Lecture Notes for "Stochastic Modeling and Computations"

M. Chertkov (lecturer), S. Belan and V. Parfeneyv (recitation instructors)
M.Sc. and Ph.D. level course at Skoltech
Moscow, March 27 - May 28, 2017
`https://sites.google.com/site/mchertkov/courses`

The course offers a soft and self-contained introduction to modern applied probability, covering theory and applications of stochastic models. Emphasis is placed on intuitive explanations of the theoretical concepts, such as random walks, law of large numbers, Markov processes, reversibility, sampling, etc., supplemented by practical/computational implementations of basic algorithms. In the second part of the course, the focus shifts from general concepts and algorithms per se to their applications in science and engineering with examples, aiming to illustrate the models and make the methods of solution, originating from physics, chemistry, machine learning, control and operations research, clear and exciting.

Contents

Brief Description. Structure. Requirements. 3

(Tentative) Schedule 4

I. Theme #1. Basic Concepts from Statistics 6
- A. Lecture #1. Random Variables: Characterization & Description. 6
  1. Probability of an event 6
  2. Sampling. Histograms. 6
  3. Moments. Generating Function. 7
  4. Probabilistic Inequalities. 7
  5. Recitation. Random Variables. Moments. Characteristic Function. 8
- B. Lecture #2. Random Variables: from one to many. 8
  1. Law of Large Numbers 8
  2. Multivariate Distribution. Marginalization. Conditional Probability. 9
  3. Bayes Theorem 10
  4. Recitation. Properties of Gaussian Distributions. Laws of Large Numbers. 10
- C. Lecture #3. Information-Theoretic View on Randomness 10
  1. Entropy. 10
  2. Independence/Dependence. Mutual Information. 12
  3. Information Channel 13
  4. Probabilistic Inequalities for Entropy and Mutual Information 13
  5. Recitation. Entropy, Mutual Information and Probabilistic Inequalities 15

II. Theme # 2. Stochastic Processes 15
- A. Lecture #4: Markov Chains [discrete space, discrete time]. 15
  1. Transition Probabilities 15
  2. Properties of Markov Chains 16
  3. Steady State Analysis 16
  4. Spectrum of the Transition Matrix & Speed of Convergence to the Stationary Distribution 17
  5. Reversible & Irreversible Markov Chains. 17
  6. Detailed Balance vs Global Balance. Adding cycles to accelerate mixing. 18
  7. Recitation. Markov Chains: Detailed Balance. Mixing time. 18
- B. Lecture #5. From Bernoulli Processes to Poisson Processes [discrete space, discrete & continuous time]. 18
  1. Bernoulli Process: Definition 18
  2. Bernoulli: Number of Successes 19
  3. Bernoulli: Distribution of Arrivals 19
  4. Poisson Process: Definition 19
  5. Poisson: Arrival Time 20
  6. From Bernoulli to Poisson: # of arrivals in t-intervals as $n \to \infty$ 21

7. Merging and Splitting Processes 21
8. Recitation. Examples of Bernoulli & Poisson Processes 22
C. Lecture #6. Monte-Carlo Algorithms: General Concepts and Direct Sampling. 22
1. Direct-Sampling by Rejection vs MCMC for ‘pebble game’ 22
2. Direct Sampling by Mapping 22
3. Direct Sampling by Rejection (another example) 23
4. Importance Sampling 23
5. Direct Brut-force Sampling 23
6. Direct Sampling from a multi-variate distribution with a partition function oracle 24
7. Ising Model 24
D. Lecture #7. Markov-Chain Monte-Carlo. 25
1. Gibbs Sampling 25
2. Metropolis-Hastings Sampling 26
3. Glauber Sampling of Ising Model 26
4. Exactness and Convergence 27
5. Exact Monte Carlo Sampling (Did it converge yet?) 27

III. Theme # 3: Graphical Models 28
A. Lecture #8. Exact & Approximate Inference. 28
1. From Ising Model to (Factor) Graphical Models 28
2. Decoding of Graphical Codes as a Factor Graph problem 29
3. Partition Function. Marginal Probabilities. Maximum Likelihood. 30
4. Kublack-Leibler Formulation & Probability Polytope 31
5. Variational Approximations. Mean Field. 32
B. Lecture #9. Inference & Learning with Belief Propagation 32
1. Bethe Free Energy & Belief Propagation 32
2. Belief Propagation & Message Passing 33
3. Sufficient Statistics 34
4. Maximum-Likelihood Estimation/Learning of GM 35
5. Recitation: Direct (Inference) and Inverse (Learning) Problems over Trees. 35

