Progress report on partial inconsistency, bitopology, and vector semantics

Michael Bukatin

Nokia Corporation, Cambridge, MA

Joint work with Ralph Kopperman and Steve Matthews

- - -

Conference on Computational Topology and Its Applications, Kent State University, Ohio, November 7, 2014

Electronic coordinates

These slides are linked from my page on partial inconsistency and vector semantics of programming languages:

http://www.cs.brandeis.edu/∼bukatin/partial_inconsistency.html

E-mail:

bukatin@cs.brandeis.edu

Outline

1. Introduction
   - Partial inconsistency landscape
   - Partially inconsistent interval numbers
   - Lawvere duality
2. Bitopology and the domain of arrows
   - *d*-frames
   - Rodabaugh representation
   - Group dual topology
   - $[R]$-valued distances and relations
3. Linear models of computation
   - MCMC sampling and probabilistic programming
   - Generalized animations

Partial inconsistency landscape

- Negative distance/probability/degree of set membership
- Bilattices
- Partial inconsistency
- Non-monotonic inference
- Bitopology
- $x = (x \wedge 0) + (x \vee 0)$ or $x = (x \wedge \bot) \sqcup (x \vee \bot)$
- Scott domains tend to become embedded into vector spaces
- Modal and paraconsistent logic and possible world models
- Bicontinuous domains
- The domain of arrows, $D^{Op} \times D$ or $C^{Op} \times D$

Partial inconsistency landscape

- Negative distance/probability/degree of set membership
- Bilattices
- Partial inconsistency
- Non-monotonic inference
- Bitopology
- $x = (x \wedge 0) + (x \vee 0)$ or $x = (x \wedge \perp) \sqcup (x \vee \perp)$
- Scott domains tend to become embedded into vector spaces
- Modal and paraconsistent logic and possible world models
- Bicontinuous domains
- The domain of arrows, $D^{Op} \times D$ or $C^{Op} \times D$

Interval numbers

Segments $[a, b]$ on real line, $a \leq b$.

What $[a, b]$ means: $[a, b]$ stands for a partially defined number x, what is known about x is the constraint $a \leq x \leq b$.

Partial order on the interval numbers:
$[a, d] \sqsubseteq [b, c]$ iff $a \leq b$ $(\leq)$ $c \leq d$.

Here $[b, c]$ is better (more precisely) defined than $[a, d]$.

Addition and weak minus

Addition: $[a_1, b_1] + [a_2, b_2] = [a_1 + a_2, b_1 + b_2]$.

Weak minus: $-[a, b] = [-b, -a]$.

These are monotonic operations:
$x \sqsubseteq y \Rightarrow x + z \sqsubseteq y + z$ and $-x \sqsubseteq -y$.

However, the minus is weak, e.g. $-[2, 3] = [-3, -2]$, so
$-[2, 3] + [2, 3] = [-1, 1] \sqsubset [0, 0]$.

So one does not get a group here.
And it would be nice to have a group.

Partially inconsistent interval numbers

Add pseudosegments $[a, b]$, such that $b < a$.

This corresponds to contradictory constraints, $x \leq b \& a \leq x$.

The new set consists of segments and pseudosegments.

Addition: $[a_1, b_1] + [a_2, b_2] = [a_1 + a_2, b_1 + b_2]$.

True minus: $-[a, b] = [-a, -b]$.

$-[a, b] + [a, b] = [0, 0]$.

This gets us a group.

True minus is antimonotonic

$x \sqsubseteq y \Rightarrow -y \sqsubseteq -x.$

True minus maps precisely defined numbers, $[a, a]$, to precisely defined numbers, $[-a, -a]$.

Other than that, true minus maps segments to pseudosegments and maps pseudosegments to segments.

In the bicontinuous setup, true minus is a bicontinuous functon from $[R]$ to $[R]^{Op}$ (or from $[R]^{Op}$ to $[R]$).

