Abstract: We consider the problem of recovering a hidden community of size K from a graph where edges between members of the community have label X drawn i.i.d. according to P and all other edges have labels drawn i.i.d. according to Q. The information limits for this problem were characterized by Hajek-Wu-Xu in 2016 in terms of the KL-divergence between P and Q. We complement their work by showing that for a broad class of distributions P and Q one may reduce to the case P = Ber(p) and Q = Ber(q) and vice versa. This implies that the computational difficulty is independent of the choice of distribution within the class (up to polynomial time computations).

Tuesday, 3:30 - 5:00, Room: L-120

Session: Sequential Decisions and Optimal Stopping

Chair: Alessandro Arlotto

Title: Design principles for multi-period flexible production systems

Presenter: Yehua Wei

Co-authors: Cong Shi (Michigan), Yuan Zhong (Chicago)

Abstract: In this talk, we develop principles to design process flexibility in a multi-period make-to-order production system. We introduce a notion of effective flexibility design, which determines whether a flexible system can achieve the same performance as full flexibility when plant capacities become highly utilized.

Using this notion, we prove that in a system with m plants and n products, we can construct an effective sparse flexibility structures with m+n arcs. We also show that the requirement of m+n arcs is necessary, as even the best flexibility structure with m+n-1 arcs cannot achieve the same notion of effectiveness.

Title: A sequential stopping rule for multivariate sectioning method

Presenter: Jing Dong

Co-authors: Peter Glynn (Stanford), Yi Zhu (Northwestern)

Abstract: Sectioning method is a cancellation method for simulation output analysis where variance estimation is difficult. Under mild assumptions on the simulation output processes, we obtain asymptotically valid confidence regions with guaranteed error bound when applying the sectioning method under a sequential stopping scheme. In particular, we characterize the scaling parameters in closed from. We also provide examples where the framework can be applied.

Title: Optimal Stopping Problems With Partial Information

Presenter: Assaf Zeevi

Co-authors: Alexander Goldenshluger, Haifa University

Abstract: Consider an optimal stopping problem where a decision maker observes sequentially independent random variables $X_1, ... ,X_n$ from a common distribution F. The goal is to design a decision rule τ that "stops" the sequence at a point that depends on all past observations, and maximizes the expected value of X_τ. If the distribution F is known, the optimal stopping rule is a threshold policy that is obtained by backward recursion. This work has a long and storied history. The problem of optimal stopping with partial information arises when the distribution F is {\it unknown}. These problems have been studied to a much lesser extent and solutions, decision rules and resulting performance, tend to rely on problem specifics (such as the parametric family of the underlying distribution). In this talk we will consider this problem in a minimax framework and develop a simple rank-based algorithm that achieves near-optimal (asymptotic) regret.

Title: A O(log n)-optimal policy for the dynamic and stochastic knapsack problem with equal values

Presenter: Alessandro Arlotto

Co-authors: Xinchang Xie (Duke University)

Abstract: We study a dynamic and stochastic knapsack problem in which a decision maker is sequentially presented with n items and needs to select which items to include in a knapsack with fixed capacity c. Arriving items have non-negative, independent sizes with common continuous distribution F, and the decision maker needs to decide whether to select or reject an item when it is first presented and its size is revealed. The decision maker seeks to maximize the expected number of selected items, subject to the capacity constraint. We propose a simple adaptive online policy and prove that under mild regularity

conditions on the distribution function F, the expected number of selections of our heuristic policy is within O (log n) of the optimal. We also discuss how the distribution of the number of selected items under such an adaptive policy compares with the number of items selected by the optimal policy.

Wednesday, 8:30 - 10:00, Room: 2110

Session: Dual-Sourcing Inventory Control with Leadtimes

Chair: Jan Van Mieghem

Title: Optimal Policies for a Dual-Sourcing Inventory Problem with Endogenous Stochastic Lead Times

Presenter: Jing-Sheng Song

Co-authors: Li Xiao (Chinese University of Hong Kong) , Hanqin Zhang (National University of Singapore), Paul Zipkin (Duke)

Abstract: We consider a single-product, two-source inventory system with Poisson demand and backlogging. Inventory can be replenished through a normal supply source, which consists of a two-stage tandem queue with exponential production time at each stage. We can also place an emergency order by skipping the first stage, for a fee. There is no fixed order cost. There are linear order, holding, and back-order costs. Through a new approach, we obtain optimal ordering policies for the discounted or long-run average cost and also characterize near-optimal heuristic policies. The approach consists of foursteps. The first step is to establish an equivalent system, in the sense that it has the same optimal policy as the original system. The second step is to construct a tandem queueing system, where costs are charged in accord with the equivalent system's cost structure. The third step derives an optimal control of the service rate at each server so as to minimize the tandem queue's system-wide cost. The fourth and final step is to translate the queue's optimal policy to an optimal policy for the equivalent system and hence the original system.

Title: Beating the curse of dimensionality in inventory problems with lead times

Presenter: David Goldberg

Co-authors: Linwei Xin, University of Illinois

Abstract: Many classical inventory models become notoriously challenging to optimize in the presence of positive lead times, since the state-space blows up and dynamic programming techniques become intractable. This includes, for example, lost sales models with positive lead times, and dual-sourcing models with positive lead time gap between the two suppliers. In this talk, we will present a new algorithmic approach to such problems, which shows that as the lead time grows large, simple policies become asymptotically optimal. These results are quite surprising, as this setting had remained an open algorithmic

challenge for over forty years. In particular, we will show that a simple constant-order policy is asymptotically optimal for lost sales models with large lead times, and provide explicit bounds on the optimality gap which demonstrate good performance even for small-to-moderate lead times. We will also show that the so-called Tailored-Base Surge heuristic for dual-sourcing problems is asymptotically optimal as the lead time gap between the two sources grows large. In both cases, our results provide a new algorithmic approach to these problems, as well as a solid theoretical foundation for the good performance of these algorithms observed numerically by previous researchers. Our approach combines ideas from the theory of random walks and queues, convex analysis, and inventory control.

Title: Robust Dual Sourcing

Presenter: Jan Van Mieghem

Co-authors: Jiankun Sun (Northwestern University)

Abstract: We present the first Robust Optimization of Inventory Control using Two Supply Sources. We present the optimal policy; demonstrate its performance relative to proferred policies in the literature.

Wednesday, 8:30 - 10:00, Room: 2120

Session: Asymptotic Analysis of Load Balancing Service Systems

Chair: Jim Dai

Title: Delay, memory, and messaging tradeoffs in distributed service systems

Presenter: Martin Zubeldia

Co-authors: David Gamarnik (MIT) and John Tsitsiklis (MIT)

Abstract: We consider a distributed service model in which arriving jobs are immediately dispatched to one of several queues associated with n identical servers. We assume that the dispatching decisions are made by a central dispatcher endowed with a finite memory, and with the ability to exchange messages with the servers.We study the fundamental resource requirements (memory bits and message exchange rate), in order to drive the expected steady-state queueing delay of a typical job to zero, as n increases. We propose a certain policy and establish (using a fluid limit approach) that it drives the delay to zero when either (i) the message rate grows superlinearly with n, or (ii) the memory grows superlogarithmically with n. Moreover, we show that any policy that has a certain symmetry property, and for which neither condition (i) or (ii) holds, results in an expected queueing delay which is bounded away from zero.Finally, using the fluid limit approach once more, we demonstrate a surprising phase transition in the expected queueing delay in steady-state. In particular, we show that for any given $\alpha>0$, if our policy only uses a linear message rate αn, the resulting asymptotic (as $n\to\infty$) expected queueing delay is

upper bounded, uniformly over all $\lambda>1$. This is a significant improvement over the usual M/M/1-queue delay scaling of $1/(1-\lambda)$, obtained when $\alpha=0$, and even over the popular "power-of-d-choices" policy, in which the expected delay scales as $\log \left(1/(1-\lambda)\right)$.

Title: Stein's Method for Heavy-Traffic Analysis of the Super-Market Model

Presenter: Lei Ying

Co-authors:

Abstract: This talk presents a recent result on the performance of the supermarket model under the-power-of-two-choices in the heavy traffic regime. Using Stein's method and the perturbation theory, we bound (1) the expected queue length of the supermarket model and (2) the mean-square distance between the stationary distribution of the supermarket model with N servers and its corresponding mean-field approximation.

Title: Universality in load balancing

Presenter: Johan van Leeuwaarden

Co-authors: Debankur Mukherjee (TU Eindhoven), Sem Borst (TU Eindhoven, Nokia Bell Labs), Phil Whiting (Macquarie University)

Abstract: Load balancing involves a stream of incoming jobs that are to be dispatched to one of N servers. JSQ(d(N)) algorithms assign an arriving job to a server with the shortest queue among d(N) randomly selected servers. Using coupling techniques, mean-field and diffusion limits we present universality classes of algorithms that achieve full resource pooling or similar delay-communication tradeoffs. We discuss algorithms where d(N) grows with N, Join-the-Idle Queue (JIQ) algorithms and algorithms that operate on networks. Joint work with Debankur Mukherjee, Sem Borst and Phil Whiting.

Wednesday, 8:30 - 10:00, Room: 2130

Session: Strategic Interactions in Queues

Chair: Sandeep Juneja

Title: The Superposition-Traffic Game

Presenter: Harsha Honnappa

Co-authors:

Abstract: Strategic behavior in queueing models is often studied under the heavy-traffic assumption. In this paper, we attempt to answer the question of whether the heavy-traffic condition is necessary as well, under strategic considerations. In particular, we consider a situation where a finite number of traffic sources compete for service at a single server queue, by choosing traffic rates that maximizes their individual utility functions. The individual utilities trade off the mean delay experienced by the sources against a positive 'network effect' experienced due to a large number of sources requesting service. Our first result shows that there exists a generalized Nash equilibrium traffic rate vector, when the sources maximize their utilities. Our second result presents some comparative statics of the equilibria and show, intriguingly, that when when the Nash equilibrium traffic rates are non symmetric it is possible to increase throughput and reduce delays, compared to a symmetric Nash equilibrium. Finally, we show that the heavy-traffic condition emerges as a consequence of how the 'network effect' scales with the number of sources.

Title: Facilitating the search for partners on matching platforms: Restricting agent actions

Presenter: Yash Kanoria

Co-authors: Daniela Saban (Stanford)

Abstract: Two-sided matching platforms, such as those for labor, accommodation, dating, and taxi hailing, control many aspects of the search for partners. We consider a dynamic model of search by strategic agents with costly discovery of pair-specific match value, and find that in many settings, the platform can mitigate wasteful competition in partner search via restricting what agents can see/do. For medium-sized screening costs, the platform should allow one side of the market to choose unilaterally (similar to Instant Book on Airbnb), whereas for large screening costs, the platform should centrally determine matches (similar to taxi hailing marketplaces). Surprisingly, restrictions can improve social welfare even when screening costs are small. In asymmetric markets where agents on one side tend to be more selective, the platform should force the more selective side of the market to reach out first, by explicitly disallowing the less selective side from doing so. This allows the agents on the less selective side to exercise more choice in equilibrium.

Title: Rational Abandonment From Observable Priority Queues

Presenter: Philipp Afche

Co-authors: Vahid Sarhangian (Columbia / U of Toronto)

Abstract: The literature on customer behavior in queueing systems largely focuses on customers' joining decisions and ignores their subsequent abandonment decisions. Such abandonment behavior is important in priority queues, which are prevalent in practice. We characterize the equilibrium joining and abandonment behavior of utility-maximizing customers in an observable priority queue. We then discuss how the abandonment process in our equilibrium model com pares to that under the standard exogenous abandonment model, and to its empirical counterpart in a real system.

Title: Rest in the lounge or directly join the queue

Presenter: Sandeep Juneja

Co-authors: D Manjunath (IIT Bombay)

Abstract: We consider games where customers arrive at a lounge either as a one shot fixed batch or as a Poisson process. All customers wish to get serviced quickly, however resting in lounge is preferred to waiting in queue. The customers can see the numbers present in the lounge as well as the queue size at all times and need to dynamically decide when to join the queue. The service times are assumed to be exponentially distributed. In this setting we arrive at a symmetric Nash equilibrium dynamic policy for each customer for both types of arrival processes. We also derive bounds on price of anarchy and discuss its sensitivity to curtailing queue as well as lounge seats.

Wednesday, 8:30 - 10:00, Room: L-130

Session: Queueing Systems and Approximations

Chair: John Hasenbein

Title: Batch sojourn times in polling systems

Presenter: Ivo Adan

Co-authors: Jelmer van der Gaast (Erasmus University), Rene de Koster (Erasmus University)

Abstract: We consider a polling system where customers arrive in batches spread over multiple queues and study the batch sojourn-time, defined as the time from the batch arrival until service completion of the last customer in the batch. For various service disciplines we derive exact expressions for the Laplace-Stieltjes transform of the batch sojourn-time and also present a mean value approach to determine its mean. Numerical results show that the best performing service discipline, in terms of minimizing the mean batch sojourn-time, depends on system characteristics.

Title: A Multiclass Queue with State-dependent Arrival Rates

Presenter: John Hasenbein

Co-authors: Philip Ernst (Rice University), Soren Asmussen (Aarhus University)

Abstract: We study a multiclass queueing model in which the arrival rate depends on the job in service. Hence, there is an arrival rate matrix, instead of a rate vector. We provide results on stability and tail asymptotics. In addition, we make a connection between this queue and classical branching processes.

Title: Naor's Model with Heterogeneous Customers and Arrival Rate Uncertainty

Presenter: Chengcheng

Co-authors: John Hasenbein (UT Austin)

Abstract: This study examines extensions of Naor's queueing model in which the arrival rate is not known with certainty by either customers or system managers. In view of a social optimizer and a revenue maximizer, we investigate both static and dynamic pricing strategies in the presence of customer heterogeneity in their economic characteristics. The solutions for social optimal problem and revenue maximization problem are compared.