IV. Theme #4: Stochastic Modeling & Optimization 35
A. Lecture #10. Space-time Continuous Stochastic Processes 35
1. Langevin equation in continuous and discrete time 36
2. From Langevin to Path Integral 36
3. From Path Integral to Fokker-Planch (through sequential Gaussian integration) 37
4. Analysis of Fokker-Planck: General Features and Examples 37
5. Recitation. Homogeneous and Forced Brownian Motion. 38
6. Recitation. First Passage Problem. Effects of Boundaries. Kramers Escape Problem. 38
B. Lecture #11. Queuing Systems. 38
1. Queuing: a bit of History & Applications 38
2. Single Open Queue = Birth/Death process. Markov Chain representation. 39
3. Generalization to (Jackson) Networks. Product Solution for the Steady State. 40
4. Heavy Traffic Limit 41
5. Recitation. Tandem Queue Example. 42
C. Lecture #12. Markov Decision Processes & Stochastic Optimal Control. 42
1. Optimal Control [discrete time, deterministic] & Dynamic Programming 42
2. Optimal Control [continuous space, continuous time, deterministic] 43
3. Stochastic Optimal Control [continuous space, continuous time, stochastic] 43
4. Markov Decision Processes [discrete space, discrete time, stochastic] 44
5. MDP: Grid World Example 46
6. Recitation. Dynamic Programming. 47

V. Subjects for Journal Club Presentations & Reports: (Incomplete) Pool of Options 48
A. General Information 48
B. Incomplete List of Suggested Subjects 48

References 50

Brief Description. Structure. Requirements.

This course is recommended to M.Sc. and Ph.D. students planning to work on the subjects containing elements of uncertainty, irregularity or what is also called 'stochasticity'. The 'stochastic' subjects are prevalent in natural sciences (physics, chemistry, biology) and engineering disciplines (electrical-, mechanical-, chemical-, industrial-, etc). The course introduces students to modeling and computational concepts, approaches, methods and algorithms which require dealing with stochasticity and uncertainty.

This is a general course recommended to Energy Systems, CDISE and other Skoltech students. The course is recommended as a core course for students who specializes in computations. It can be chosen as elective by students who use computations and algorithms in their work, however not as the prime focus.

There will be 12 lectures, 12 recitations, 2 homework assignments and journal club presentations/reports and, finally, the exam.

Lectures. Lecture notes are to be provided (online) before the actual lecture. Lecturer will mainly be using white-board, sometimes supplemented by computer demonstrations in IJulia.

Recitations. Recitation notes are to be provided (online) after the actual recitations. Recitations will be lead by two instructors (alternating) with the use of whiteboard and computer demonstrations (ijulia notebooks). Students in the class may be called to lead discussions/solving the problems.

Homework assignments. Two assignments will be given. Each homework will consist of ~ 6 problems, including multiple ($\sim 2-4$) sub-problems of varying difficulty. First homework will be distributed in the beginning of the second week and will be collected by Sun, Apr 23, 11:59pm Moscow time. Second homework will be distributed in the beginning of the 4th week and will be collected by Sun, May 21, 11:59pm Moscow time. Problems in the homework will be similar in principle, but different in details from these discussed in lectures and recitations (prior to the homework distribution). Solutions from the homework will be discussed at the recitations after the homework collection. It is encouraged to use electronic formats (latex and/or ipython/ijulia) for the homework reports. Submission of the homework(s) is electronic only.

Each student will be required to choose a subject for **journal club presentation and report**. Suggested subjects are listed below in the document. In terms of picking a subject – the policy is 'first come first served'. The list is not meant to be complete or exclusive. In particular, the students are encouraged to suggest additional subjects linked to the course material and possibly related to student's own research focus/interest. All additional subjects should be discussed with and approved by the lecturer. Subjects should be presented during the presentation session (tentatively) scheduled for May 22 (extra sessions may be added, if needed). Each presentation is 20 mins. All reports should be submitted by May 28, 11:59pm. Reports are individual, should be at least 10 pages but not longer that 20 pages. Presentations and reports will be graded together.

A written **exam** will be administered. The exam will include $3-4$ problems similar to these discussed at the recitations and contained in the homework. Format of the exam (in class or take home) will be decided at later time depending on how the class progresses.

The three **books** referred extensively in lectures, recitations and homework are [1–3]. In addition, many relevant reviews and papers available online are cite in the lecture notes. Students may also find it useful to check [4–7] for related (but often alternative) explanations. (A number of hard copies of all the aforementioned books are available at the edu@skoltech library.