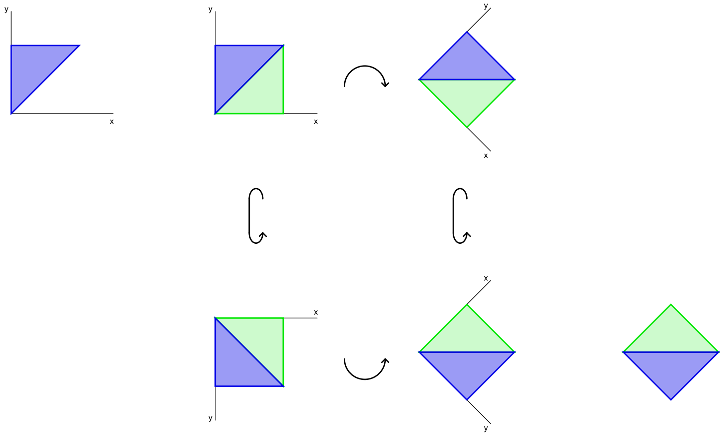
Multiple rediscoveries

Known under various names: Kaucher interval arithmetic, directed interval arithmetic, generalized interval arithmetic, modal interval arithmetic, interval algebraic extensions, etc.

First mention we know: M. Warmus, Calculus of Approximations. Bull. Acad. Pol. Sci., Cl. III, 4(5): 253-259, 1956, http://www.cs.utep.edu/interval-comp/warmus.pdf

A comprehensive repository of literature on the subject is maintained by Evgenija Popova: The Arithmetic on Proper & Improper Intervals (a Repository of Literature on Interval Algebraic Extensions), http://www.math.bas.bg/∼epopova/directed.html

From Cartesian to Hasse representation

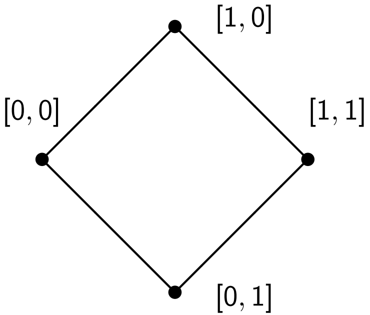


Partially inconsistent interval numbers as a domain of arrows

$[R] = \mathbb{R} \times \mathbb{R}^{Op}$

(There is a tension between the group structure on $\mathbb{R}$ and $[R]$ and the axioms of domains requiring $\bot$ and $\top$ elements which can be satisfied by restricting to a segment of reals, or by adding $-\infty$ and $+\infty$. I am mostly being ambiguous about this in this slide deck, but this is something to keep in mind.)

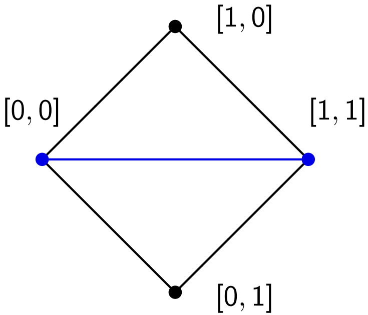
Partially inconsistent interval numbers within a segment



$[a, b] \leq [c, d]$ iff $a \leq c, b \leq d$

$[a, d] \sqsubseteq [b, c]$ iff $a \leq b, c \leq d$

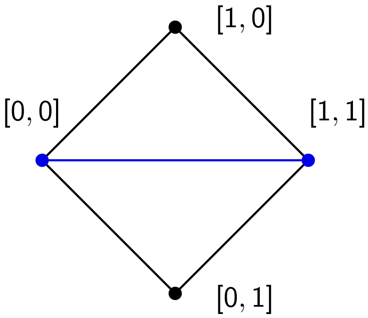
Partially inconsistent interval numbers within a segment



