

# **34th Conference on Uncertainty in Artificial Intelligence 2018**

Monterey, California, USA  
6-10 August 2018

Volume 1 of 2

**Editors:**

**Amir Globerson**

**Ricardo Silva**

ISBN: 978-1-5108-7160-1

**Printed from e-media with permission by:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571



**Some format issues inherent in the e-media version may also appear in this print version.**

Copyright© (2018) by Association For Uncertainty in Artificial Intelligence (AUAI)  
All rights reserved.

Printed by Curran Associates, Inc. (2018)

For permission requests, please contact Association For Uncertainty in Artificial Intelligence (AUAI)  
at the address below.

Association For Uncertainty in Artificial Intelligence (AUAI)  
Department of Computer Science  
University of British Columbia  
Vancouver BC Canada V6T 1Z4

Phone: (703) 787-8044

Fax: (703) 478-0829

[www.auai.org](http://www.auai.org)

**Additional copies of this publication are available from:**

Curran Associates, Inc.  
57 Morehouse Lane  
Red Hook, NY 12571 USA  
Phone: 845-758-0400  
Fax: 845-758-2633  
Email: [curran@proceedings.com](mailto:curran@proceedings.com)  
Web: [www.proceedings.com](http://www.proceedings.com)

# Contents

<b>Organizing Committee</b>	<b>vii</b>
<b>Acknowledgments</b>	<b>ix</b>
<b>Sponsors</b>	<b>xvii</b>
<b>1 Proceedings</b>	<b>1</b>
Testing for Conditional Mean Independence with Covariates through Martingale Difference Divergence. <i>Ze Jin, Xiaohan Yan, David S. Matteson</i> . . . . .	2
Analysis of Thompson Sampling for Graphical Bandits Without the Graphs. <i>Fang Liu, Zizhan Zheng, Ness Shroff</i> . . . . .	13
Structured nonlinear variable selection. <i>Magda Gregorova, Alexandros Kalousis, Stephane Marchand-Maillet</i> . . . . .	23
Identification of Strong Edges in AMP Chain Graphs. <i>Jose M. Peña</i> . . . . .	33
A Univariate Bound of Area Under ROC. <i>Siwei Lyu, Yiming Ying</i> . . . . .	43
Efficient Bayesian Inference for a Gaussian Process Density Model. <i>Christian Donner, Manfred Opper</i> . . . . .	53
Comparing Direct and Indirect Temporal-Difference Methods for Estimating the Variance of the Return. <i>Craig Sherstan, Dylan R. Ashley, Brendan Bennett, Kenny Young, Adam White, Martha White, Richard S. Sutton</i> . . . . .	63
How well does your sampler really work?. <i>Ryan Turner, Brady Neal</i> . . . . .	73
Learning Deep Hidden Nonlinear Dynamics from Aggregate Data. <i>Yisen Wang, Bo Dai, Lingkai Kong, Sarah Monazam Erfani, James Bailey, Hongyuan Zha</i>	83
Revisiting differentially private linear regression: optimal and adaptive prediction & estimation in unbounded domain. <i>Yu-Xiang Wang</i> . . . . .	93
Imaginary Kinematics. <i>Sabina Marchetti, Alessandro Antonucci</i> . . . . .	104
From Deterministic ODEs to Dynamic Structural Causal Models. <i>Paul K. Rubenstein, Stephan Bongers, Joris M. Mooij, Bernhard Schölkopf</i> . . . . .	114
Frank-Wolfe Optimization for Symmetric-NMF under Simplicial Constraint. <i>Han Zhao, Geoff Gordon</i> . . . . .	124
Learning Time Series Segmentation Models from Temporally Imprecise Labels. <i>Roy Adams, Benjamin M. Marlin</i> . . . . .	135
Multi-Target Optimisation via Bayesian Optimisation and Linear Programming. <i>Alistair Shilton, Santu Rana, Sunil Gupta, Svetha Venkatesh</i> . . . . .	145
Stochastic Learning for Sparse Discrete Markov Random Fields with Controlled Gradient Approximation Error. <i>Sinong Geng, Zhaobin Kuang, Jie Liu, Stephen Wright, David Page</i> . . . . .	156