On pre-requisites and requirements. All necessary concepts from statistics, probability theory and statistical mechanics will be introduced in the course self-consistently (no formal pre-requisites in these disciplines are required). However, solid preparation in practical math (ability to solve problems in linear algebra, calculus, and differential equations) will be required from anybody taking this course.

We will mainly be using in lectures and recitations for computations and illustrations Julia `http://julialang.org/` under IJulia/Ipython-notebook environment `https://github.com/JuliaLang/IJulia.jl`. A very convenient online web-service for Julia/Ijuia (which does not require any local installations) is available at `https://juliabox.com/`. Students are encouraged to self-learn and use Julia and IJulia. However, computations (e.g. for homework and exam) in any other (reasonably common and transparent) programming languages will also be accepted.

Grading:

- Homework – 30%
- Exam – 30%
- Journal Club Presentation & Report – 20%
- Participation – 20%

(Tentative) Schedule

Mon & Fri 16:00–17:30 + 17:30–19:00 (two periods 1.5 hours each)
l=lecture, r=recitation

First week:

March 27, Mon
16:00–17:30 l#1 Random Variables: Characterization and Description
17:30-19:00 l#2 Random Variables: Operations & Transformations

March 31, Fri
16:00-17:30 l#3 Information-Theoretic View on Randomness
17:30-19:00 l#4 Markov Chains [discrete space, discrete time]

Second week:

April 3, Mon
16:00-17:30 r#1 Moments/Averages/Cumulants/Generation Function on Examples
17:30-19:00 r#2 Example of Gaussian Variables: Matrix Inversion, Normalization, Moments

Apr 7, Fri
16:00-17:30 l#5 From Bernoulli Processes to Poisson Processes [discrete space, discrete & continuous time]
17:30-19:00 l#6 Monte-Carlo Algorithms: General Concepts and Direct Sampling

Third week:

Apr 10, Mon
16:00-17:30 l#7 Markov-Chain Monte-Carlo
17:30-19:00 l#8 Exact & Approximate Inference

Apr 14, Fri
16:00-17:30 r#3 Entropy, Mutual Information and Probabilistic Inequalities on Example (Communication over Noisy Channel)
17:30-19:00 r#4 Markov Chains: Detailed Balance. Mixing time.

Fourth week:

Apr 17, Mon
16:00-17:30 r#5 Examples of Bernoulli & Poisson Processes
17:30-19:00 r#6 MC and MCMC on example of the Ising model

Apr 21, Fri
16:00-17:30 l#9 Inference & Learning with Belief Propagation
17:30-19:00 l#10 Space-time Continuous Stochastic Processes

Fourth week:

Apr 24, Mon
16:00-17:30 l#11 Queuing Systems
17:30-19:00 l#12 Markov Decision Processes & Stochastic Optimal Control

Apr 28, Tue
16:00-17:30 r#7 Inference & Learning on Trees
17:30-19:00 r#8 Homogeneous and Forced Brownian Motion

Fifth & Six weeks:

Holidays, homework, work on projects

Seven week:

May 15, Mon
16:00-17:30 r#9 First Passage Problem and Effects of Boundaries
17:30-19:00 r#10 Queuing Systems

May 19, Fri
16:00-17:30 r#11 Markov Decision Processes & Stochastic Optimal Control

10:30-12:00 r#12 TBD

Eight week:

May 22, Mon
16:00-17:30 Project presentations
17:30-19:00 Project presentations

May 23, Tue
16:00-17:30 Project presentations (if needed)
17:30-19:00 Project presentations (if needed)

May 26, Fri
16:00-19:00 Exam (in class or take-home TBD later)

V. SUBJECTS FOR JOURNAL CLUB PRESENTATIONS & REPORTS: (INCOMPLETE) POOL OF OPTIONS

A. General Information

Each student will be required to choose a subject for **journal club presentation and report**. List of suggested subjects is listed below. In terms of picking a subject – the policy is 'first come first served'. Please e-mail the lecturer as soon as possible.

The list is not meant to be complete or exclusive. In particular, the students are encouraged to suggest additional subjects linked to the course material and possibly related to their own research focus/interest. All additional subjects should be discussed with and approved by the lecturer.

Subjects should be presented during the presentation session which will be scheduled for May 22, Mon. (Extra sessions may be added during the week of May 22, if needed.) Each presentation is 20 mins. All reports should be submitted by May 28, 11:59pm.

Reports are individual, should be at least 10 pages but not longer that 20 pages. Presentations and reports will be graded together. See `http://www.people.fas.harvard.edu/~rpoddar/Papers/ldpc.pdf` for an exemplary student report.