$[a,b] \leq [c,d]$ iff $a \leq c, b \leq d$

$[a,d] \sqsubseteq [b,c]$ iff $a \leq b, c \leq d$

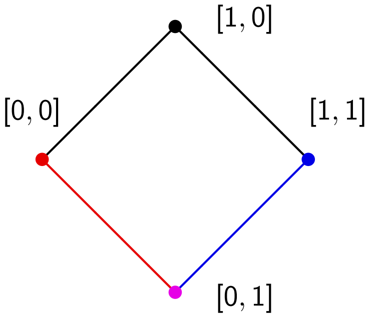
Partially inconsistent interval numbers within a segment



blue – precisely defined numbers

pseudosegments are above the blue

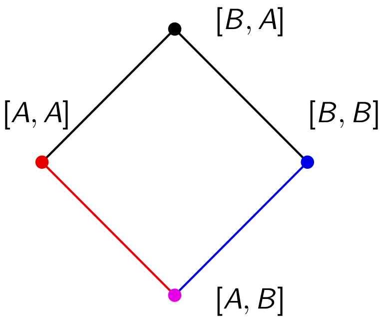
Negative and positive subspaces



Negative – space of upper bounds $[0, x]$

Positive – space of lower bounds $[x, 1]$

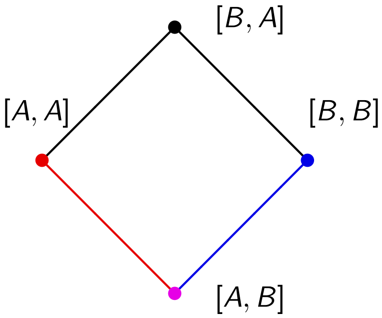
Negative and positive subspaces



Negative – space of upper bounds $[A, x]$

Positive – space of lower bounds $[x, B]$

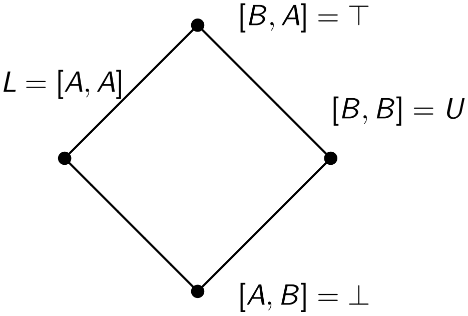
Negative and positive subspaces



We require $A < B$.

We can even allow $A = -\infty, B = \infty$.

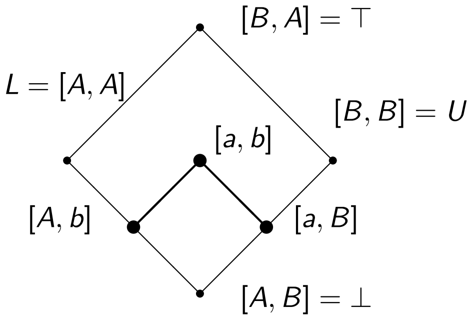
Negative and positive subspaces



We require $A < B$.

We can even allow $A = -\infty, B = \infty$.

Decomposition into negative and positive subspaces



$$[A, b] = [a, b] \wedge \bot = [a, b] \sqcap L$$
$$[a, B] = [a, b] \vee \bot = [a, b] \sqcap U$$
$$[a, b] = [A, b] \sqcup [a, B]$$

Lawvere duality

Order-theoretic duality between metric and logical structures.

(E.g. see Bukatin, Kopperman, Matthews, Some Corollaries of the Correspondence between Partial Metrics and Multivalued Equalities, *Fuzzy Sets and Systems* **256** (2014) 57-72.)

Lawvere duality

Quasi-metrics and partial metrics.
Fuzzy partial orders and multivalued equalities.

0 is the smallest possible distance.
0 is the maximal possible degree of equality.

Partial ultrametrics are valued in closed sets.
Ω-sets are valued in open sets.

But distances in the domain theory are hybrid between metric and logical, e.g. to ensure their Scott continuity, 0 must be the largest element among non-negative reals, etc.