Active Information Acquisition for Linear Optimization. <i>Shuran Zheng, Bo Waggoner, Yang Liu, Yiling Chen</i> . . . . .	167
Transferable Meta Learning Across Domains. <i>Bingyi Kang, Jiashi Feng</i> . . . . .	177
Learning the Causal Structure of Copula Models with Latent Variables. <i>Ruifei Cui, Perry Groot, Moritz Schauer, Tom Heskes</i> . . . . .	188
$f_{BGD}$ : Learning Embeddings From Positive Unlabeled Data with BGD. <i>Fajie YUAN, Xin Xin, Xiangnan He, Guibing Guo, Weinan Zhang, CHUA Tat-Seng, Joemon Jose</i> . . . . .	198
Soft-Robust Actor-Critic Policy-Gradient. <i>Esther Derman, Daniel J Mankowitz, Timothy A Mann, Shie Mannor</i> . . . . .	208
Constant Step Size Stochastic Gradient Descent for Probabilistic Modeling. <i>Dmitry Babichev, Francis Bach</i> . . . . .	219
Discrete Sampling using Semigradient-based Product Mixtures. <i>Alkis Gotovos, Hamed Hassani, Andreas Krause, Stefanie Jegelka</i> . . . . .	229
Combining Knowledge and Reasoning through Probabilistic Soft Logic for Image Puzzle Solving. <i>Somak Aditya, Yezhou Yang, Chitta Baral, Yiannis Aloimonos</i> . . . . .	238
Nesting Probabilistic Programs. <i>Tom Rainforth</i> . . . . .	249
Scalable Algorithms for Learning High-Dimensional Linear Mixed Models. <i>Zilong Tan, Kimberly Roche, Xiang Zhou, Sayan Mukherjee</i> . . . . .	259
Constraint-based Causal Discovery for Non-Linear Structural Causal Models with Cycles and Latent Confounders. <i>Patrick Forré, Joris M. Mooij</i> . . . . .	269
Marginal Weighted Maximum Log-likelihood for Efficient Learning of Perturb-and-Map models. <i>Tatiana Shpakova, Francis Bach, Anton Osokin</i> . . . . .	279
Variational Inference for Gaussian Processes with Panel Count Data. <i>Hongyi Ding, Young Lee, Issei Sato, Masashi Sugiyama</i> . . . . .	290
A unified probabilistic model for learning latent factors and their connectivities from high-dimensional data. <i>Ricardo Pio Monti, Aapo Hyvarinen</i> . . . . .	300
Improved Stochastic Trace Estimation using Mutually Unbiased Bases. <i>Jack Fitzsimons, Michael Osborne, Stephen Roberts, Joseph Fitzsimons</i> . . . . .	310
Unsupervised Multi-view Nonlinear Graph Embedding. <i>Jiaming Huang, Zhao Li, Vincent W. Zheng, Wen Wen, Yifan Yang, Yuanmi Chen</i> . . . . .	319
Graph-based Clustering under Differential Privacy. <i>Rafael Pinot, Anne Morvan, Florian Yger, Cedric Gouy-Pailler, Jamal Atif</i> . . . . .	329
GaAN: Gated Attention Networks for Learning on Large and Spatiotemporal Graphs. <i>Jiani Zhang, Xingjian Shi, Junyuan Xie, Hao Ma, Irwin King, Dit-yan Yeung</i> . . . . .	339
Causal Learning for Partially Observed Stochastic Dynamical Systems. <i>Søren Wengel Mogensen, Daniel Malinsky, Niels Richard Hansen</i> . . . . .	350
Variational zero-inflated Gaussian processes with sparse kernels. <i>Pashupati Hegde, Markus Heinonen, Samuel Kaski</i> . . . . .	361
KBIn: End-to-End Learning of Knowledge Base Representations with Latent, Relational, and Numerical Features. <i>Alberto Garcia-Duran, Mathias Niepert</i> . . . . .	372
Probabilistic AND-OR Attribute Grouping for Zero-Shot Learning. <i>Yuval Atzmon, Gal Chechik</i> . . . . .	382
Sylvester Normalizing Flows for Variational Inference. <i>Rianne van den Berg, Leonard Hasenclever, Jakub Tomczak, Max Welling</i> . . . . .	393
Holistic Representations for Memorization and Inference. <i>Yunpu Ma, Marcel Hildebrandt, Volker Tresp, Stephan Baier</i> . . . . .	403
Simple and practical algorithms for $\ell_p$ -norm low-rank approximation. <i>Anastasios Kyrillidis</i> . . . . .	414