Title: Large Deviations Asymptotics for Brownian Queues

Presenter: Rob Wang

Co-authors: Peter Glynn (Stanford)

Abstract: In this talk, we discuss large deviations asymptotics for the departure process of a single-station Brownian queue. Both infinite and finite buffer contexts are considered. Our work is motivated by the fact that understanding rare-event behavior for departure processes is fundamental to understanding rare events in networked systems (since the endogenous arrivals to one station are departures from other stations). We extend existing literature on effective bandwidths by placing emphasis on the study of conditional queue dynamics given an unusual number of departures.

Wednesday, 8:30 - 10:00, Room: L-120

Session: Energy & Uncertainty

Chair: Adam Wierman

Title: Thinking fast and slow: Optimization decomposition across timescales

Presenter: Adam Wierman

Co-authors: Gautam Goel (California Institute of Technology), Niangjun Chen (California Institute of Technology), Adam Wierman (California Institute of Technology)

Abstract: Many real-world control systems, such as the smart grid, software defined networking, and human sensorimotor systems, have decentralized components that react quickly to stochastic noise using local information and centralized components that provide slow timescale planning using a more global view. Our results provide a theoretical framework for the design of such multi-timescale controllers using a novel form of optimization decomposition. We illustrate the approach using an example from power systems: decomposing generation planning into economic dispatch and frequency regulation.

Title: Optimal Control of Battery Storage under Cycle-Based Degradation Models

Presenter: Baosen Zhang

Co-authors: Yuanyuan Shi (UW), Bolun XU (UW), Daniel Kirschen (UW)

Abstract: Battery energy storage systems are becoming increasingly important for the operation of the power system. At the same time, these batteries have complex internal chemistries that control their degradation. Traditionally, these degradation were either ignored or greatly simplified in control and optimization problems. Even under simplified degradation models, optimal control of batteries is non-trivial, especially in online settings. In this talk, we present that a cycle-based cost model can capture a wide range of degradation behaviors. We show that this cost model is convex. Then we describe an approximately optimal online algorithm for fast regulation services.

Title: Learning to Price in Demand Response Programs

Presenter: Eilyan Bitar

Co-authors:

Abstract: Electric power utilities offer a diverse array of demand response (DR) programs where consumers are paid to curtail their electricity consumption during periods when electricity is scarce and expensive. There are several challenges a utility faces in implementing such programs, the most basic of which is the prediction of how customers will adjust their aggregate demand in response to different prices (or rebates). If the offered price is too low, consumers may be unwilling to curtail their demand; if the offered price is too high, the utility pays too much and gets more reduction than needed. More generally, the degree to which customers are willing to forego or shift their consumption, in exchange for monetary compensation, is contingent on variety of idiosyncratic and stochastic factors Ñ the majority of which are initially unknown or not directly measurable by the utility. The utility must therefore design its pricing policy to balance the tradeoff between the need to learn the unknown customer response model (exploration) and maximize its payoff (exploitation) over time. In this lecture, we will discuss the role that techniques from adaptive control might play in guiding the design of pricing polices that effectively balance this tradeoff.

Title: On wholesale electricity market design under uncertainty

Presenter: Subhonmesh Bose

Co-authors: Eilyan Bitar (Cornell), Khaled Alshehri (UIUC), Tamer Basar (UIUC)

Abstract: Wholesale electricity markets are networked marketplaces for trading in energy, mediated by a system operator. Current market design paradigms are not particularly suited to handle the deepening penetration of renewable supply. I will discuss the challenges in forward market designs that stem from the variability characteristics of renewables such as wind and solar resources. Then, I will present recent work on two possible directions to tackle the challenges: a contingent pricing approach and a centralized mechanism for trading cash-settled call options.

Wednesday, 8:30 - 10:00, Room: L-070

Session: Advances in Simulation and Stochastic Optimization

Chair: Henry Lam

Title: Dynamic Resource Provisioning in Data Centers under Demand Uncertainty

Presenter: Chaithanya Bandi

Co-authors:

Abstract: Data centers like Google and AWS face uncertain demand requirements that need to be satisfiedsequentially with the smallest possible cost of hardware and power. We propose a dynamic robustoptimization approach to model the uncertainty in the demand. Since, this dynamic multi-stageproblem is NP-hard, we consider affine policies approximation and surprisingly show that they are notnecessarily optimal even for simplex uncertainty sets in multi-stage unlike two-stage. We show alogarithmic bound approximation on the number of stages if the uncertainty sets are nested over timeand we give a general approximation bound for general uncertainty sets.

Title: Bayesian Optimization with Gradients

Presenter: Peter Frazier

Co-authors: Jian Wu (Cornell), Matthias Poloczek (Cornell), Andrew Gordon Wilson (Cornell)

Abstract: In recent years, Bayesian optimization has proven successful for global optimization of expensive-to-evaluate multimodal noisy objective functions. However, unlike most optimization methods, Bayesian optimization typically does not use derivative information. In this talk we show how Bayesian optimization can exploit derivative information to decrease the number of objective function evaluations required for good performance. In particular, we develop a novel Bayesian optimization algorithm, the derivative-enabled knowledge-gradient (dKG), for which we show one-step Bayes-optimality, asymptotic consistency, and greater one-step value of information than is possible in the derivative-free setting. Our procedure accommodates noisy and incomplete derivative information, and comes in both sequential and batch forms.

Title: Bayesian calibration of inexact computer models

Presenter: Matthew Plumlee

Co-authors:

Abstract: Bayesian calibration is used to study computer models in the presence of both a calibration parameter and model bias. The parameter in the predominant methodology is left undefined. This results in an issue where the posterior of the parameter is sub-optimally broad. There have been no generally accepted alternatives to date. This talk proposes and studies a relatively straightforward fix for Bayesian calibration where the prior distribution on the bias is made orthogonal to the gradient of the computer model. Problems associated with Bayesian calibration are shown to be mitigated through analytic results in addition to examples.

Title: Distributionally Robust Stochastic Optimization with Fixed Marginals

Presenter: Rui Gao

Co-authors: Anton Kleywegt (Georgia Tech)

Abstract: A central difficulty faced by decision-making under high-dimensional uncertainty is that the joint distribution of correlated random variables can hardly ever be estimated accurately, although estimates of the one-dimensional marginals often are relatively accurate. To deal with such problems, classical approaches find the worst-case distribution over all distributions with fixed marginals, but such approaches often lead to over-conservative decisions. We propose a distributionally robust, nonparametric approach, which hedges against a family of distributions with similar dependence structure and fixed marginals. Similarity of the dependence structure is captured through the Wasserstein distance to some nominal model, such as an empirical distribution or independent product distribution. Tractability of our new formulation is enhanced by a novel constructive proof of strong duality, combining ideas from the theory of optimal transport. Numerical experiments demonstrate how the proposed approach outperforms conventional approaches.

Wednesday, 8:30 - 10:00, Room: 2410

Session: Brownian Models and Stochastic Control Applications

Chair: Nur Sunar

Title: Managing production and buffer size with drift control

Presenter: Melda Ormeci Matoglu

Co-authors:

Abstract: We model the problem of managing capacity in a build-to-order environment as a Brownian drift control problem and seek a policy that minimizes the long-term average cost. We assume the controller can, at some cost, shift the processing rate by, for example, adding or removing staff, increasing or reducing the number of shifts or opening or closing production lines. The controller can also reject orders

when the order queue is too long or idle the system. We seek a policy that minimizes long-term average cost of control and holding cost. We show that a simple control band policy is optimal and determine its parameters.

Title: Stochastic Game of Investment in Common Goods

Presenter: Youngsoo Kim

Co-authors: George Georgiadis (Northwestern), H. Dharma Kwon (UIUC)

Abstract: Motivated by manufacturers' quality investment in shared suppliers, we study an investment game with spillover. Firms decide when to invest and restore the declining quality repeatedly. The resulting game is a repeated stochastic war of attrition. We characterize pure and mixed strategy equilibria, and discuss its managerial implications.

Title: Optimal Dynamic Product Development and Launch for a Network of Customers

Presenter: Nur Sunar

Co-authors: John Birge (University of Chicago)

Abstract: We consider a firm that dynamically chooses its effort to develop a product for a network of customers represented by a connected graph. The product's technological development is governed by a stochastic process dependent on the firm's efforts. In addition to dynamically choosing its development effort, the firm chooses when to launch or abandon the product. If the firm launches the product, the firm also chooses a selling price, a promotional price and a target customer to offer promotion. Once the target customer adopts the product, the product diffuses over the customer network based on the topology of the graph and the selling price. The product provides local network benefits to its adopters. The expected local network benefit of adoption is proportional to the number of neighbor customers that have already adopted the product. In a continuous-time setting, we explicitly solve the firm's jointly-optimal development, launch and post-launch strategies for any connected network, which required us to solve a stochastic control problem mixed with optimal stopping. We extend our analysis to consider demand heterogeneity in the customer network and multiple target customers.

Wednesday, 8:30 - 10:00, Room: 2420

Session: Learning and Applications

Chair: Bora Keskin

Title: Optimal A-B Testing

Presenter: Ciamac Moallemi

Co-authors: Nikhil Bhat (Columbia), Vivek Farias (MIT)

Abstract: We consider the problem of sequential A-B testing when the impact of a treatment is marred by a large number of covariates. Our main contribution is a tractable algorithm for the online allocation of test subjects to either treatment with the goal of maximizing the efficiency of the estimates treatment effect under a linear model, which due to a surprising state space collapse, reduces to solving a low dimensional dynamic program. Our approach is robust and covers many variations of the problem, including cases where there are budget constraints on individual treatments, where the number of trials is to be endogenously decided, and where the objective is to balance a tradeoff between efficiency and bias.

Title: Dynamic Learning and Pricing with Model Misspecification and Endogeneity Effect

Presenter: He Wang

Co-authors: Mila Nambiar (MIT), David Simchi-Levi (MIT)

Abstract: We study a dynamic pricing problem with contextual information where the seller may assume an incorrect demand model. The seller sequentially observes demand, estimates model parameters, and then chooses price. In this setting, model misspecification can cause price endogeneity, i.e., the correlation between price and the error term in the demand function. The endogeneity effect in turn leads to inconsistent estimation of price elasticity and bad pricing decisions, a phenomenon known as the "spiral-down effect" in revenue management. To address the endogeneity effect we propose a "Random Price Shock" (RPS) algorithm that dynamically generates independent price shocks to estimate price elasticity while maximizing revenue. We show that the RPS algorithm is simple to use, has strong theoretical performance guarantees, and is robust to model misspecification.

Title: Dynamic learning in the MNL-Bandit problem

Presenter: Vashist Avadhanula

Co-authors: Shipra Agrawal (Columbia), Vineet Goyal (Columbia), Assaf Zeevi (Columbia)

Abstract: We consider a dynamic assortment selection problem, where in every round the retailer offers a subset (assortment) of N substitutable products to a consumer, who selects one of these products according to a multinomial logit (MNL) choice model. The retailer observes this choice and the objective is to dynamically learn the model parameters, while optimizing cumulative revenues over a selling horizon of length T. We refer to this exploration-exploitation formulation as the MNL-Bandit problem. We present an efficient algorithm that simultaneously explores and exploits, achieving performance independent of the underlying parameters. The algorithm can be implemented in a fully online manner, without knowledge of the horizon length T. Furthermore, the algorithm is adaptive in the sense that its performance is near-optimal in both the 'well separated' case (where the gap between revenues corresponding to optimal assortment and sub-optimal alternatives is not small), as well as the general parameter setting where this separation need not hold.

Title: On Incomplete Learning and Certainty-Equivalence Control

Presenter: Bora Keskin

Co-authors: Assaf Zeevi (Columbia)

Abstract: Motivated by dynamic pricing applications, we consider a dynamic control-and-estimation problem. The decision-maker sequentially chooses controls and observes responses that depend on both the chosen controls and an unknown parameter. The decision-maker uses a certainty-equivalence policy, and we characterize the asymptotic accuracy performance of this policy.

Wednesday, 8:30 - 10:00, Room: 2430

Session: Matching Models

Chair: Jean Mairesse

Title: Advances on the stability problem of stochastic matching models on general graphs

Presenter: Pascal Moyal

Co-authors: Jean Mairesse (CNRS/UPMC), Ana Busic (INRIA), Ohad Perry (Northwestern University)

Abstract: We consider an extension of the bipartite matching model, in which the compatibility graph is general (i.e. non-necessarily bipartite) and the arrivals are simple, instead of pairwise. After providing a natural maximal stability region (Ncond), we illustrate the crucial influence of the matching policy on the stability of the model. We show, first, that aside for a particular class of graphs, there always exists a matching policy rendering the stability region strictly smaller than Ncond. Then, by adapting arguments developed for the Extended Bipartite matching model, we obtain an original product form Theorem, showing that the stability region of the policy First Come, First Matched is maximal. We conclude by proving a sub-additive Theorem, valid for most usual matching policies, that is the pilar of a Loynes-type coupling result.

Title: Fluid Models of Parallel Service Systems under FCFS

Presenter: Hanqin ZHANG

Co-authors: Yuval Nov (Department of Statistics, The University of Haifa, Israel), Gideon Weiss (Department of Statistics, The University of Haifa, Israel)

Abstract: We study deterministic fluid approximation models of parallel service systems operating under first come first served policy (FCFS) when the service time distributions may depend both on the server and

on the customer type. We explore the relations between fluid models and the three properties of stability, resource pooling, and matching rates. We find that stability and resource pooling are determined by the unique fluid model only in three cases: when service rates depend only on the server, when service rates depend only on the type of customer, and when the bipartite compatibility graph is a tree, a complete graph or a hybrid of the two. In general, when service rates depend on both server and customer type and the graph is not one of the above we find that the fluid model may not be unique, and stability and resource pooling cannot be determined from first moment information. Matching rates between pairs of compatible server and customer types cannot be determined from the fluid model, unless the compatibility graph is of one of the above forms. In particular, we discuss an example of Foss and Chernova (Queueing Systems, Vol29(1998), 55-73), and show by simulation that matching rates and stability depend on the service time distributions beyond first moments. Further simulations show that matching rates depend on distributions of service times even when service times depend only on the server type and the fluid model is unique. On the other hand, we solve a static planning linear program similar to Harrison and Lopez (Queueing Systems, Vol.33(1999), 339-368), and obtain a maximum throughput compatibility sub-graph that is a tree or a forest, and show that using only links of this sub-graph, FCFS is a throughput optimal policy.