Lawvere duality and the domain of arrows

For distances and orders/equalities **valued in arrow domains**, this duality takes an especially simple structure:

$$D^{Op} \times D \leftrightarrow D \times D^{op}.$$

Bitopology and *d*-frames

Achim Jung, M. Andrew Moshier. On the bitopological nature of Stone duality. Technical Report CSR-06-13. School of Computer Science, University of Birmingham, December 2006, 110 pages.

This text has a lot of very interesting material. I am only touching a bit of it here.

d-frames

Take two frames L_+ and L_- (the informal intent is for their elements correspond to open sets where the predicates are true and where they are false).

$L = L_+ \times L_-$ is a bilattice.

Introduce $Con, Tot \subseteq L$ with the informal intent that for pairs of open sets $U = \langle U_+, U_- \rangle$, $U \in Con$ when $U_+ \cap U_- = \emptyset$, and $U \in Tot$ when $U_+ \cup U_-$ covers the whole space.

This allows to handle partial inconsistency and the bilattice pattern does appear. (L_+, L_-, Con, Tot) is called a *d*-frame.

Bitopological Stone duality

This paper studies Stone duality modified to apply to bitopological spaces and *d*-frames.

It also demonstrates that a number of classical dualities, namely dualities of Stone, Ehresmann-Bénabou, and Jung-Sünderhauf, actually have bitopological nature, namely they are special cases of the Stone duality between bitopological spaces and *d*-frames.

d-frame for the (lower, upper) bitopology on $\mathbb{R}$

d-frame elements are pairs $\langle L, U \rangle$ of open rays, $\langle (-\infty, a), (b, +\infty) \rangle$ (a and b are allowed to take $-\infty$ and $+\infty$ as values).

Non-overlapping pairs of open rays are consistent ($a \leq b$), overlapping pairs of open rays ($b < a$) are total.

Correspondence with partially inconsistent interval numbers

The bilattice isomorphism between d-frame elements and partially inconsistent interval numbers with "infitinity crust":
$\langle(-\infty, a), (b, +\infty)\rangle$ corresponds to a partially inconsistent interval number $[a, b]$.

Consistent, i.e. non-overlapping, pairs of open rays ($a \leq b$) correspond to segments. Total, i.e. covering the whole space, pairs of open rays ($b < a$) correspond to pseudosegments.

Correspondence with partially inconsistent interval numbers

The bilattice isomorphism between d-frame elements and partially inconsistent interval numbers with "infitinity crust": $\langle(-\infty, a), (b, +\infty)\rangle$ corresponds to a partially inconsistent interval number $[a, b]$.

Consistent, i.e. non-overlapping, pairs of open rays ($a \leq b$) correspond to segments. Total, i.e. covering the whole space, pairs of open rays ($b < a$) correspond to pseudosegments.

In the context of Lawvere duality: interval numbers are similar to closed sets, but are ordered by reverse inclusion.

Lawvere duality and $[R]$

In the context of Lawvere duality: interval numbers are similar to closed sets, but are ordered by reverse inclusion as a domain. Cf. distances in the domain theory:

0 is the smallest possible distance.
0 is the maximal possible degree of equality.

Partial ultrametrics are valued in closed sets.
Ω-sets are valued in open sets.

But distances in the domain theory are hybrid between metric and logical, e.g. to ensure their Scott continuity, 0 must be the largest element among non-negative reals, etc.

Pseudosegments and negative set membership

Consider $\mathbb{R}$ as a characteristic function, and subtract from it characteristic functions of $(-\infty, a)$ and $(b, +\infty)$.

If $a \leq b$, we get the usual charactertic function for $[a, b]$.

However if $b < a$, we get the generalized characteristic function which takes value -1 on (a, b) and 0 elsewhere.

Topological asymmetry

Algebraically we can say that totally defined numbers $[a, a]$ belong to both segments and pseudosegments, or to neither.

But topologically (and via characteristic functions), this symmetry must be broken.

We brake it in favor of the “natural” viewpoint: totally defined numbers are segments, and not pseudosegments.