Quantile-Regret Minimisation in Infinitely Many-Armed Bandits. <i>Arghya Roy Chaudhuri, Shivaram Kalyanakrishnan</i> . . . . .	425
Variational Inference for Gaussian Process Models for Survival Analysis. <i>Minyoung Kim, Vladimir Pavlovic</i> . . . . .	435
A Cost-Effective Framework for Preference Elicitation and Aggregation. <i>Zhibing Zhao, Haoming Li, Junming Wang, Jeffrey O. Kephart, Nicholas Mattei, Hui Su, Lirong Xia</i> . . . . .	446
Incremental Learning-to-Learn with Statistical Guarantees. <i>Giulia Denevi, Carlo Ciliberto, Dimitris Stamos, Massimiliano Pontil</i> . . . . .	457
Bandits with Side Observations: Bounded vs. Logarithmic Regret. <i>Rmy Degenne, Evrard Garcelon, Vianney Perchet</i> . . . . .	467
Sampling and Inference for Beta Neutral-to-the-Left Models of Sparse Networks. <i>Benjamin Bloem-Reddy, Adam Foster, Emile Mathieu, Yee Whye Teh</i> . . . . .	477
Clustered Fused Graphical Lasso. <i>Yizhi Zhu, Oluwasanmi Koyejo</i> . . . . .	487
Unsupervised Learning of Latent Physical Properties Using Perception-Prediction Networks. <i>David Zheng, Vinson Luo, Jiajun Wu, Joshua Tenenbaum</i> . . . . .	497
Subsampled Stochastic Variance-Reduced Gradient Langevin Dynamics. <i>Difan Zou, Pan Xu, Quanquan Gu</i> . . . . .	508
Finite-State Controllers of POMDPs using Parameter Synthesis. <i>Sebastian Junges, Nils Jansen, Ralf Wimmer, Tim Quatmann, Leonore Winterer, Joost-Pieter Katoen, Bernd Becker</i> . . . . .	519
Identification of Personalized Effects Associated With Causal Pathways. <i>Ilya Shpitser, Eli Sherman</i> . . . . .	530
Fast Counting in Machine Learning Applications. <i>Subhadeep Karan, Matthew Eichhorn, Blake Hurlburt, Grant Iraci, Jaroslaw Zola</i> . . . . .	540
A Dual Approach to Scalable Verification of Deep Networks. <i>Krishnamurthy Dvijotham, Robert Stanforth, Sven Gowal, Timothy Mann, Pushmeet Kohli</i>	550
Understanding Measures of Uncertainty for Adversarial Example Detection. <i>Lewis Smith, Yarin Gal</i> . . . . .	560
Causal Discovery in the Presence of Measurement Error. <i>Tineke Blom, Anna Klimovskaia, Sara Magliacane, Joris M. Mooij</i> . . . . .	570
IDK Cascades: Fast Deep Learning by Learning not to Overthink. <i>Xin Wang, Yujia Luo, Daniel Crankshaw, Alexey Tumanov, Fisher Yu, Joseph E. Gonzalez</i>	580
Learning Fast Optimizers for Contextual Stochastic Integer Programs. <i>Vinod Nair, Dj Dvijotham, Iain Dunning, Oriol Vinyals</i> . . . . .	591
Differential Analysis of Directed Networks. <i>Min Ren, Dabao Zhang</i> . . . . .	601
Sparse-Matrix Belief Propagation. <i>Reid Bixler, Bert Huang</i> . . . . .	611
Sequential Learning under Probabilistic Constraints. <i>Amirhossein Meisami, Henry Lam, Chen Dong, Abhishek Pani</i> . . . . .	621
Abstraction Sampling in Graphical Models. <i>Filjor Broka, Rina Dechter, Alexander Ihler, Kalev Kask</i> . . . . .	632
Meta Reinforcement Learning with Latent Variable Gaussian Processes. <i>Steindor Saemundsson, Katja Hofmann, Marc Peter Deisenroth</i> . . . . .	642
Non-Parametric Path Analysis in Structural Causal Models. <i>Junzhe Zhang, Elias Bareinboim</i> . . . . .	653
Stochastic Layer-Wise Precision in Deep Neural Networks. <i>Griffin Lacey, Graham W. Taylor, Shawki Areibi</i> . . . . .	663
Estimation of Personalized Effects Associated With Causal Pathways. <i>Razieh Nabi, Phyllis Kanki, Ilya Shpitser</i> . . . . .	673
High-confidence error estimates for learned value functions. <i>Touqir Sajed, Wesley Chung, Martha White</i> . . . . .	683