Title: Reward Maximization in General Dynamic Matching Systems

Presenter: Alexander Stolyar

Co-authors: Mohammadreza Nazari (Lehigh University)

Abstract: A system with random arrivals of items of different types is considered. Items stay in the system until they are 'matched'. There is a finite number of 'matchings' which can be applied, each requires certain numbers of different-type items, and brings a certain 'reward'. Applications include assemble-to-order systems, Internet advertising, matching web portals, etc. We propose a dynamic matching scheme for maximizing the long-term average reward subject to the queues' stability, and prove its asymptotic optimality. The key element of the scheme is a virtual matching system, where the item queues are allowed to be negative. The scheme does not require a priori knowledge of the item arrival rates.

Title: Mean waiting times in systems with multi-type jobs and multi-type servers

Presenter: Ivo Adan

Co-authors: Marko Boon (Eindhoven Univeristy of Technology), Sigrid van Hoek (Eindhoven University of Technology)

Abstract: We study a parallel service queueing model of multi-type servers, serving multi-type customers under the policy FCFS-ALIS. In the exponential model an exact expression is available for the LST of the waiting time for each customer type. This expression has an interesting probabilistic interpretation. In this talk we explore approximations for the mean waiting time in the non-exponential system, by exploiting the exact exponential results.

Wednesday, 10:30 - 12:00, Room: 2110

Session: Stochastic Models in Inventory Management

Chair: Jiheng Zhang

Title: Optimal and Heuristic Ordering Policies for an Inventory System with Positive Leadtimes and an All-or-Nothing Yield Pattern

Presenter: Frank Chen

Co-authors: Fei Hu (CityU-HK), Candace Yano (Berkeley)

Abstract: We study a single-item, periodic-review inventory system with stochastic demand, positive delivery lead times and all-or-nothing random yields: whenever a shipment arrives, the firm takes a sample and conducts a quality test; if the sample(s) do not pass the test, the entire shipment is returned or destroyed. Although the problem is impossible to solve exactly except in special cases, we introduce the partial symmetry to the transformed model and then are able to derive properties of the optimal ordering policy. We use these properties in designing two heuristic policies. We also develop lower/upper bounds on the minimum cost and the optimal orders, which converge as the yield rate approaches 1. A numerical study illustrates the performance of the heuristics and bounds.

Title: Population Monotonicity in Newsvendor Games

Presenter: Zhenyu Hu

Co-authors: Xin Chen (UIUC), Qiong Wang (UIUC), Xiangyu Gao (UIUC)

Abstract: It is well-known that the core of the newsvendor game is non-empty and one can use duality theory in stochastic programming to construct an allocation belonging to the core, which we refer to asdual-based allocation scheme. In this work, we identify conditions under which the dual-based allocation scheme is a population monotonic allocation scheme (PMAS), which also requires eachplayer's cost decreases as the coalition to which she belongs grows larger. Specifically, we show that independent and log-concave demand is sufficient to guarantee this. In general, the dual-basedallocation scheme is a PMAS if the growth of the coalition does not increase the dependence structure between each player and the coalition.

Title: Management of a Shared Spectrum Network in Wireless Communications - A Queueing Approach

Presenter: Shining Wu

Co-authors: Jiheng Zhang (The Hong Kong University of Science and Technology), Rachel Q. Zhang (The Hong Kong University of Science and Technology)

Abstract: We consider a band of the electromagnetic spectrum with a finite number of identical channels shared by both licensed and unlicensed users. Such a network differs from most many-server, two-class queues in service systems including call centers due to the restrictions imposed on the unlicensed users in order to limit interference to the licensed users. We first approximate the key performance indicators, namely the throughput rate of the system and the delay probability of the licensed users under the asymptotic regime, which requires the analysis of both scaled and unscaled processes simultaneously using the averaging principle. Our analysis reveals a number of distinctive properties of the system. We then study the optimal sharing decisions of the system to maximize the system throughput rate while maintaining the delay probability of the licensed users below a certain level when the system is overloaded.

Title: NEAR OPTIMAL CONTROL FOR PERISHABLE INVENTORY

Presenter: Zhang Hailun

Co-authors: Jiheng Zhang (HKUST), Rachel Zhang (HKUST)

Abstract: We study joint replenishment and clearance of perishable products when the demand rate is large. We proposes two control policies. The fluid based policy can achieve asymptotic optimality with the gap explicitly computed. The policy based on news vendor problem can significantly improve the gap when the initial inventory is small. When the initial inventory is large, we propose an algorithm, which is much simpler than the original dynamic programming, to achieve asymptotic optimality with the improved gap. Numerical experiments shows that our proposed policy works well comparing to the true optimal policy by solving the stochastic dynamic programming.

Wednesday, 10:30 - 12:00, Room: 2110

Session: Stationarity in Stochastic Processes

Chair: Yanting Chen

Title: Constructions of Markov processes in random environments which lead to a product form of the stationary measure

Presenter: Anirban Das

Co-authors:

Abstract: Recently Belopolskaya and Suhov (2015) studied Markov processes in a random environment, where the environment changes in a Markovian manner. They introduced constructions allowing the process to interact with an environment. This was done in such a manner that the combined process has the product of the stationary measures of the individual processes as it's stationary measure . In my recent paper, a new construction is implemented, related to aproduct form for the stationary measure. This

construction can be carried out with almost no conditions. However it requires the use of an additional state, denoted by c. The extent to which the combined process uses state c indicates how far this process is from naturally having a product form for the stationary measure. The construction gives rise to a wide class of processes where the environment and system interact leading to a product form of the stationary measure.

Title: On the extinction of lower Hessenberg branching processes with countably many types

Presenter: Sophie Hautphenne

Co-authors: Peter Braunsteins (The University of Melbourne)

Abstract: We consider the extinction events of a subclass of branching processes with countably infinitely many types which we refer to as Lower Hessenberg branching processes (LHBPs). These are multitype Galton-Watson processes whose typeset corresponds to the nonnegative integers, in which individuals of type i cannot give birth to type j>i+1 offspring. For the class of LHBPs we completely characterise the set of fixed points of the progeny generating function, and we identify which elements in this set correspond to the partial and global extinction probability vectors. Under some additional conditions we then derive a global extinction criterion in the difficult case where there is almost sure partial extinction.

Title: Inhomogeneous perturbation and error bounds for the stationary performance of random walks in the quarter plane

Presenter: Xinwei Bai

Co-authors: Jasper Goseling (University of Twente)

Abstract: A continuous-time random walk in the quarter plane with homogeneous transition rates is considered. Given a non-negative reward function on the state space, we are interested in the expected stationary performance. Since a direct derivation of the stationary probability distribution is not available in general, the performance is approximated by a perturbed random walk, whose transition rates on the boundaries are changed such that its stationary probability distribution is known in closed form. A perturbed random walk for which the stationary distribution is a sum of geometric terms is considered and the perturbed transition rates are allowed to be inhomogeneous. It is demonstrated that such rates can be constructed for any sum of geometric terms that satisfies the balance equations in the interior of the state space. The inhomogeneous transitions relax the pairwise-coupled structure on these geometric terms that would be imposed if only homogeneous transitions are used.An explicit expression for the approximation error bound is obtained using the Markov reward approach, which does not depend on the values of the inhomogeneous rates but only on the parameters of the geometric terms. Numerical experiments indicate that inhomogeneous perturbation can give smaller error bounds than homogeneous perturbation.

Title: Invariant measures and error bounds for random walks in the quarter-plane based on sums of geometric terms

Presenter: Yanting Chen

Co-authors: Richard J. Boucherie (University of Twente), Jasper Goseling (University of Twente)

Abstract: We consider homogeneous random walks in the quarter-plane. The necessary conditions which characterize random walks of which the invariant measure is a sum of geometric terms are provided in Chen et al. (arXiv:1304.3316, 2013, Probab Eng Informational Sci 29(02):233-251, 2015). Based on these results, we first develop an algorithm to check whether the invariant measure of a given random walk is a sum of geometric terms. We also provide the explicit form of the invariant measure if it is a sum of geometric terms. Second, for random walks of which the invariant measure is not a sum of geometric terms, we provide an approximation scheme to obtain error bounds for the performance measures. Our results can be applied to the analysis of two-node queueing systems. We demonstrate this by applying our results to a tandem queue with server slow-down.

Wednesday, 10:30 - 12:00, Room: 2130

Session: Statistics of Queues

Chair: Yoni Nazarathy

Title: Inferring queue transition dynamics by boosting hazard regression with time-varying covariates

Presenter: Donald Lee

Co-authors: Ningyuan Chen (HKUST)

Abstract: The customer transition dynamics for a variety of queuing networks can be modelled as a failure process. In order to better understand these dynamics from a data-driven viewpoint, we propose a practical gradient boosting procedure for flexibly estimating the transition intensities. Our procedure comes with statistical oracle guarantees when flexible regression tree-based models are used. We apply it to shed new light on an existing empirical operations question regarding the impact of staff workload on service rates in an emergency department.

Title: Impact of Callers' History on Abandonment: Model and Implications

Presenter: Seyed Emadi

Co-authors: Jayashankar Swaminathan (UNC Kenan-Flagler Business School)

Abstract: Caller abandonment could depend on their past waiting experiences. Using Cox regressions we show that callers who abandoned or waited for a shorter time in the past abandon more in the future. However, Cox regression approach does not shed light on callers' prior belief about the duration of their delays. Moreover, Cox regressions cannot separate the impact of callers' parameters such as their waiting

costs on their abandonment behavior from the impact of their beliefs about their delay durations, which are affected by their past waiting experiences. To tease out the impact of callers' waiting experiences on their abandonment behavior, we use a structural estimation approach in a Bayesian learning framework. We estimate the parameters of this model from a call center data set with multiple priority classes. We show that in this call center new callers who do not have any experience with the call center are optimistic about their delay in the system and underestimate its length irrespective of their priority class. We also show that our bayesian learning model not only has a better fit to the data set compared to the rational expectation model in Aksin et al. (2013), Aksin et al. (2016) and Yu et al. (2016) but also outperforms the rational expectation model in out-of-sample tests. In addition, our bayesian framework does not lead to biased estimates, which would happen under the rational expectation assumption if callers' belief about their waiting durations does not match their actual waiting time distribution. Our bayesian framework has managerial implications at both tactical and operational levels such as managing customer expectation about their delays in the system, and implementation of patience-based priority policies such as Least-Patience-First and Most-Patience-First scheduling.

Title: Queueing estimation insights from 200 papers applied to 200,000 patient journey observations.

Presenter: Yoni Nazarathy

Co-authors:

Abstract: We use a dataset describing patient journeys for 200,000 patients in an Australian hospital, recorded over 4 years. Armed with queueing theory we attempt to use insights from the theory for describing phenomena in the data. Specifically we consider about 200 research papers dealing with parameter and state estimation of queues. For each paper, we attempt to classify the usefulness of the results and methods for the dataset under study.

Wednesday, 10:30 - 12:00, Room: 2410

Session: Computational Methods for Markov Decision Processes

Chair: Eugene Feinberg

Title: Easy Affine MDPs with Generalized Decomposability

Presenter: Matthew J. Sobel

Co-authors: Jie Ning (Case Western Reserve)

Abstract: Markov decision processes with vector-valued continuous states and actions are generally impractical to solve numerically except with drastic discretization which is subject to the curse of dimensionality. Easy affine MDPs are a subclass with affine rewards and dynamics and decomposable

constraints. Such MDPs with discounted criteria can be solved exactly and easily via auxiliary equations. Thus, many applications have become accessible. This talk is an overview of easy affine MDPs accompanied by a weakening of the decomposability assumption.

Title: On the reduction of total cost and average cost MDPs to discounted MDPs

Presenter: Jefferson Huang

Co-authors: Eugene Feinberg

Abstract: We provide conditions under which total-cost and average-cost Markov decision processes (MDPs) can be reduced to discounted ones. Results are given for transient total-cost MDPs with transition rates whose values may be greater than one, as well as for average-cost MDPs with transition probabilities satisfying the condition that there is a state such that the expected time to reach it is uniformly bounded for all initial states and stationary policies. In particular, these reductions imply sufficient conditions for the validity of optimality equations and the existence of stationary optimal policies for MDPs with undiscounted total cost and average-cost criteria. When the state and action sets are finite, these reductions lead to linear programming formulations and complexity estimates for MDPs under the aforementioned criteria.

Title: Censored Markov Chains and Recursive Algorithms

Presenter: Isaac M Sonin

Co-authors:

Abstract: An important, though not well-known tool for the study of Markov chains (MCs) is the notion of a Censored (Embedded) MC. It is based on a simple and insightful idea of Kolmogorov and Doeblin: a MC observed only on a subset of its state space is again a MC with a reduced state space and a new transition matrix. The sequential application of this idea leads to an amazing variety of important algorithms in Probability Theory and its Applications. Three of them will be discussed: the State Elimination algorithm for the problem of optimal stopping of MC, an algorithm to calculate the well-known Gittins Index and the Generalized Gittins Index, and a polynomial algorithm to calculate a crucial characteristic in the MC Tree theorem.

Title: Stochastic switching for partially observable dynamics

Presenter: Juri Hinz

Co-authors: Yes

Abstract: In industrial applications, optimal control problems frequently appear in the context of decision-making under incomplete information. In such framework, decisions must be adapted dynamically to account for possible regime changes of the underlying dynamics. Using stochastic filtering theory, Markovian evolution can be modeled in terms of latent variables, which naturally leads to high-dimensional state space, making practical solutions to these control problems notoriously challenging. In our talk, we

utilize a specific structure of this problem class to present a solution in terms of simple, reliable, and fast algorithms. The algorithms presented in this paper have already been implemented in an R package.