But one can brake it in favor of the dual viewpoint, by considering dual d-frames of closed sets (and stipulating that characteristic functions of segments take value 1 only on their interiors).

Rodabaugh representation

S. Rodabaugh, Functorial Comparisons of Bitopology with Topology and the Case for Redundancy of Bitopology in Lattice-valued Mathematics, Applied General Topology **9**(1), 77–108 (2008)

L-valued bitopology can be understood as L^2-valued topology, and, in particular, that ordinary bitopology can be understood as 4-valued topology. The 4-valued set here is the standard bilattice of 4 elements playing the same role in bitopology as the Sierpinski space plays in topology.

The L^2 in general is also a bilattice, with $\sqsubseteq$ being obtained from the product $(L, \sqsubseteq) \times (L, \sqsubseteq)$ and the material order, $\leq$, being obtained from the product of the dual lattice by the original one, $(L, \sqsupseteq) \times (L, \sqsubseteq)$.

Real-valued bitopology (hand-waved)

Hence fuzzy bitopology valued in $\mathbb{R}$ can be represented as a fuzzy topology valued in $[R]$

.
Let's make this construction more crisp in the next few slides.

(L, M)-valued bitopology

Slight generalization: (L, M)-valued bitopology can be understood as $L \times M$-valued topology.

(Via $L^X \times M^X \cong (L \times M)^X$ isomorphism.)

(lower,upper)-valued bitopology

Consider the lower and upper topologies on $\mathbb{R}$.

The (lower, upper)-valued bitopology can be understood as the topology valued in the (lower, upper) d-frame, hence the topology valued in $[R]$ (with the "infinity crust").

Let's look a bit more at the intuition behind the choice of (lower,upper)-valued bitopology.

Intuition behind (lower,upper)-valued bitopology

Consider ordinary (crisp) bitopology and one of its topologies.

For a topology open sets correspond to their characteristic functions, which are continuous functions into the Sierpinski space.

These characteristic functions are monotonic with respect to the specialization order.

The bitopological cases of interest tend to have the opposite specialization orders for the topologies involved (in particular, this is so for bicontinuous domains).

Intuition behind (lower,upper)-valued bitopology

In these cases, it is natural to identify pairs of open sets with pairs of characteristic functions to "Sierpinski spaces with opposite orders".

Fuzzyfication of this situation leads to (lower,upper)-valued bitopologies.

This suggests possibilities of studying how various constructions from [Jung-Moshier] would work for *L*-valued topological systems studied by Denniston, Melton, and Rodabaugh.

Another route to bitopology

A bitopology with two specialization orders pointing in the opposite directions is what seems to be required to handle antimonotonic functions well.

The group negation would typically be a pairwise continuous function from (X, T, T^{-1}) to (X, T^{-1}, T),
where T^{-1} is a group dual topology of T.

S. Andima, R. Kopperman, P. Nickolas, An Asymmetric Ellis Theorem, Topology and Its Applications **155**, 146–160 (2007)

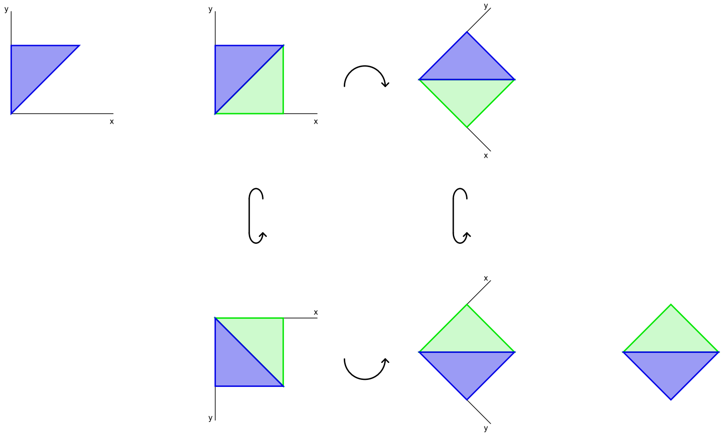
Antimonotonic bicontinuous group inverse

$- : \mathbb{R} \to \mathbb{R}$.