Combinatorial Bandits for Incentivizing Agents with Dynamic Preferences. <i>Tanner Fiez, Shreyas Sekar, Liyuan Zheng, Lillian Ratliff</i> . . . . .	693
Sparse Multi-Prototype Classification. <i>Vikas K. Garg, Lin Xiao, Ofer Dekel</i> . . . . .	704
Fast Stochastic Quadrature for Approximate Maximum-Likelihood Estimation. <i>Nico Piatkowski, Katharina Morik</i> . . . . .	715
Finite-sample Bounds for Marginal MAP. <i>Qi Lou, Rina Dechter, Alexander Ihler</i> . . . . .	725
Acyclic Linear SEMs Obey the Nested Markov Property. <i>Ilya Shpitser, Robin Evans, Thomas S. Richardson</i> . . . . .	735
A Unified Particle-Optimization Framework for Scalable Bayesian Sampling. <i>Changyou Chen, Ruiyi Zhang, Wenlin Wang, Bai Li, Liqun Chen</i> . . . . .	746
An Efficient Quantile Spatial Scan Statistic for Finding Unusual Regions in Continuous Spatial Data with Covariates. <i>Travis Moore, Weng-Keen Wong</i> . . . . .	756
Stable Gradient Descent. <i>Yingrue Zhou, Sheng Chen, Arindam Banerjee</i> . . . . .	766
Learning to select computations. <i>Frederick Callaway, Sayan Gul, Paul M. Krueger, Thomas L. Griffiths, Falk Lieder</i> . . . . .	776
Per-decision Multi-step Temporal Difference Learning with Control Variates. <i>Kristopher De Asis, Richard S. Sutton</i> . . . . .	786
The Indian Buffet Hawkes Process to Model Evolving Latent Influences. <i>Xi Tan, Vinayak Rao, Jennifer Neville</i> . . . . .	795
Battle of Bandits. <i>Aadirupa Saha, Aditya Gopalan</i> . . . . .	805
Adaptive Stochastic Dual Coordinate Ascent for Conditional Random Fields. <i>Rémi Le Priol, Alexandre Piché, Simon Lacoste-Julien</i> . . . . .	815
Adaptive Stratified Sampling for Precision-Recall Estimation. <i>Ashish Sabharwal, Yexiang Xue</i> . . . . .	825
Fast Kernel Approximations for Latent Force Models and Convolved Multiple-Output Gaussian processes. <i>Cristian Guarnizo, Mauricio Álvarez</i> . . . . .	835
Fast Policy Learning through Imitation and Reinforcement. <i>Ching-An Cheng, Xinyan Yan, Nolan Wagener, Byron Boots</i> . . . . .	845
Hyperspherical Variational Auto-Encoders. <i>Tim Davidson, Luca Falorsi, Nicola De Cao, Thomas Kipf, Jakub M. Tomczak</i> . . . . .	856
Dissociation-Based Oblivious Bounds for Weighted Model Counting. <i>Li Chou, Wolfgang Gatterbauer, Vibhav Gogate</i> . . . . .	866
Averaging Weights Leads to Wider Optima and Better Generalization. <i>Pavel Izmailov, Dmitrii Podoprikin, Timur Garipov, Dmitry Vetrov, Andrew Gordon Wilson</i> . . . . .	876
Block-Value Symmetries in Probabilistic Graphical Models. <i>Gagan Madan, Ankit Anand, Mausam, Parag Singla</i> . . . . .	886
Max-margin learning with the Bayes factor. <i>Rahul G. Krishnan, Arjun Khandelwal, Rajesh Ranganath, David Sontag</i> . . . . .	896
Densified Winner Take All (WTA) Hashing for Sparse Datasets. <i>Beidi Chen, Anshumali Shrivastava</i> . . . . .	906
Lifted Marginal MAP Inference. <i>Vishal Sharma, Noman Ahmed Sheikh, Happy Mittal, Vibhav Gogate, Parag Singla</i> . . . . .	917
PAC-Reasoning in Relational Domains. <i>Ondrej Kuzelka, Yuyi Wang, Jesse Davis, Steven Schockaert</i> . . . . .	927
Pure Exploration of Multi-Armed Bandits with Heavy-Tailed Payoffs. <i>Xiaotian Yu, Han Shao, Michael R. Lyu, Irwin King</i> . . . . .	937
Counterfactual Normalization: Proactively Addressing Dataset Shift Using Causal Mechanisms. <i>Adarsh Subbaswamy, Suchi Saria</i> . . . . .	947