Wednesday, 10:30 - 12:00, Room: 2420

Session: New Directions in Learning

Chair: Yonatan Gur

Title: Learning in Repeated Auctions with Budgets: Regret Minimization and Equilibrium

Presenter: Yonatan Gur

Co-authors: Santiago R. Balseiro (Duke)

Abstract: In online advertising markets, advertisers often purchase ad placements through bidding in repeated auctions based on realized viewer information. We study how budget-constrained advertisers may bid in the presence of competition, when there is uncertainty about future bidding opportunities as well as competitors' heterogeneous preferences and budgets. We formulate this problem as a sequential game of incomplete information, where bidders know neither their own valuation distribution, nor the budgets and valuation distributions of their competitors. We introduce a family of dynamic bidding strategies we refer to as adaptive pacing strategies, in which advertisers adjust their bids throughout the campaign according to the sample path of observed expenditures. We analyze the performance of this class of strategies under different assumptions on competitors' behavior. Under arbitrary competitors' bids, we establish through matching lower and upper bounds the asymptotic optimality of this class of strategies as the number of auctions grows large. When adopted by all the bidders, the dynamics converge to a tractable and meaningful steady state. Moreover, we show that these strategies constitute an approximate Nash equilibrium in dynamic strategies: The benefit of unilaterally deviating to other strategies, including ones with access to complete information, becomes negligible as the number of auctions and competitors grows large. This establishes a connection between regret minimization and market stability, by which advertisers can essentially follow equilibrium bidding strategies that also ensure the best performance that can be guaranteed off-equilibrium.

Title: Online learning in repeated auctions

Presenter: Jonathan Weed

Co-authors: Vianney Perchet (ENS Paris-Saclay), Philippe Rigollet (MIT)

Abstract: Motivated by online advertising auctions, we consider repeated Vickrey auctions where goods of unknown value are sold sequentially and bidders only learn (potentially noisy) information about a good's value once it is purchased. We adopt an online learning approach with bandit feedback to model this

problem and derive bidding strategies for two models: stochastic and adversarial. In the stochastic model, the observed values of the goods are random variables centered around the true value of the good. In this case, logarithmic regret is achievable when competing against well behaved adversaries. In the adversarial model, the goods need not be identical. Comparing our performance against that of the best fixed bid in hindsight, we show that sublinear regret is also achievable in this case. For both the stochastic and adversarial models, we prove matching minimax lower bounds showing our strategies to be optimal up to lower-order terms. To our knowledge, this is the first complete set of strategies for bidders participating in auctions of this type.

Title: Deep Exploration in Reinforcement Learning via Randomized Value Functions

Presenter: Daniel Russo

Co-authors: Ian Osband (DeepMind), Zheng Wen (Adobe Research), Benjamin Van Roy (Stanford)

Abstract: The field of reinforcement learning develops algorithms that learn to optimize performance in complicated MDPs by using observed rewards and state-transitions. A major challenge in the field is the ability to collect the right training data, and even important breakthroughs have still relied on completely random exploration, or the imitation of human experts who have mastered the control task.This work studies the use of randomized value functions to guide deep exploration in reinforcement learning. This offers an elegant means for synthesizing statistically and computationally efficient exploration with common practical approaches to value function learning. We present several reinforcement learning algorithms that leverage randomized value functions and demonstrate their efficacy through computational studies. We also prove a regret bound that establishes statistical efficiency with a tabular representation

Title: The value of foresight

Presenter: Quan Zhou

Co-authors: P.A. Ernst (Rice University), L.C.G. Rogers (University of Cambridge)

Abstract: Suppose you have one unit of stock, currently worth 1, which you must sell before time T. The Optional Sampling Theorem tells us that whatever stopping time we choose to sell, the expected discounted value we get when we sell will be 1. Suppose however that we are able to see a units of time into the future, and base our stopping rule on that; we should be able to do better than expected value 1. But how much better can we do? And how would we exploit the additional information? The optimal solution to this problem will never be found, but in this paper we establish remarkably close bounds on the value of the problem, and we derive a fairly simple exercise rule that manages to extract most of the value of foresight.

Wednesday, 10:30 - 12:00, Room: 2430

Session: Mean Field Limits and Their Applications

Chair: Rami Atar and Asaf Cohen

Title: A Large Scale Analysis of Unreliable Stochastic Networks

Presenter: Philippe Robert

Co-authors: Reza Aghajani (UCSD) Wen Sun (INRIA France)

Abstract: The problem of reliability of a large distributed system is analyzed via a new mathematical model. A typical framework is a system where a set of files are duplicated on several data servers. When one of these servers breaks down, all copies of files stored on it are lost. They can be retrieved afterwards if copies of the same files are stored on some other servers. In the case where no other copy of a given file is present in the network, it is definitively lost. The efficiency of such a network is therefore directly related to the performances of the mechanism used to duplicate files on servers. We study the asymptotic behavior of this system in a mean-field context, i.e. when the number N of servers is large. The analysis is complicated by the large dimension of the state space of the empirical distribution of the state of the network. We introduce a stochastic model of the evolution of the network which has values in state space whose dimension does not depend on N. This description does not have the Markov property but it turns out that it is converging in distribution, as N gets large, to a nonlinear Markov process. Additionally, this asymptotic process gives a limiting result on the rate of decay of the network which is the key characteristic of interest of these systems. Convergence results are established and we derive a lower bound on the exponential decay, with respect to time, of the fraction of the number of initial files with at least one copy. Stochastic calculus with marked Poisson processes, technical estimates and mean-field results are the main ingredients of the proofs of the results.

Title: Mean-field approximation of large banking system with defaults

Presenter: Tomoyuki Ichiba

Co-authors: Romuald Elie (Universite Paris-Est Marne-la-Vallee) and Mathieu Lauriere (NYU Shanghai)

Abstract: In this paper, we consider the dynamics of the cash reserves of an interconnected banking system. Whenever a bank defaults (by letting its reserve reach a given threshold), each bank is impacted negatively, via an instantaneous jump on its reserve level. We also take into account the arrival of new financial institutions in the system. The underlying dynamics of such system is written in the spirit of the spiking neural network models as studied by Delarue, Inglis, Rubenthaler and Tanr\'e. We study the mean field limit of such system and focus in particular on its stationary distribution. Following the approach of Fouque, Carmona and Sun, we try to identify such dynamics as the equilibrium of a mean field game between banks in interaction. This is a joint work with R. Elie and M. Lauriere.

Title: Mean Field Equilibria of Pricing and Work-Quality Selection Games in Internet Marketplaces

Presenter: Vijay Subramanian

Co-authors: V. R. Raja (TAMU) and S. Shakkottai (TAMU)

Abstract: We model an Internet marketplace using a set of servers that choose prices for performing jobs. Each server has a queue of unfinished jobs, and is penalized for delay by the market maker. Each server is allowed to choose the work-quality when performing a job, with a job done at a higher work-quality than its inherent value incurring a cost and jobs truthfully report the "quality" with which they were completed. The best estimate of quality based on these reports is the "reputation" of the server. A server bases its pricing decision on the distribution of its competitors prices and reputations. An entering job chooses the best server based on a combination of price and reputation. We seek to understand how prices would be determined in such a marketplace using Mean Field Equilibrium. We show the existence of a MFE and impact of reputation in allowing servers to declare larger prices than their competitors. We also illustrate our results by a numerical study of an existing marketplaces.

Title: Rate Control under Heavy Traffic with Strategic Servers

Presenter: Asaf Cohen

Co-authors: Erhan Bayraktar (Michigan), Amarjit Budhiraja (Chapel-Hill)

Abstract: We consider a large queueing system that consists of many strategic servers that are weakly interacting. Each server processes jobs from its unique critically loaded buffer and controls the rate of arrivals and departures associated with its queue to minimize its expected cost. The rates and the cost functions in addition to depending on the control action, can depend, in a symmetric fashion, on the size of the individual queue and the empirical measure of the states of all queues in the system. In order to determine an approximate Nash equilibrium for this finite player game we construct a Lasry-Lions type mean-field game (MFG) for certain reflected diffusions that governs the limiting behavior. Under conditions, we establish the convergence of the Nash-equilibrium value for the finite size queuing system to the value of the MFG. In general closed form solutions of such MFG are not available and thus numerical methods are needed. We use the Markov chain approximation method to construct approximations for the solution of the MFG and establish convergence of the numerical scheme.

Wednesday, 1:30 - 3:00, Room: 2110

Session: Stochastic Models for Service Management and Control

Chair: Zeynep Aksin Karaesmen

Title: Delay Announcements in Service Systems with Customer Priorities

Presenter: Rouba Ibrahim

Co-authors: A. Bassamboo (Northwestern)

Abstract: We investigate the accuracy of announcing the waiting time of the Last customer to Enter Service (LES) in the context of a queueing model with multiple customer classes and a priority service discipline. We present ways of exploiting this historical information to design new and improved announcements and supplement our theoretical results with an extensive simulation study to generate practical managerial insights.

Title: Front-office multitasking between service encounters and back-office tasks

Presenter: Zeynep Aksin

Co-authors: Benjamin Legros (PSB Paris School of Business), Oualid Jouini (CentraleSupelec Universite Paris-Saclay), Ger Koole (VU University Amsterdam)

Abstract: We model the work of a front-line service worker, who interacts with customers in a multi-stage process. Some stages of this service encounter require an interaction between server and customer, while other stages are performed by the customer as a self service task or with the help of another resource. In addition to customer interactions, the server needs to deal with back-office tasks, or tasks that do not require interaction with the customer. The latter tasks are of lower priority. The server's work is represented by a queue with high priority tasks, and an infinitely backlogged amount of low priority tasks. The server needs a switching time when switching between the two types of tasks. The server can treat back-office tasks during the interludes of a service encounter or between successive encounters. The objective is to maximize the expected proportion of time spent on low priority tasks subject to a constraint on the high priority task waiting time. Hence, a good tradeoff has to be found between two conflicting performance measures in a context where switching times may discourage frequent changes. Under certain parameter values, working on the back-office tasks during interludes is found to be valuable. We find that switching times between tasks are best controlled by a queue length dependent threshold type policy during breaks, and by a static service probability during interludes.

Title: A Model of Managing Chronic Care with Patient Activation Measure (PAM)

Presenter: Odysseas Kanavetas

Co-authors: Evrim Gunes (Koc University), Lerzan Ormeci (Koc University)

Abstract: We develop a MDP framework to manage care for patients with multiple chronic conditions via a complex care hub. Complex care provision influences the evolution of PAM, an indicator for healthy behavior, which affects the evolution of health state of patients. We explore optimal policies to minimise healthcare costs.

Title: The effect of customer heterogeneity in transportation stations

Presenter: Athanasia Manou

Co-authors: Fikri Karaesmen (Koc University), Pelin G. Canbolat (Koc University)

Abstract: We consider a transportation station, where customers arrive according to a Poisson process. A transportation facility with unlimited capacity visits the station according to a renewal process and at each visit it serves all presentcustomers. We assume that the arriving customers decide to use the transportation facility or not. A customer who chooses not to use the facility earns no rewards and incurs no costs. A customer who chooses to use it earns a reward upon service completion, pays a service fee, and incurs a waiting cost. Customers differ in their sensitivity in delays and reward from service. This situation can be considered as a game among heterogeneous customers. We study this game under two different levels of information provided to the customers upon arrival. We obtain the equilibrium customer behavior and the utilities of the customers and the administrator of the system under equilibrium. Finally, we explore the effect of heterogeneity on customer behavior and the utilities of the customers and the administrator.

Wednesday, 1:30 - 3:00, Room: 2120

Session: Heavy-Traffic Analysis and Control

Chair: Kavita Ramanan

Title: Heavy-traffic approximations for a layered network with limited resources

Presenter: Maria Vlasiou

Co-authors: A. Aveklouris (TU/e), J. Zhang (HKUST), B. Zwart (CWI)

Abstract: Motivated by a web-server model, we present a queuing network consisting of two layers. Thefirst layer incorporates the arrival of customers at a network of two FCFS single-server nodes. We assume general distributions and Markovian routing. At the second layer, active servers act as jobs that are served by a single LPS server. Our main result is a diffusion approximation for the process describing the number of customers in the system.

Title: Equilibrium behavior of randomized load balancing algorithms

Presenter: Pooja Agarwal

Co-authors: Kavita Ramanan (Brown)

Abstract: Randomized load-balancing algorithms play an important role in large-scale networks. We consider a network of N parallel queues in which incoming jobs that have an iid general service distribution with a density and finite mean are routed on arrival using the join-the-shortest-of-d-queues routing algorithm. We show that, under subcriticality, the hydrodynamic limit of the network has a unique equilibrium point. We also discuss convergence of the stationary distributions to this equilibrium point for

various classes of service distributions, thus complementing results on tail behavior for power-law distributions by Bramson-Lu-Prabhakar. The proofs entail the analysis of a coupled system of deterministic measure-valued equations, which may be of independent interest.

Title: Analysis of Processor Sharing Queues via Relative Entropy

Presenter: Amber L. Puha

Co-authors: Ruth J. Williams (University of California San Diego)

Abstract: Processor sharing is a mathematical idealization of round-robin scheduling algorithms commonly used in computer time-sharing. ÊIt is a fundamental example of a non-head-of-the-line service discipline. ÊFor such disciplines, it is typical that any Markov description of the system state is infinite dimensional. Due to this, measure-valued stochastic processes are becoming a key tool used in the modeling and analysis of stochastic network models operating under various non-head-of-the-line service disciplines.In this talk, we discuss a new approach to studying the asymptotic behavior of fluid model solutions (formal functional law of large numbers limits) for critically loaded processor sharing queues. ÊFor this, we introduce a notion of relative entropy associated with measure-valued fluid model solutions. ÊThis approach is developed with idea that similar notions involving relative entropy may be helpful for understanding the asymptotic behavior of critical fluid model solutions for stochastic networks operating under protocols naturally described by measure valued processes.