The minus is bicontinuous from the (lower, upper) bitoplogy to the (upper, lower) bitopology and vice versa.

The corresponding inverse image map between the d-frames is very similar to the weak minus on $[R]$ (Ginsberg involution), except that the order of bitopological components also needs to be swapped to respect bitopological duality ($\mathbb{R} \times \mathbb{R}^{Op} \leftrightarrow \mathbb{R}^{Op} \times \mathbb{R}$).

From Cartesian to Hasse representation



Antimonotonic bicontinuous group inverse

In a simular fashion, the true minus operation on $[R]$ is bicontinuous between a (T, T^{-1}) bitopology on $[R]$ and its dual (T^{-1}, T) bitopology, and vice versa.

Here T and T^{-1} must be group dual topologies of each other.

The main case: T is the Scott topology corresponding to $\sqsubseteq$, and T^{-1} is the Scott topology corresponding to $\sqsupseteq$.

Negative self-distance

The standard partial metric on the interval numbers is
$p([a_1, b_1], [a_2, b_2]) = \max(b_1, b_2) - \min(a_1, a_2)$.

Hence the self-distance for $[a, b]$ is $b - a$.

If we extend this formula to pseudosegments, the self-distance of pseudosegments turns out to be negative.

Weak vs strong axioms

Partial metrics can be understood as upper bounds for "ideal distances".

One often has to trade the tightness of those bounds for nicer sets of axioms.

E.g. the natural upper bound for the distance between $[0, 2]$ and $[1, 1]$ is 1, and there is a weak partial metric which yields that.

However, if one wants to enjoy the axiom of small self-distances, $p(x, x) \leq p(x, y)$, one has to accept $p([0, 2], [1, 1]) = 2$, since $p([0, 2], [0, 2]) = 2$.

Lower bounds

A similar trade can be made for lower bounds. The standard interval-valued relaxed metric produces the gap between non-overlapping segments as their lower bound, but takes 0 as the lower bound for the distance between overlapping segments (hence 0 is also the lower bound for self-distance).

If one settles for a less tight lower bound and allows the lower bound to be negative in those cases, one can obtain a distance with much nicer properties:

$l([a_1, b_1], [a_2, b_2]) = \max(a_1, a_2) - \min(b_1, b_2).$

An $[R]$-valued distance on $[R]$

We think about the pair $\langle l, p \rangle$ as a relaxed metric valued in $[R]$.

The self-distance of $[a, b]$ is $[a - b, b - a]$ and the self-distance of a pseudosegment is a pseudosegment.

The map $[a, b] \mapsto [b, a]$ expressing the symmetry between segments and pseudosegments also transforms $\langle l, p \rangle$ into $\langle p, l \rangle$.

Theoretical open problems: domain equations

First steps towards higher-order theory are made in [Kozen, Semantics of Probabilistic Programs] and also in [Keimel, Bicontinuous Domains and Some Old Problems in Domain Theory].

However, in this context I have not seen anything coming close to the solution of domain equations, such as $D \cong [D \to D]$.

If one follows the appoach by Kozen, where programs denote linear operators, and if one focuses on reversible programs, what seems to be called for here is **applicative representation theory**.

Theoretical open problems: type theory

The type theory when one has D^{Op} constructor, and when the unusual form of polymorphism is present: the group minus has both the type $D \times D^{Op} \to D^{Op} \times D$ and the type $D^{Op} \times D \to D \times D^{Op}$.

This should be done together with the usual set of constructors and the ability to define recursive types (i.e. solve domain equations).

Applied hopes

Consider denotational semantics: $[\![]\!] : \{\text{Programs}\} \to \{\text{Meanings}\}$.

If the space of meanings is sufficiently rich to allow standard methods of applied math, one would hope to be able to pull back those methods from meanings to programs.