Projects resulting in julia/ijulia programs/illustrations on the subjects linked to the lectures, which can be used as a basis for illustrations in the course in the future, are especially encouraged.

B. Incomplete List of Suggested Subjects

Large Deviation for Multiplicative Processes

Stretching and Rotations of clouds and particles, ordered exponentials, long time statistics of Lyapunov exponents. Cramer/entropy function. `http://arxiv.org/abs/cond-mat/0105199`

The Noisy Channel Coding (Shannon) Theorem

Sec. 9.3 and 10 of [2]

Compressed Sensing and its many uses (How l_1 norm promotes sparcity?)

Pick a review from the extended list available at `https://en.wikipedia.org/wiki/Compressed_sensing` An original option is `http://statweb.stanford.edu/~candes/papers/DecodingLP.pdf`

Slice Sampling MCMC

See `https://en.wikipedia.org/wiki/Slice_sampling`. Recommended review is Neal, Radford M. (2003). "Slice Sampling". Annals of Statistics 31 (3): 705767.

Simulated Annealing Sampling

Important idea and algorithm allowing to explore seriously non-convex problems – rugged landscape with multiple valleys, saddle points, minima and peaks. The original paper is Kirkpatrick, S.; Gelatt Jr, C. D.; Vecchi, M. P. (1983). "Optimization by Simulated Annealing". Science 220 (4598): 671680. See also `https://en.wikipedia.org/wiki/Simulated_annealing` and references there in.

Hamiltonian MCMC

MCMC which is capable to accelerate sampling by adding additional degrees of freedom - related to controlled inertia/momenta expressed through a Hamiltonian description (from physics) — thus the name. Recommended review `http://www.cs.utoronto.ca/~radford/ftp/ham-mcmc.pdf`

Irreversible Monte Carlo algorithms for efficient sampling

The original paper is `http://arxiv.org/abs/0809.0916`.

Warm Algorithm in Classical and Quantum Statistical Physics

The original paper is `http://scholarworks.umass.edu/cgi/viewcontent.cgi?article=2194&context=physics_faculty_pubs`. See also `http://wiki.phys.ethz.ch/quantumsimulations/_media/lecture_101007.pdf`.

Gillespie algorithm

Sampling from stochastic equations (Langevin type) which proceeds by jumps. See the original paper Gillespie, Daniel T. (1977). "Exact Stochastic Simulation of Coupled Chemical Reactions". The Journal of Physical Chemistry 81 (25): 23402361 and also check `https://en.wikipedia.org/wiki/Gillespie_algorithm`.

Sequential Monte Carlo for Importance Sampling & Inference

Recommended paper `https://www.irisa.fr/aspi/legland/ensta/ref/doucet00b.pdf`.

Ising models and Other Graphical Models in Image Analysis

Recommended tutorial `https://www.math.ntnu.no/~joeid/TMA4250/image_ana.pdf`.

Efficient Exact Inference in Planar Ising Model

Recommended paper `http://arxiv.org/pdf/0810.4401.pdf`.

Stochastic Resonances

Curious physics phenomena important in optics & communications which explains how noise/randomness allows to amplify signal and observe what otherwise would be difficult to detect. Recommended paper is Benzi, R.; Sutera, A.; and Vulpiani, A. ”The Mechanism of Stochastic Resonance.” J. Phys. A 14, L453-L457, 1981.

Decoding of Low Density Parity Check Codes

Section 47 of [2]. Implementation of a message passing decoding in julia/ijulia is especially encouraged.

Analytic and Algorithmic Solution of Satisfiability Problem

The original paper is `http://cacs.usc.edu/education/cs653/Mezard-RSAT-Science02.pdf` Also check the book of Mezard and Montanari + papers/reviews of Parisi, Mezard and Zechina.

Neural Network Learning

Part V of [2].

Jackson Networks of Queues

Recommended paper is Kelly, F. P. (Jun 1976). ”Networks of Queues”. Advances in Applied Probability 8 (2): 416432. See alos `https://en.wikipedia.org/wiki/Jackson_network` and references there in. It may also be useful to consult with the recent book: ”Stochastic Networks” by E. Yudovina and F. Kelly, Cambridge University Press, 2014. Implementation of a julia/ijulia illustration/program for this subject is especially encouraged.

Path Integral Control & Reinforcement Learning

Recommended review `http://www.snn.ru.nl/~bertk/kappen_granada2006.pdf` Implementation of a julia/ilulia illustration/program for this subject is especially encouraged.