Wednesday, 1:30 - 3:00, Room: 2130

Session: Strategic Customers in Service Operations

Chair: Yehua Wei

Title: On the Efficacy of Static Prices for Revenue Management in the Face of Strategic Customers

Presenter: Yiwei Chen

Co-authors: Vivek Farias (MIT)

Abstract: We consider a canonical revenue management problem wherein a monopolist seller seeks to maximize revenues from selling a fixed inventory of a product to customers who arrive over time. We assume that customers are forward looking and strategize on the timing of their purchase, an empirically confirmed aspect of modern customer behavior. In the event that customers were myopic, foundational work by Gallego and Van Ryzin [1994] established that static prices were asymptotically optimal for this problem. In stark contrast, for the case where customers are forward looking, available results in mechanism design and dynamic pricing offer no such simple solution and are also constrained by restrictive assumptions on customer type. We study the revenue management problem while assuming forward

looking customers. We demonstrate that for a broad class of customer utility models, static prices surprisingly continue to remain asymptotically optimal in the regime where inventory and demand grow large. We further show that irrespective of regime, an optimally set static price guarantees the seller revenues that are within at least 63.2% of that under an optimal dynamic mechanism. The class of customer utility models we consider is parsimonious and enjoys empirical support. It also subsumes many of the utility models considered for this problem in existing mechanism design research; we allow for multi-dimensional customer types. We also allow for a customer's disutility from waiting to be positively correlated with his valuation. Our conclusions are thus robust and provide a simple solution to what is considered a challenging problem of dynamic mechanism design.

Title: Multi-agent Mechanism Design without Money

Presenter: Santiago Balseiro

Co-authors: Huseyin Gurkan (Duke University) and Peng Sun (Duke University)

Abstract: We consider a principal repeatedly allocating a single resource in each period to one of multiple agents without relying on monetary payments over an infinite horizon. Agents' private values are independent and identically distributed. We show that as the discount factor approaches one, the optimal dynamic mechanism without money achieves the first-best efficient allocation (the welfare-maximizing allocation as if values are public). As part of the proof, we provide an incentive compatible dynamic mechanism that achieves asymptotic optimality.

Title: Pricing In A Two-sided Market With Time-sensitive Customers And Suppliers

Presenter: Philipp Afche

Co-authors: Mustafa Akan (Carnegie Mellon)

Abstract: We consider a firm that matches stochastically arriving and time-sensitive customers and suppliers. We characterize the structure and performance of the profit-maximizing and socially optimal pricing policies.

Title: Tight Competitive Ratios for Online Matching/Assortment Problems with a Fixed Set of Edge-weights/Prices

Presenter: Will Ma

Co-authors:

Abstract: Online bipartite matching, introduced by Karp, Vazirani, and Vazirani in 1990, is a classical problem in the study of online algorithms and competitive analysis. It has since had many generalizations, including online vertex-weighted matching, Adwords, and online assortment, which have found application in internet advertising, personalized e-commerce, and ride-sharing. These problems can be abstracted as follows: there are fixed resources, which must be allocated on-the-fly, without assuming anything about

future demand. Two types of algorithms---"ranking" and "balance"---have been developed for these problems, and achieve a tight competitive ratio of 1-1/e under the integral and fractional, asymptotic settings, respectively. A key assumption in the aforementioned problems is that each resource is sold at a fixed rate when it is allocated. In this paper, we study these problems when each resource could be sold at multiple, known prices. We derive a tradeoff function using the idea of protection levels for different prices, which enables us to generalize the "ranking" and "balance" algorithms. Furthermore, we construct a family of examples which shows that this results in the optimal competitive ratio, for every possible set of prices. As a concrete example of our results, if each resource has two potential prices, then 1-1/sqrt(e) is the tight competitive ratio. Our analysis also provides a system to obtain competitive ratio bounds in the fractional, non-asymptotic setting, which improves existing bounds in the single-price as well.

Wednesday, 1:30 - 3:00, Room: L-130

Session: Design and Optimal Control of Queues

Chair: Yunan Liu

Title: Data-Driven Control of Queueing Networks Using the P-Model

Presenter: Shuangchi He

Co-authors: Melvyn Sim (NUS), Meilin Zhang (NUS), Shasha Han (NUS)

Abstract: We study customer routing and sequencing problems in queueing networks, which are motivated by applications in healthcare systems such as patient scheduling in emergency departments and inpatient bed assignment in hospital wards. These problems are usually subject to constraints on customer waiting times or throughput times. We propose an approach based on the P-model first studied by Charnes and Cooper (1963), in order to obtain near-optimal control policies. In numerical experiments, the proposed P-model approach outperforms an asymptotically optimal scheduling policy.

Title: Incentive Based Service System Design: Staffing and Compensation to Trade Off Speed and Quality

Presenter: Dongyuan Zhan

Co-authors: Amy R. Ward (USC)

Abstract: Most common queueing models used for service system design assume the servers work at fixed (possibly heterogeneous) rates. However, real-life service systems are staffed by people, and people may change their service speed in response to their compensation incentives. The delicacy is that the resulting employee service rate affects the staffing, but also the staffing affects the resulting employee service rate. Our objective in this paper is to find a joint staffing and compensation policy that induces optimal service system performance.We do this under the assumption that there is a trade-off between

service speed and quality, and employees are paid based on both. The employees each selfishly choose their own service speed in order to maximize their own expected utility (which depends on the staffing through their busy time). We prove the existence of an equilibrium service speed under a simple piece-rate compensation policy, and show the convergence to a unique limit as the customer arrival rate becomes large. The endogeneous service rate assumption leads to a centralized control problem in which the system manager jointly optimizes over the staffing and service rate. That centralized control problem (solved under fluid scaling) leads to conditions on the system manager's cost function under which a critically loaded, efficiency-driven, quality-driven, or intentional idling regime - in which there is simultaneous customer abandonment and server idling - are economically optimal operating regimes. Finally, noting that exact first best cannot be achieved without uniqueness (or an added equilibrium selection criteria), we provide a limiting first best policy.

Title: Asymptotically optimal control for crisscross networks

Presenter: Xin Liu

Co-authors: Amarjit Budhiraja (UNC-Chapel Hill), Subhamay Saha (IIT, Guwahati)

Abstract: A crisscross network consists of two stations. There are two classes of jobs arriving from outside to Station 1. Class 1 jobs leave the system once their service is completed, and Class 2 jobs after being served at Station 1 proceed to Station 2, where they are re-designated as Class 3 jobs and get served at Station 2. We study a scheduling control problem for crisscross networks, using formal diffusion approximations under suitable temporal and spatial scaling known as Brownian control problems (BCP) and their equivalent workload formulations (EWF). In the regime considered here, the singular control problem corresponding to the EWF does not have a simple form explicit solution. However, by considering an associated free boundary problem one can give a representation for an optimal controlled process as a two dimensional reflected Brownian motion in a Lipschitz domain whose boundary is determined by the solution of the free boundary problem. Using the form of the optimal solution we propose a sequence of control policies, given in terms of suitable thresholds, for the scaled stochastic network control problems and prove that this sequence of policies is asymptotically optimal. As suggested by the solution of the EWF, the policy we propose requires a server to idle under certain conditions which are specified in terms of thresholds determined from the free boundary.

Title: Simple and explicit bounds for multi-server queues with universal 1/(1 - rho) scaling

Presenter: David Goldberg

Co-authors: Yuan Li (Georgia Tech)

Abstract: We consider the FCFS GI/GI/n queue, and prove the first simple and explicit bounds that scale gracefully and universally as 1/(1-rho), with rho the corresponding traffic intensity. Our main results are bounds for the tail of the steady-state queue length and the steady-state probability of delay, where the strength of our bounds (e.g. in the form of tail decay rate) is a function of how many moments of the inter-arrival and service distributions are assumed finite. In contrast to all existing bounds in the literature, our

simple and explicit bounds scale gracefully even when the number of servers grows large and the traffic intensity converges to unity simultaneously, as in the Halfin-Whitt scaling regime.

Wednesday, 1:30 - 3:00, Room: L-120

Session: Performance and Optimization of Power Systems

Chair: Maria Vlasiou

Title: Markov-modulated models for electricity pricing

Presenter: Giang Nguyen

Co-authors: Nigel Bean (The University of Adelaide), Angus Lewis (The University of Adelaide)

Abstract: Markov-modulated (or regime-switching) models are often used for modelling electricity price, as they can capture both the sporadic spikes (caused by supply shortage due to various reasons) and the mean-reverting behaviour during normal times. In this talk, we use Markov-modulated models to analyse electricity price in Australian markets, and compare procedures for estimating the model parameters.

Title: Electric vehicle charging - a queueing approach

Presenter: Angelos Aveklouris

Co-authors: Yorie Nakahira (Caltech), Maria Vlasiou (Eindhoven University of Technology), Bert Zwart (Centrum Wiskunde and Informatica, Eindhoven University of Technology)

Abstract: The number of electric vehicles is expected to increase. As a consequence, more vehicles will need charging, potentially causing congestion in the power grid.Motivated by this, we consider a parking lot with finitely many positions, in which electric vehicles arrive in order to get charged. A vehicle has a random parking time and a random charging time. Furthermore, the total capacity of energy is limited. Thus, the charging rate that a vehicle receives is also limited. We are interested in finding the fraction of vehicles that get fully charged, which gives the probability a vehicle leaves the parking lot with fully charged battery. We develop several bounds and asymptotic approximations for this performance measure and compare these results with numerical outcomes.

Title: Opportunities for Price Manipulation by Aggregators in Electricity Markets

Presenter: Navid Azizan-Ruhi

Co-authors: Krishnamurthy Dvijotham (Caltech), Niangjun Chen (Caltech), Adam Wierman (Caltech)

Abstract: Aggregators of distributed generation are playing an increasingly crucial role in the integration of renewable energy in power systems. However, the intermittent nature of renewable generation makes market interactions of aggregators difficult to monitor and regulate, raising concerns about potential market manipulation by aggregators. In this paper, we study this issue by quantifying the profit an aggregator can obtain through strategic curtailment of generation in an electricity market. We show that, while the problem of maximizing the benefit from curtailment is hard in general, efficient algorithms exist when the topology of the network is radial (acyclic). Further, we highlight that significant increases in profit are possible via strategic curtailment in practical settings.

Title: Power-law of cascading failures in power systems

Presenter: Fiona Sloothaak

Co-authors: Sem Borst (Eindhoven University of Technology, Nokia Bell Labs), Vitali Wachel (Augsburg University), Bert Zwart (CWI, Eindhoven University of Technology)

Abstract: Cascading failure models are used to describe systems of interconnected components where failures possibly trigger subsequent failures of other components. Despite the deceptively simple appearance of these models, they capture an extraordinary richness of different behaviors and have therefore proven to be effective in a wide range of practical applications. Our inspiration is drawn from power outages in electric transmission systems. As these networks continue to increase in complexity and volatility, a fundamental understanding of the risks of cascading failures becomes of critical importance to guarantee and maintain a high reliability. In this talk, we consider large-scale systems where small disruptions in the load distribution potentially lead to severe reliability issues through a cascading failure mechanism. We particularly explore settings under which the tail of the number of failures exhibits power-law behavior, as is commonly encountered in empirical data analysis of blackout sizes. Exploiting an equivalent critical random-walk representation, we describe a framework where the power-law can be identified.

Wednesday, 1:30 - 3:00, Room: L-070

Session: Stochastic Optimization and Simulation

Chair: Susan Hunter

Title: The epsilon-constraint method for integer-ordered bi-objective simulation optimization

Presenter: Kyle Cooper

Co-authors: Susan R. Hunter (Purdue), Kalyani Nagaraj (Oklahoma State)

Abstract: Consider the context of integer-ordered bi-objective simulation optimization, in which the feasible region is a subset of the integer lattice. We propose a framework to identify the Pareto set that involves solving a sequence of stochastically constrained problems (via the epsilon-constraint method) and that is designed for deployment on a parallel computing platform. We discuss the design principles that make our framework efficient.

Title: Probability Ratio Testing for Multiple-Objective Ranking and Selection

Presenter: Wenyu Wang

Co-authors: Hong Wan

Abstract: In this paper, we introduce a sequential procedure for the Multi-Objective Ranking and Selection (MOR&S) problems that identifies the Pareto front with a guaranteed probability of correct selection (PCS). In particular, the proposed procedure is fully sequential using the test statistics built upon the generalized sequential probability ratio test (GSPRT). The main features of the new proposed procedure are: 1) a unified framework, the new procedure treats the multi-objective problems in the same way as the single-objective problems; 2) an indifference-zone-free formulation, the new procedure eliminates the necessity of indifference-zone parameter; 3) asymptotically optimality, the GSPRT achieves asymptotically the shortest expected sample size among all sequential tests; 4) general distribution, the procedure uses the empirical likelihood for generally distributed observation. A numerical evaluation demonstrates the efficiency of the new procedure.

Title: Logarithmically Efficient Simulation For Misclassification Probabilities In Sequential Multiple Testing

Presenter: Yanglei Song

Co-authors: Georgios Fellouris (UIUC)

Abstract: We consider the problem of estimating via Monte Carlo simulation the misclassification probabilities of two sequential multiple testing procedures. The first one stops when all local test statistics exceed simultaneously either a positive or a negative threshold. The second assumes knowledge of the true number of signals, say m, and stops when the gap between the top m test statistics and the remaining ones exceeds a threshold. For each multiple testing procedure, we propose an importance sampling algorithm for the estimation of its misclassification probability. These algorithms are shown to be logarithmically efficient when the data for the various statistical hypotheses are independent, and each testing problem satisfies an asymptotic stability condition and a symmetry condition. Our theoretical results are illustrated by a simulation study in the special case of testing the drifts of Gaussian random walks.

Wednesday, 1:30 - 3:00, Room: 2410

Session: Optimal Control of Stochastic Systems

Chair: Mark Squillante

Title: Achievable Performance of Blind Policies in Heavy Traffic

Presenter: Bart Kamphorst

Co-authors: Bert Zwart (CWI, TU/e), Nikhil Bansal (CWI, TU/e)

Abstract: We are interested in scheduling policies for the GI/GI/1 queue where we wishto minimize the average sojourn time. It is well-known that this objectiveis minimized by the Shortest Remaining Processing Time (SRPT) policy;however, this policy needs to know all job sizes upon arrival in the system. Ifthis information is not available then the server needs to resort to so-calledblind policies.Naturally, we now wonder which blind policy to choose in order to perform asgood as possible, and how big the performance gap is when compared to theSRPT policy. This talk addresses these two questions and presents a knownblind policy that is (in some sense) optimal. The proof of this result displays apromising hybrid of competitive analysis and applied probability techniques.This is joint work with prof.dr. Bert Zwart and prof.dr. Nihil Bansal.