Decentralized Planning for Non-dedicated Agent Teams with Submodular Rewards in Uncertain Environments. <i>Pritee Agrawal, Pradeep Varakantham, William Yeoh</i> . . . . .	958
A Forest Mixture Bound for Block-Free Parallel Inference. <i>Neal Lawton, Greg Ver Steeg, Aram Galstyan</i> . . . . .	968
Causal Identification under Markov Equivalence. <i>Amin Jaber, Jiji Zhang, Elias Bareinboim</i> . . . . .	978
The Variational Homoencoder: Learning to learn high capacity generative models from few examples. <i>Luke B. Hewitt, Maxwell I. Nye, Andreea Gane, Tommi Jaakkola, Joshua B. Tenenbaum</i>	988
Probabilistic Collaborative Representation Learning for Personalized Item Recommendation. <i>Aghiles Salah, Hady W. Lauw</i> . . . . .	998
Reforming Generative Autoencoders via Goodness-of-Fit Hypothesis Testing. <i>Aaron Palmer, Dipak Dey, Jinbo Bi</i> . . . . .	1009
Towards Flatter Loss Surface via Nonmonotonic Learning Rate Scheduling. <i>Sihyeon Seong, Yegang Lee, Youngwook Kee, Dongyoon Han, Junmo Kim</i> . . . . .	1020
A Lagrangian Perspective on Latent Variable Generative Models. <i>Shengjia Zhao, Jiaming Song, Stefano Ermon</i> . . . . .	1031
Bayesian optimization and attribute adjustment. <i>Stephan Eismann, Daniel Levy, Rui Shu, Stefan Bartzsch, Stefano Ermon</i> . . . . .	1042
Join Graph Decomposition Bounds for Influence Diagrams. <i>Junkyu Lee, Alexander Ihler, Rina Dechter</i> . . . . .	1053
Causal Discovery with Linear Non-Gaussian Models under Measurement Error: Structural Identifiability Results. <i>Kun Zhang, Mingming Gong, Joseph Ramsey, Kayhan Batmanghelich, Peter Spirtes, Clark Glymour</i> . . . . .	1063