Adaptive Importance Sampling

Paragraph 10.5 and 10.6 from `http://statweb.stanford.edu/~owen/mc/Ch-var-adv.pdf` Implementation/demonstration of a julia/ilulia illustration/program for this subject is especially encouraged.

Cross-Entropy Method

for statistical model checking `http://people.rennes.inria.fr/sean.sedwards/publications/CAV2012.pdf`

for estimations `https://web.stanford.edu/~glynn/papers/2013/KroeseRubinsteinG13.pdf`

Adaptive Multiple Importance Sampling

`https://arxiv.org/pdf/0907.1254.pdf`

Multiple Importance Sampling for Gaussian Processes

for Gaussian Processes `https://arxiv.org/pdf/1508.01050.pdf`

[1] N. van Kampen, *Stochastic Processes in Physics and Chemistry (Third Edition)*, third edition ed. Amsterdam: Elsevier, 2007. [Online]. Available: http://www.sciencedirect.com/science/article/pii/B9780444529657500003

[2] D. J. C. Mackay, *Information theory, inference, and learning algorithms*. Cambridge: Cambridge University Press, 2003. [Online]. Available: http://www.inference.phy.cam.ac.uk/itprnn/book.html

[3] C. Moore and S. Mertens, *The Nature of Computation*. New York, NY, USA: Oxford University Press, Inc., 2011. [Online]. Available: http://www.nature-of-computation.org/

[4] H. E. Taylor and S. Karlin, *An Introduction to Stochastic Modeling*, 3rd ed. Academic Press, Feb. 1998. [Online]. Available: http://www.ime.usp.br/~fmachado/MAE5709/KarlinTaylorIntrodStochModeling.pdf

[5] C. W. Gardiner, *Handbook of stochastic methods for physics, chemistry and the natural sciences*, 3rd ed., ser. Springer Series in Synergetics. Berlin: Springer-Verlag, 2004, vol. 13.

[6] B. L. Nelson, *Stochastic modeling - analysis and simulation (reprint from 1995).* Dover Publications, 2002.

[7] E. Cinclar, *Introduction to stochastic processes*. Prentice-Hall, 1975. [Online]. Available: http://gso.gbv.de/DB=2.1/CMD?ACT=SRCHA&SRT=YOP&IKT=1016&TRM=ppn+021423008&sourceid=fbw_bibsonomy

[8] T. Richardson and R. Urbanke, *Modern Coding Theory*. Cambridge University Press, 2008.

[9] J. S. Yedidia, W. T. Freeman, and Y. Weiss, "Constructing free-energy approximations and generalized belief propagation algorithms," *Information Theory, IEEE Transactions on*, vol. 51, no. 7, pp. 2282–2312, 2005.

[10] M. J. Wainwright and M. I. Jordan, "Graphical models, exponential families, and variational inference," *Found. Trends Mach. Learn.*, vol. 1, no. 1-2, pp. 1–305, Jan. 2008. [Online]. Available: https://www.eecs.berkeley.edu/~wainwrig/Papers/WaiJor08_FTML.pdf

[11] F. P. Kelly, *Reversibility and stochastic networks*, ser. Wiley series in probability and mathematical statistics. New York, Chichester: Wiley, 1979. [Online]. Available: http://opac.inria.fr/record=b1117013

[12] L. Kleinrock, *Queueing Systems Volume 1: Theory*. Wiley, 1975.

[13] J. F. C. Kingman, "On queues in heavy traffic," *Journal of the Royal Statistical Society. Series B (Methodological) (Wiley)*, vol. 24 (2), p. 383392, 1962.

[14] W. Iglehart, Donald L. Ward, "Multiple channel queues in heavy traffic. ii: Sequences, networks, and batches," *Advances in Applied Probability (Applied Probability Trust)*, vol. 2 (2), p. 355369, 1970.

[15] J. Köllerström, "Heavy traffic theory for queues with several servers i," *Journal of Applied Probability (Applied Probability Trust)*, vol. 11 (3), pp. 544–552, 1974.

[16] D. L. Iglehart, "Limiting diffusion approximations for the many server queue and the repairman problem," *Journal of Applied Probability*, vol. 2, no. 2, pp. 429–441, 1965.

[17] A. A. Borovkov, "On limit laws for service processes in multi-channel systems," *Siberian Mathematical Journal*, vol. 8 (5), pp. 746–763,, 1967.

[18] D. L. Iglehart, "Weak convergence in queueing theory," *Advances in Applied Probability*, vol. 5, no. 3, pp. 570–594, 1973.

[19] The algorithm is justified through a standard Dynamic Programming arguments, of the type discussed above.