Title: On the performance of Tailored Base-Surge policies: evidence from Walmart.com

Presenter: Linwei Xin

Co-authors: John Bowman (Walmart Labs), Huijun Feng (Capital One), Long He (National University of Singapore), Zhiwei (Tony) Qin (Didi Research), Jagtej Bewli (Walmart Labs)

Abstract: We consider the following dual-sourcing inventory problem: one supplier is reliable but has a longer lead time; the other one is not always reliable but has a shorter lead time. It is motivated by a real-world problem at Walmart.com and the lead time differences of many import items could be as large as 12 weeks. We prove that a Tailored-Base Surge (TBS) policy is asymptotically optimal as the lead time difference grows. We also test the performance of TBS by using data from Walmart.com. Our result shows that Tailored-Base Surge outperforms other heuristics.

Title: On Optimal Weighted-Delay Scheduling in Input-Queued Switches

Presenter: Mark S. Squillante

Co-authors: Yingdong Lu (IBM Research), Siva Theja Maguluri (Ga Tech), Tonghoon Suk (IBM Research)

Abstract: Motivated by relatively few delay-optimal scheduling results, in comparison to results on throughput optimality, we investigate an input-queued switch scheduling problem in which the objective is to minimize a linear function of the queue-length vector. Theoretical properties of variants of the well-known MaxWeight scheduling algorithm are established within this context, which includes showing that these algorithms exhibit optimal heavy-traffic queue-length scaling. For the case of 2 by 2 input-queued switches, we derive an optimalscheduling policy and establish its theoretical properties, demonstrating fundamental differences with the variants of MaxWeight scheduling. Our theoretical results are expected to be of interest

more broadly than input-queued switches. Computational experiments demonstrate and quantify thebenefits of our optimal scheduling policy.

Title: Workflow Re-design to Improve Patient Service in an Emergency Department

Presenter: Yuan Zhong

Co-authors: David Yao (Columbia), Ting Zhu

Abstract: We present a queueing network model specifically developed to improve the overall performance of a medical emergency department, taking real data from Columbia University Medical Center (CUMC). The model leads to important insights as to where to best allocate additional resource and how to re-allocate existing resource so to achieve load balancing and improved patient service. The analytical results are validated against a detailed simulation model.

Wednesday, 1:30 - 3:00, Room: 2420

Session: Statistical Learning and Sequential Decision

Chair: Karthyek Murthy

Title: Modelling and analysis of maintenance decision policies

Presenter: Stella Kapodistria

Co-authors:

Abstract: In this talk, we start with a comparison of the two classical forms of maintenance policies (age based and condition based maintenance) using the framework of renewal theory and drawing conclusions on the structure of the Markov decision process solution. In the sequel, we investigate the effect of the combination of the two policies on the modelling and the analysis, and investigate the optimality of control limit policies. We conclude the talk with a Bayesian extension for data-driven tailor made policies.

Title: Forward Selection Convergence Rates

Presenter: Damian Kozbur

Co-authors:

Abstract: Forward regression is a statistical model selection and esti- mation procedure which inductively selects covariates that add predictive power into a working statistical regression model. Once a model is selected, unknown regression parameters are estimated by least squares. This pa- per analyzes forward

regression in high-dimensional sparse linear models. Probabilistic bounds for prediction error norm and number of selected co- variates are proved. The analysis in this paper gives sharp rates and does not require beta-min or irrepresentability conditions.

Title: A Bivariate Mixture of Negative Binomial Distribution and Its Applications

Presenter: Deepak Singh

Co-authors: Somesh Kumar (Indian Institute of Technology Kharagpur, India)

Abstract: We introduce a new bivariate mixture of negative binomial distribution to describe the correlated count data more efficiently. Various properties of the mixture of distributions are examined including the derivation of conditional distribution. The correlation coefficient between marginals of bivariate distribution, and joint probability generating function are also explored. Three numerical examples with comparison are carried out at the end to illustrate the effectiveness of our theory.

Title: Quantifying distributional model risk via optimal transport

Presenter: Karthyek Murthy

Co-authors: Jose Blanchet (Stanford)

Abstract: The objective is to quantify the impact of model misspecification when computing general expected values of interest. The methodology that we propose is applicable in great generality, in particular, we consider examples involving path-dependent expectations of stochastic processes. Our approach consists in computing bounds for the expectation of interest regardless of the probability measure used, as long as the measure lies within a prescribed tolerance measured in terms of a flexible class of distances from a suitable baseline model. These distances, based on optimal transportation between probability measures, include Wasserstein's distances as particular cases. The proposed methodology is well-suited for risk analysis, as we demonstrate with its applications to computing ruin probabilities.

Wednesday, 1:30 - 3:00, Room: 2430

Session: Asymptotic Analysis on Large Networks

Chair: Wen Sun

Title: Analysis of optical fibre networks with a void-avoiding schedule

Presenter: Jan-Pieter Dorsman

Co-authors: Dieter Fiems (Ghent University), Wouter Rogiest (Ghent University)

Abstract: The major growth of personalised video streaming and the paradigm shift towards big data all add to the bandwidth requirements of internet users, urging network providers to provision them with more capacity. While current optical fibre networks in theory offer great capacities in excess of 10 Tbit/s per fibre, such transmission speeds are rarely achieved in practice. This is due to the fact that optical packets need to traverse multiple intermediary nodes on their path from origin to destination, while it is hard to temporarily store the packets in case these nodes are congested. A possible solution to this problem is to send the packets into fibre delay loops when the transmission line of an intermediary node is occupied. This buffering strategy is different from classical buffering in the sense that once the transmission line becomes available, a waiting packet first needs to traverse the remainder of its delay loop before it is ready for possible transmission.In this presentation, we analyse the stationary number of optical packets being delayed in fibre loops simultaneously at an intermediary node, under the assumption that a transmission is initiated as soon as the transmission line and any packet is available for transmission (the so-called void-avoiding schedule). The analysis turns out to have surprising links with particular queueing systems, and results turn out to be strikingly simple and familiar.

Title: Community detection in networks: algorithms, complexity, and information limits

Presenter: Jiaming Xu

Co-authors: Bruce Hajek (UIUC), Yihong Wu (Yale)

Abstract: Many datasets can be viewed as networks where nodes represent objects and edges encode pairwise interactions between objects. An interesting problem is to identify communities consisting of similar objects based on the network topology. Learning communities in a large-scale network is both statistically and computationally challenging, and many different algorithms have been proposed over the years. Nevertheless, it remains unclear when it is computationally feasible to infer the communities. This talk will present an overview and our recent results toward understanding the information theoretic and computational limits of community detection in networks.

Title: A New Behavioral SIR Model, with Applications to the Swine Flu Epidemic

Presenter: Jussi Keppo

Co-authors: Elena Quercioli (University of Texas), Lones Smith (University of Wisconsin-Madison)

Abstract: Contagious diseases are passed on when contagious and susceptible individuals meet. This paper introduces and explores a new matching game, characterized by individuals meeting pairwise, possibly unwittingly passing along a disease in a contagion-like fashion. We assume that individuals can expend costly effort to avoid acquiring it. In this population game, efforts are strategic substitutes: The harder other individuals try, the more lax one can be. We solve for the unique Nash equilibrium when individuals are heterogeneous. We then estimate this structural model and show that it improves on the explanation of the data without endogenous behaviour.

Title: A mean-field study of placement algorithms in large networks

Presenter: Wen Sun

Co-authors: Philippe Robert (INRIA)

Abstract: This paper is focused on the analysis of the efficiency of replication mechanisms in large distributed systems. These algorithms determine the location of the copies of a given file in the servers of the network. They may be at the origin of the variability of the loads of the nodes of the network. We investigate for two such policies: Random Choice and Power of Choice.¥ Random Choice.Each of the regenerated files will be assigned to another server chosen atrandom among the servers in the neighborhood of the crashed server.¥ Power of d Choice.For each regenerated file, the system will choose d servers at random in theneighborhood and assign the file to the least loaded one.We study the asymptotic behaviors of these two policies in a mean-field context,i.e. when the number N of servers is large. We show that the load on each serverconverges in distribution to a nonlinear Markov Jump Process as N large. We show an interesting finite support property when the load per server is large.

Wednesday, 3:30 - 5:00, Room: 2110

Session: Production Scheduling

Chair: Justus Arne Schwarz

Title: A Novel Dynamic Scheduling Method for Manufacturing Systems with Unreliable Machines and Time Window Constraints

Presenter: Cheng-Hung Wu

Co-authors:

Abstract: This research studies scheduling problems for production systems with time-window constraints. Under time-window constraints, the waiting times of Work-in-Processes (WIPs) are constrained above before certain processing steps. In semiconductor manufacturing, hundreds of processing steps are required to transfer silicon wafers into semiconductor chips and more than 40 percent of those processing steps adopt time-windows constraints to prevent harmful deposits and oxidation from developing on wafers. Violation of time constraints causes quality degradation and may result in reworks or scraps. Because of low machine reliability in semiconductor manufacturing, the variance of waiting time increases and the scheduling problems become challenging. While aggressive production control leads to high risks of violating time window constraints, conservative control sacrifices cycle time and throughput performance. A Markov decision processes (MDP) model is developed to balance the need for low time window constraint violation and high operation efficiency. Since machine reliability and congestions are two major causes of time constraint violation, real time machine reliability and WIP distribution is explicitly considered in the model. To mitigate the 'curse-of-dimensionality' of large MDP problems, monotone structures on optimal

scheduling policies are proved. In numerical study, the proposed method reduces average production costs and the risks of time constraint violation by 15% and 50% respectively.

Title: Condition-Based Repair Prioritization in Repairable Inventory Supply Systems

Presenter: Chiel van Oosterom

Co-authors: Joachim Arts (Eindhoven University of Technology), Geert-Jan van Houtum (Eindhoven University of Technology)

Abstract: It is common to use information collected via condition monitoring of systems for optimizing replacement decisions, but this information also has great potential to improve control of the resources that are needed to perform replacements, such as spare parts and service tools. In this spirit, we propose a model for exploiting condition monitoring information on the installed base to dynamically prioritize repairs of failed repairable spare parts in a capacitated repair shop. Specifically, we consider a repair shop that supports a series system with a number of different repairable component, which all deteriorate over time according to continuous-time Markov chains. The system is down whenever a component fails and no ready-for-use spare part is available. The objective in prioritizing repairs is to maximize the long-run availability of the system. We analytically establish that it is optimal to prioritize based on shortest expected lifetime in the special case of exponentially distributed component lifetimes (i.e., the continuous-time Markov chain deterioration models all have just two statesÑworking and failedÑand no condition information can be obtained), and numerically assess the value of real-time information on the component's condition when multiple states can be distinguished.

Title: Structural properties of flow production with time-dependent processing rates

Presenter: Justus Arne Schwarz

Co-authors: Raik Stolletz (University of Mannheim)

Abstract: Flow lines process workpieces sequentially on multiple stations. The processing times are often stochastic, hence buffers are installed to decouple the stations. For these systems, structural properties characterize the relationship between design variables such as buffer capacities and the performance measures expected throughput and expected work in process inventory. The identification of structural properties is important because of their algorithmic consequences for flow line design approaches. We review structural properties of flow lines with constant processing rates that have been proven or are numerically observed under steady-state conditions. Moreover, new monotonicity results for systems with time-dependent processing rates are introduced. These properties are based on sample-path arguments for the case of exponentially distributed processing times and supported by numerical evidence for general distributions.

Wednesday, 3:30 - 5:00, Room: 2120

Session: Blocking Servers in Loss Networks

Chair: Vianney Boeuf

Title: Fluid limits and the batched processor sharing model

Presenter: Katelynn Kochalski

Co-authors: Christian Gromoll (University of Virginia)

Abstract: We consider a sequence of single-server queueing models with renewal arrivals and general i.i.d. service times operating under a service policy that incorporates batches into processor sharing. Each model is described by a measure-valued process that keeps track of the residual service times for all jobs present in the system and evolves according to a family of dynamic equations. Under mild conditions and a law-of-large-numbers scaling, we prove that the sequence of measure-valued processes converges in distribution to an essentially deterministic limit process. We show that this limit process obeys periodic dynamics that are easy to describe as a function of the initial condition.

Title: Martingale approach for tail asymptotic problems in the generalized Jackson network

Presenter: Masakiyo Miyazawa

Co-authors:

Abstract: We study the tail asymptotic of the stationary joint queue length distribution for a generalized Jackson network (GJN for short), assuming its stability. We aim to derive their upper and lower bounds in the logarithmic sense for the marginal distributions in given directions as well as for the stationary probabilities of state sets of small volumes. It is shown that those bounds are identical for the two node case. Our tool is a martingale, which enables to use change of measure. A key ingredient here is to see how the GJN is changed under the new measure, which will be a GJN again, but some of their nodes are unstable.

Title: An Asymptotic Analysis of Blocking in a Finite Capacity Network with Two Levels

Presenter: Vianney Boeuf

Co-authors: Philippe Robert, INRIA Paris

Abstract: In this paper a stochastic model of a finite capacity system with a 2-level architecture is analyzed. A first-level pool of operators answers calls, handles non-urgent calls. If a call is identified as urgent it requires a specific service and it is transfered to specialized second level operators if one of them is available. When the operators of the second level are all busy, the operator of first level is blocked until the urgent call can be assigned at the second level. Such blocking procedure has some similarities with classical loss networks but, in this case, the invariant probability distribution of our system does not seem to

have a closed form expression. We investigate, under a scaling assumption, Kelly's regime, the evolution of the number of urgent calls blocked at level 1. It is shown that if the ratio of the number of operators at level 2 and 1 is greater than some constant depending on the parameters of the traffic, then the system operates without congestion (no urgent call is blocked) with probability 1 after some finite time, otherwise it is proved that a positive fraction of calls are blocked at level after some time. Stochastic calculus with Poisson processes, coupling arguments and generalized Skorokhod problems are the main mathematical tools to establish these convergence results.

Wednesday, 3:30 - 5:00, Room: 2130

Session: Managing Queueing Systems: Abandonment, Learning and Priorities

Chair: Philipp Afeche and Bora Keskin

Title: Coverage, Coarseness and Classification: Determinants of Social Efficiency in Priority Queues

Presenter: Martin Lariviere

Co-authors: Itai Gurvich (Northwestern), Can Ozkan (Northwestern)

Abstract: We examine differences in how a revenue maximizer and a social planner manage a priority queue. We consider a single server queue with customers that draw their valuations from a continuous distribution and have a per-period waiting cost that is proportional to their realized valuation. The decision maker posts a menu offering a finite number of waiting time-price pairs, which determines coverage (i.e., how many customer in total to serve), coarseness (i.e., how many classes of service to offer), and classification (i.e., how to map customers to priority levels).We show that differences between the decision makers' priority policies are all about classification. Both are content to offer very coarse schemes with just two priority levels, and they will have negligible differences in coverage. However, differences in classification are persistent. A revenue maximizer may -- relative to the social planner -- have too few or too many high priority customers. Whether the revenue maximizer over- or under-stuffs the high priority class depends on a measure of consumer surplus that is captured by the mean residual life function of the valuation distribution. In addition we show that there is a large class of valuation distributions for which a move from first-in, first-out service to a priority scheme that places those with higher waiting costs at the front of the line reduces consumer surplus.

Title: Observational learning and abandonment in congested sstems

Presenter: Costis Maglaras

Co-authors: John Yao and Assaf Zeevi (Columbia)

Abstract: Demand systems used in operations management and service operations settings often assume that system parameters that may affect user decisions, e.g., to join a system or purchase a service, are known or accurately communicated to the market. In several practical settings, this need not be the case, but users may still form estimates of these system parameters through their own observations or experiences in the system. In this talk, we will study the effect of observational learning on user behavior and equilibrium system performance in the context of a queueing model. Specifically, we analyze a congested service system in which delay-sensitive customers have no a priori knowledge of the service rate, but instead join the system and observe their progress through the queue in order to learn the system's service rate, estimate remaining waiting times, and make abandonment decisions.

Title: Managing Services where service times and valuations are correlated

Presenter: Chenguang Allen Wu

Co-authors: Achal Bassamboo (Northwestern), Ohad Perry (Northwestern)

Abstract: Motivated by recent empirical evidence, we consider a large service system in which the patience time of each customer depends on his service requirement. Our goal is to study the impact of such dependence on key performance measures, such as expected waiting times and average queue length, as well as on optimal capacity decisions. Since the dependence structure renders exact analysis intractable, we employ a stationary fluid approximation that is based on the entire joint distribution of the service and patience times. Our results show that even moderate dependence has significant impacts on system performance, so considering the patience and service times to be independent when they are in fact dependent is futile. We further demonstrate that Pearson's correlation coefficient, which is commonly used to measure and rank dependence, is an insufficient statistic, and that the entire joint distribution is required for comparative statics. Thus, we propose a novel framework, incorporating the fluid model with bivariate dependence orders and copulas, to study the impacts of the aforementioned dependence. We then demonstrate how that framework can be applied to facilitate revenue optimization when staffing and abandonment costs are incurred. Finally, the effectiveness of the fluid-based approximations and optimal-staffing prescriptions is demonstrated via simulations.

Title: Learning and Earning for Congestion-prone Service Systems

Presenter: Bora Keskin

Co-authors: Philipp Afeche (University of Toronto)

Abstract: Consider a firm selling a service in a congestion-prone system to price- and delay-sensitive customers. The firm faces Bayesian uncertainty about the consumer demand for its service and can dynamically make noisy observations on the demand. We characterize the structure and performance of the myopic Bayesian policy and well-performing variants.

Wednesday, 3:30 - 5:00, Room: L-130

Session: Optimal Scheduling

Chair: Siva Theja Maguluri

Title: Optimally Scheduling Jobs with Multiple Tasks of Unknown Duration

Presenter: Ziv Scully

Co-authors: Guy Blelloch (Carnegie Mellon University), Mor Harchol-Balter (Carnegie Mellon University), Alan Scheller-Wolf (Carnegie Mellon University)

Abstract: We consider optimal job scheduling where each job consists of multiple tasks, each of unknown duration, with precedence constraints between tasks. A job is not considered complete until all of its tasks are complete. Traditional heuristics, such as favoring the job of shortest expected remaining processing time, are suboptimal in this setting. Furthermore, even if we know which job to run, it is not obvious which task within that job to serve. In this talk, we characterize the optimal policy for a class of such scheduling problems. We show the policy is simple to compute in many practical cases.

Title: Near Delay-Optimal Job Scheduling and Task Replications over Parallel Machines: A Novel Sampling-Path Method

Presenter: Yin Sun

Co-authors: C. Emre Koksal (OSU), Ness B. Shroff (OSU)

Abstract: In modern computer systems, long-running jobs are divided into small tasks and executed on multiple servers. Empirical observations in practical cloud systems suggest that the task service times are highly random and the delay to complete a job (including the waiting time in the queue and the service time to complete all tasks of the job) is bottlenecked by the slowest task. One approach to tame the long job delay is to replicate tasks over multiple servers such that one of the copies is completed early. Google has reported that task replications can reduce the job delay significantly, e.g., in one study replications reduce the 99.9%-th percentile delay from 1,800 ms to 74 ms. However, in many other systems, replications actually increase the delay, because redundant replications can increase the system load. So far, little is understood on how to optimize the scheduling decisions (e.g., when to replicate, which servers to replicate on, and which job to serve first) to minimize the job delay.In this talk, I will present a comprehensive study on delay-optimal job scheduling and task replications in multi-server systems. I will describe low-complexity scheduling policies that we have developed that make scheduling decisions based on the type of service time distribution and the delay metric to be optimized. For arbitrary number, sizes, arrival times, and due times (soft deadlines) of the jobs, these scheduling policies are proven to be delay-optimal or near delay-optimal in a stochastic ordering sense among all non-preemptive and causal policies. These results can characterize (near) delay optimality for arbitrary arrival processes, for both transient- and steady-state systems, and for minimizing several classes of delay metrics, including average delay, maximum delay,

jitter probability, and maximum lateness. Novel sample-path ordering and coupling methods are developed to prove these general results. The design principles proposed in this study can be used to obtain dramatic delay reduction without sacrificing throughput. The multi-server model is also applicable to communication networks and cloud storage systems.

Title: Delay Performance of Scheduling Algorithms for Data Center Networks and Input Queued Switches

Presenter: Siva Theja Maguluri

Co-authors: R. Srikant (UIUC), Sai Kiran Burle (UIUC)

Abstract: Today's era of cloud computing is powered by massive data centers hosting servers that are connected by high speed networks. It is therefore desirable to design scheduling algorithms for data packets that have low computational complexity and result in small average packet delays. We consider the scheduling problem in an input-queued switch, which is a good abstraction for a data center network. We present low complexity scheduling algorithms that have optimal queue length (equivalently, delay) behavior in the heavy traffic regime. We also present bounds on the queue length in light traffic. These results are obtained using drift based arguments.

Title: Bandwidth Sharing with Phase-Type File Size Distribution

Presenter: Weina Wang

Co-authors: Siva Theja Maguluri (Georgia Tech), R. Srikant (UIUC), Lei Ying (ASU)

Abstract: We consider the Massouli-Roberts model for bandwidth sharing, where file sizes have a phase-type distribution and proportionally fair resource allocation is used. We analyze the expected number of files in steady-state by setting the steady-state drift of an appropriately chosen Lyapunov function equal to zero. We obtain asymptotically tight bounds on the expected number of files in the system in the heavy-traffic regime, thus complementing the diffusion approximation result of Vlasiou, Zhang, and Zwart.

Wednesday, 3:30 - 5:00, Room: L-120

Session: Performance and Optimization of Power Systems

Chair: Stella Kapodistria

Title: Optimal energy storage operations in power grids under uncertainty

Presenter: Alessandro Zocca

Co-authors: Bert Zwart (CWI, Amsterdam)

Abstract: Power grids are increasingly affected by uncertainty due to the intermittent nature of renewable generation. In this talk I will present a stylized model for energy network under uncertainty, aiming to get insight in the interplay between renewable energy and grid reliability.The physical network is modeled by a weighted graph G, where nodes represent buses and edges represent transmission lines. The power injected or consumed in the network nodes is described by a power injection vector p, modeled as a random vector or multidimensional stochastic process, that uniquely determines the current flows f in the network edges under the DC power flow approximation. Using this model in a joint work with B. Zwart we investigate stochastic optimization of energy storage. More specifically, we consider a scenario where energy storage devices ("batteries") that can coordinate their operations are added to the energy network. Such batteries can both charge using the network current excess or discharge to meet the network current demand. Either way, the presence of batteries can be leveraged to mitigate the intrinsic uncertainty in the power generation and demand and, hence, transport the energy more efficiently through the network. The performance metric that we consider is the expected total heat loss $\mathbb{E} H(\alpha)$ when the battery load-sharing control α is used, where the total heat loss H is a random variable (or stochastic process) can be expressed as a quadratic for of the power injection vector p. I will show how the optimal control α^* depends on the network structure, on the correlations between the power injections, as well as on the number of available batteries and on their displacement in the network.If time allows, I will also present extensions to a dynamic setting, in which we derive the optimal control $\alpha(t)$ over a finite time interval in the case where the power injections are modeled by Ornstein-Uhlenbeck processes.

Title: Managing Stored Energy in Microgrids via Multistage Stochastic Programming

Presenter: Arnab Bhattacharya

Co-authors: Jeffrey P. Kharoufeh (University of Pittsburgh)

Abstract: Energy storage systems are used to mitigate adverse effects of renewable sources in a microgrid where procurement and storage decisions are made under uncertain demand, renewable supply and prices. A multistage stochastic programming (SP) model is formulated to minimize the expected total costs in a microgrid. To improve computational tractability, a customized stochastic dual-dynamic programming (SDDP) algorithm is employed to obtain high-quality solutions within a reasonable time. A numerical study highlights significant cost reductions and computational benefits of the enhanced SDDP algorithm.

Title: Energy imbalance market call options and the valuation of storage

Presenter: John Moriarty

Co-authors: Jan Palczewski (Leeds)

Abstract: The use of energy storage to balance electric grids is increasing and, with it, the importance of operational optimisation from the twin viewpoints of cost and system stability. In this paper we assess the real option value of balancing reserve provided by an energy-limited storage unit. The contractual arrangement is a series of American-style call options in an energy imbalance market (EIM), physically covered and delivered by the store, and purchased by the power system operator. We take the EIM price

as a general regular one-dimensional diffusion and impose natural economic conditions on the option parameters. In this framework we derive the operational strategy of the storage operator by solving two timing problems: when to purchase energy to load the store (to provide physical cover for the option) and when to sell the option to the system operator. We give necessary and sufficient conditions for the finiteness and positivity of the value function -- the total discounted cash flows generated by operation of the storage unit. We also provide a straightforward procedure for the numerical evaluation of the optimal operational strategy (EIM prices at which power should be purchased) and the value function. This is illustrated with an operational and economic analysis using data from the German Amprion EIM. The talk is based on joint work with Jan Palczewski, and I will also mention some recent extensions.

Title: Wind turbine and wind farm power output modelling

Presenter: Sndor Kolumbn

Co-authors: Nazanin Nooraee (TU/e), Stella Kapodistria (TU/e)

Abstract: We are interested in the modelling of the power output of a single wind turbine and the combined power output of a wind farm for short and long term predictions.As a first step, towards short term predictions, and following the guidelines of the existing literature we explored various parametric and non-parametric techniques for the modelling of the wind turbine power curve (WTPC) of a single turbine. All of these techniques seem to have an intrinsic limitation in terms of accuracy, making the corresponding models inappropriate for short term forecasting. To avoid this conundrum, we show that adding a properly scaled autoregressive-moving-average (ARMA) modelling layer increases short term prediction performance while keeping the long term prediction capabilities of WTPC models given wind information.Afterwards, we show how to explore the wind park aggregated data and we investigate if the joint distribution of power output and wind speed can be modelled using a Gaussian setting.

Wednesday, 3:30 - 5:00, Room: L-070

Session: Stochastic Optimization and Simulation

Chair: Krzysztof Bisewski

Title: Subsolution approach for the simulation of a rare event in the GIIGII1 tandem queue

Presenter: Anne Buijsrogge

Co-authors: Pieter-Tjerk de Boer (University of Twente), Werner Scheinhardt (University of Twente)

Abstract: In this work we study the event that the total number of customers in a GIIGII1 tandem queue reaches some high level in a busy cycle of the system. Using simulation, we want to estimate the probability of this event. As this event is rare when the level is high, we choose to use importance sampling

to speed up the simulation. Therefore we need to find a change of measure that is asymptotically efficient.For a Markovian tandem queue a state-dependent change of measure has previously been proven to be asymptotically efficient by using the so-called subsolution approach. Our goal is to generalize this to a GIIGII1 tandem queue in order to build more realistic models.The approach we use is similar to the existing subsolution approach. To obtain a Markov process, we add the residual inter-arrival time and the residual service time to the state space description. This approach seems to work well in order to find a state-dependent change of measure for the GIIGII1 tandem queue and to prove asymptotic efficiency. For the case of a single GIIGII1 queue, this approach enables us to find an alternative proof for asymptotic efficiency of a state-independent change of measure.

Title: A Random Monotone Operator Framework for Stochastic Optimization

Presenter: Rahul Jain

Co-authors: William Haskell (NUS)

Abstract: Analysis of every algorithm for stochastic optimization seems to require a different convergence proof. It would be desirable to have a unified mathematical framework within which with minimal extra effort, proof of convergence and its rate could be obtained. We first present a random monotone operator-based unified convergence analysis framework for iterative algorithms for strongly convex stochastic optimization. The framework offers both versatility and simplicity, and allows for clean and straightforward analysis of many algorithms for stochatic convex minimization, saddle-point problems and variational inequalities. We show convergence of the random operator to a probabilistic fixed point, and obtain non-asymptotic rates of convergence. We then consider the non-strongly convex case. The analysis technique relies on a simple but powerful stochastic dominance technique wherein we construct an easy to analyze Markov chain.

Title: Efficient Sampling Methods for Stochastic Min-max Optimization

Presenter: Soumyadip Ghosh

Co-authors: Mark S Squillante (IBM Research), Ebisa D Wollega (Colorado State)

Abstract: Motivated by an energy systems application, we study efficientalgorithms to find approximately optimal solutions to stochasticmin-max formulations. The standard approach is to solve a large-sampleapproximation to the true problem. Many common large-scale statisticallearning formulations (e.g. regularized risk minimization, multiplekernel learning, robust machine learning) are special instances ofthis large-sample approximate form. Standard cutting-plane or columngeneration methods are applied to this large-scale formulation, butcomputational speed suffers from having to generate sub-gradientinformation over the entire large sample set. We propose analternative stochastic first-order (gradient following) recursionprocedure to solve the outer (minimization) problem, where gradientinformation is gathered from the inner (maximization) problem similarto the cutting-plane methods. However, under this approach, it issufficient to compute the gradients from small subsamples of theentire dataset, and convergence is guaranteed as long as the sampleset size growth satisfies some regularity conditions. The method canalso be computationally efficient (in a certain precise sense) if therate at which this set grows is carefully

controlled. We characterizethe regimes when the fastest possible convergence rates areachieved. Preliminary numerical results show the efficacy of ourapproach.

Title: Minimizing time discretization error for the supremum of Brownian Motion

Presenter: Krzysztof Bisewski

Co-authors: Daan Crommelin (Centrum Wiskunde en Informatica), Michel Mandjes (Kortweg de Vries Institute for Mathematics, University of Amsterdam)

Abstract: We consider the error arising from time discretization when estimating w(b) - the tail of the distribution of a supremum of a real stochastic process over a finite time interval. For the standard Brownian Motion we demonstrate that the error can be significantly reduced by using other discretization grids than the commonly used equidistant grids. In particular, we show that in order to control the error as b grows large, it suffices to properly shift the gridpoints instead of refining the grid with more and more points. At the same time, controlling the error using equidistant grids requires quadratic (in b) growth of the number of gridpoints, as b grows large. The discretization grids that we develop can be used to construct a strongly efficient algorithm for the estimation of w(b).

Wednesday, 3:30 - 5:00, Room: 2410

Session: Recent Developments in Markov Decision Processes

Chair: Flora Spieksma and Michael Katehakis

Title: Kolmogorov's equations for jump Markov processes with unbounded jump rates

Presenter: Manasa Mandava

Co-authors: Eugene Feinberg (SunySB), Albert Shiryaev (Steklov Mathematical Institute)

Abstract: As is well-known, transition probabilities of jump Markov processes satisfy Kolmogorov's backward and forward equations. In the seminal 1940 paper, William Feller investigated solutions of Kolmogorov's equations for jump Markov processes. Recently the authors solved the problem studied by Feller and showed that the minimal solution of Kolmogorov's backward and forward equations is the transition probability of the corresponding jump Markov process if the transition rate at each state is bounded. This paper presents more general results. For Kolmogorov's backward equation, the sufficient condition for the described property of the minimal solution is that the transition rate at each state is locally integrable, and for Kolmogorov's forward equation the corresponding sufficient condition is that the transition rate at each state is locally bounded.

Title: Optimal and Heuristic Policies for Dual-Speed Search Problems

Presenter: Jake Clarkson

Co-authors: Kevin Glazebrook (Lancaster), Peter Jacko (Lancaster), Christopher Kirkbride (Lancaster), Kyle Lin (Naval Postgraduate School)

Abstract: A hidden object needs to be found in many real-life situations, some of which involve large costs and significant consequences with failure. Therefore, efficient search methods are paramount. Further, there is often a choice regarding search speed. Areas can be covered slowly, or quickly, with a faster search using less time but increasing the probability of missing the object. This trade-off is core to this research; we model a search in multiple discrete areas with two available search speeds, fast and slow.When only one speed is available, a policy which minimises the expected total search time is known. Here, each area has an associated search time t and detection probability q with which the object, if located there, is found upon searching that area. The searcher originally believes the object is in some area with prior probability p, with p continually updated to a posterior p' as the search progresses. The optimal policy assigns each area an index p'q/t, searching the one with the largest index.In the dual-speed problem, each area has fast and slow detection probabilities, search times and, hence, indices. Further, any policy additionally needs to determine search speeds. We have proven that if the slow index is greater than or equal to its fast counterpart, the corresponding area is optimally always searched slowly. Further, we know that once the fast index becomes sufficiently greater than the slow, with 'sufficient' depending on the size of the slow detection probability, it is optimal to always search the area at the fast speed. Below this sufficient level, the optimal choice of speed may change. Determining this speed can be complicated, it depends on the size of the area's detection probabilities, the current posterior probabilities and the other areas available to search.Hence, here, we look for heuristic policies, of which we have two candidates. The first involves solving a simplified two area problem whenever we are required to make a choice of speed. The second assigns each area a fixed speed based upon a function of its detection probabilities. We have also developed lower bounds on the optimal search time and an upper bound on how suboptimal the second policy can be. A numerical study has been constructed to compare the two heuristics and assess the tightness of these bounds. In the future, we intend to move the problem to a network setting, using previous results to derive more heuristic policies.

Title: Recent Developments in Markov Decision Processes and Beyond Motivated by Inventory Control

Presenter: Eugene A Feinberg

Co-authors:

Abstract: The talk describes the recent progress in the theory of Markov Decision Processes (MDPs). This progress is motivated by inventory control applications of MDPs. The progress became possible because of recent generalizations of two facts in real analysis: Fatou's lemma and Berge's maximum theorem. In addition to inventory control applications, the talk describes new results on game theory and robust optimization.

Title: Markov decision processes with a special lumpable structure: server farm optimization

Presenter: Floske Spieksma

Co-authors: Herman Blok, Eindhoven University of Technology, Netherlands

Abstract: In this talk we consider a Markov decision process, that can be modeled as a controlled quasi-birth-death process with the following special lumpable structure:the level sets of the quasi-birth-death process each contain one state that is both the entrance state and the exit state of the level set.If the level sets are finite, this structure can be exploited to derive an efficient algorithm that at each iteration computes the optimal policy per level, both for the discounted and average cost optimality criteria.We will apply this to a server farm optimization problem.

Wednesday, 3:30 - 5:00, Room: 2420

Session: Statistical Learning and Optimization

Chair: Umit Deniz Tursun

Title: Yule's "Nonsense Correlation" Solved!

Presenter: Philip Ernst

Co-authors: L.A. Shepp (Wharton), A.J. Wyner (Wharton)

Abstract: Abstract: In this talk, I will discuss how I recently resolved a longstanding open statistical problem. The problem, formulated by the British statistician Udny Yule in 1926, is to mathematically prove Yule's 1926 empirical finding of "nonsense correlation.' We solve the problem by analytically determining the second moment of the empirical correlation coefficient of two independent Wiener processes. Using tools from Fredholm integral equation theory, we calculate the second moment of the empirical correlation to obtain a value for the standard deviation of the empirical correlation of nearly .5. The "nonsense" correlation, which we call "volatile" correlation, is volatile in the sense that its distribution is heavily dispersed and is frequently large in absolute value. It is induced because each Wiener process is "self-correlated" in time. This is because a Wiener process is an integral of pure noise and thus its values at different time points are correlated. In addition to providing an explicit formula for the second moment of the empirical correlation, we offer implicit formulas for higher moments of the empirical correlation. The full paper is currently in press at The Annals of Statistics and can be found at http://www.imstat.org/aos/AOS1509.pdf.

Title: Probabilistic CART with Proper Scoring

Presenter: Sara Shashaani

Co-authors: Matthew Plumlee (UMich), Seth Guikema (UMich)

Abstract: Predictive distributions are often preferred over point predictions for systems with large variance in the outputs. Interestingly a simply calibrated predictive distribution, that is when the CDF-inverse is uniformly distributed, does not yield the 'best' predictive distribution. This is because the 'best' predictive distribution is surprisingly ill-defined in the presence of covariates; in other words, calibrated predictive distributions are not unique. To evaluate probabilistic forecast methodologies, proper scoring rules can provide feedback about forecast deficiencies. Naturally one can raise the question whether proper scoring rules combined with modern machine learning methods can result in improved distributional predictive capabilities. We study this question through the construction of Classification and Regression Trees (CART). Currently the default algorithms in building CART are based on implicit distributional assumptions on the data; e.g. least squares is obtained by assuming normality. This is not guaranteed to give good performance in terms of predictive distributions. We investigate the improvement in the predictive performance by incorporating a proper scoring as a split condition in construction of CART. Preliminary testing has revealed encouraging results that this method can produce trees with superior capabilities for probabilistic prediction.

Title: Representing Markov processes as dynamic, copula-based Bayesian networks

Presenter: Thomas G. Yeung

Co-authors: Alex Kosgodagan (IMT-Atlantique)

Abstract: In multivariate statistics, recent attractive approaches include copula-based graphical models and specifically so-called pair-copula Bayesian networks. Their attractiveness is largely due to the flexibility that copula models provide, whereby the marginal distributions can be modelled arbitrarily, and any dependence captured by the copula. However, very little attention has been given for these models to fit within a full probabilistic framework and for which inference could desirably be used. In this paper, we first prove that any k-th order Markov process can be represented as a dynamic pair copula-based Bayesian network. Dependence is formulated as (conditional) bivariate time-copulas derived from the Markov process as well as the corresponding (conditional) rank correlation. Second, we explicitly show the requirements in order to perform analytical conditioning. We finally illustrate our findings through an example focused on Brownian motion.

Title: RANDOM PROJECTION METHODS FOR STOCHASTIC CONVEX MINIMIZATION

Presenter: Umit Deniz Tursun

Co-authors: Angelia Nedich (Arizona State University)

Abstract: The focus of this talk is to solve a stochastic convex minimization problem over an arbitrary family of nonempty, closed and convex sets. The problem has random features. Gradient or subgradient of objective function carries stochastic errors. Number of constraint sets can be extensive or infinitely many.

Constraint sets might not be known apriori yet revealed through random realizations or randomly chosen from a collection of constraint sets throughout the horizon as in online learning concept. We showed that projecting onto a random subcollection of them using our algorithm with diminishing stepsize is sufficient to converge to the solution set almost surely. Also the convergence of the algorithm for constant and nondiminishing nonsummable stepsizes are proved within an error bound.

Wednesday, 3:30 - 5:00, Room: 2430

Session: Asymptotics of Large Random Graphs and Matrices

Chair: Pascal Moyal

Title: Inhomogeneous random digraphs

Presenter: Mariana Olvera-Cravioto

Co-authors:

Abstract: The talk will describe a large class of inhomogeneous directed random graphs for modeling complex networks such as the web graph, Twitter, ResearchGate, and other social networks. This class of graphs includes as a special case the classical Erdos-Renyi model, and can be used to replicate almost any type of predetermined degree distributions, in particular, power-law degrees such as those observed in most real-world networks. The talk will cover the basic connectivity properties of this family as well as the behavior of algorithms such as Google's PageRank.

Title: Scaling Limits and Generic Bounds for Exploration Processes

Presenter: Jaron Sanders

Co-authors: Paola Bermolen (UdelaR), Matthieu Jonckheere (UBA)

Abstract: In this talk we will consider exploration algorithms of the random sequential adsorption type both for homogeneous random graphs and random geometric graphs based on spatial Poisson processes. In these exploration algorithms we select at each step a vertex of the graph, which becomes active and its neighboring nodes become explored. We will discuss how, given an initial number of vertices N growing to infinity, we can study statistical properties of the proportion of explored nodes in time using scaling limits. More precisely, we can obtain exact limits for homogeneous graphs and prove an explicit central limit theorem for the final proportion of active nodes, known as the jamming constant, through a diffusion approximation for the exploration process. Next we will focus on bounding the trajectories of such exploration processes on random geometric graphs, i.e. random sequential adsorption. As opposed to homogeneous random graphs, these do not allow for a reduction in dimensionality. Instead we can build on a fundamental relationship between the number of explored nodes and the discovered volume in the spatial

process, and obtain generic bounds: bounds that are independent of the dimension of space and the detailed shape of the volume associated to the discovered node. And by constructing coupled exploration processes that have the same fluid limits, we can give trajectorial interpretations of these bounds.This talk is based on a preprint available at:https://arxiv.org/abs/1612.09347

Title: High Dimensional Linear Regression with Few Samples

Presenter: Ilias Zadik

Co-authors: D. Gamarnik (MIT)

Abstract: In this talk we will talk about the sparse high dimensional regression Y=X\beta^{*}+W where X is n\times p matrix with i.i.d. standard normal entries, Wis n\times 1 vector with i.i.d. N(0,\sigma^{2}) entries and \beta^{*} is p\times 1 binary vector with k entries equal to unity (k-sparse). The goal is recovering with high probability (w.h.p) the vector \beta^{*} from observed X and Y. In the literature some remarkable papers by Wainwright, Donoho, Candes, Tao and others, show that LASSO (l_1-constrained quadratic programming) and Compressed Sensing techniques, such as Basis Pursuit Denoising Scheme, can exactly recover w.h.p. \beta^* as long as n > 2k \log p but no mechanism can recover w.h.p. \beta^* with n0, n> (1+\epsilon)n^* then the maximum likelihood estimator w.h.p. is recovering almost all the support of \beta^*, but if n<(1-\epsilon)n^* w.h.p. it fails to recover any positive fraction of the true support.This is joint work with David Gamarnik.

Title: The impact of degree variability on connectivity properties of large networks

Presenter: Lasse Leskela

Co-authors: Hoa Ngo (Aalto University)

Abstract: The goal of this work is to study how increased variability in the degreedistribution impacts the global connectivity properties of a largenetwork. We approach this question by modeling the network as a uniformrandom graph with a given degree sequence. We analyze the effect of thedegree variability on the approximate size of the largest connectedcomponent using stochastic ordering techniques. A counterexample showsthat a higher degree variability may lead to a larger connected component,contrary to basic intuition about branching processes. When certainextremal cases are ruled out, the higher degree variability is shown todecrease the limiting approximate size of the largest connected component.(Based on joint work with Hoa Ngo, Aalto University; arXiv:1508.03379).