In particular, one would hope to find better ways to solve machine learning problems over spaces of programs (this is known as the problem of **program learning** or **symbolic regression**).

The choice of a domain with Scott topology as the space of meaning seems attractive in this sense, because of the richness of domain theory.

Applied hopes

Two main obstacles for this use of domain theory seem to be:

- defective linear algebra (e.g. for ordinary interval numbers)
- inability to pull the constructions made in domains back to the realm of programs in the ways which would be computationally effective.

Linear models seem to fix the first of these problems.

To address the second of these problems we'll focus on software architectures which allow to take linear combinations of actual computational processes, rather than just of program meanings.

Kozen semantics

One can think about probabilistic programs as transformers from the probability distributions on the space of inputs to the probability distributions on the space of outputs.

It is fruitful to replace the space of probability distributions by the space of signed measures.

D. Kozen, Semantics of Probabilistic Programs, Journal of Computer and System Sciences **22** (3), 328–350 (1981)

Kozen semantics

One defines $\nu < \mu$ iff $\mu - \nu$ is a positive measure.

The space of signed measures is a vector lattice (a Riesz space) and a Banach space, so people call this structure a Banach lattice.

Denotations of programs are continuous linear operators with finite norms.

The probabilistic powerdomain is embedded into the positive cone of this Banach lattice.

Hahn-Jordan decomposition

$\mu^+ = \mu \vee 0, \mu^- = \mu \wedge 0, \mu = \mu^+ + \mu^-$, holds, since it's a theorem for all lattice-ordered groups.

Defining $\nu \sqsubseteq \mu$ iff $\nu^+ \leq \mu^+$ and $\nu^- \leq \mu^-$, one also obtains $\mu = \mu^+ \sqcup \mu^-$, making this an instance of the "bilattice pattern".

Linear combinations of programs

Take $0 < \alpha < 1$ and **random** being a generator of uniformly distributed reals between 0 and 1.

if random $< \alpha$ **then P else Q** yields a linear combination of programs **P** and **Q**.

To allow negative coefficients one needs to consider computing a negative and a positive channel in parallel (computations are marked as negative or positive).

But one cannot obtain linear combinations of single runs under this approach (because only one sample from a distribution is taken during a particular run).

Sampling semantics

The situations when one can consider linear combinations of single execution runs should be especially attractive.

Sampling semantics is one possibility here – the input is literally a sampling of a probability distribution, and so is the output, and the whole thing is, in some sense, a "probabilistic dataflow architecture".

To implement linear combinations of probabilistic programs with positive coefficients one can simply execute those programs in parallel, and the values of coefficients can be controlled by changing the relative execution speed of those programs.

To allow negative coefficients one again needs to use a negative and a positive channel

Sampling semantics

What is the right way to talk about higher-order probabilistic programming?

The tradition is to talk about probabilistic lambda-calculus, but I am not sure it is a convenient formalism.

In any case, higher-order probabilistic programming is the ability to take samplings of probabilistic programs as inputs and to produce samplings of probabilistic programs as outputs.

We'll talk about this when we discuss "sampling the samplers" approach to program learning.

Fuzzy sampling and animations

Fuzzy samplings where points are taken with real coefficients might be even more attractive.

One can think about them as generalized animations, where points might be indexed by a more sophisticated index set than a discretized rectangle.

Here one might allow negative coefficients and avoid the need for a separate channel (speaking in terms of animation this means that 0 is at some grey level, between black and white).

One can leverage existing animations, digital and physical (such as light reflections and refractions in water), as computational oracles.

Expressive power

Music is a fast animation (typically on the index set of 2 points for usual stereo).

Very short programs can express complex dynamics.

A way to incorporate aesthetic criteria into software systems.

Probabilistic sampling and evolutionary programming

The connections between probabilistic programming and evolutionary/genetic programming are much tighter than it is usually acknowledged.

MCMC is essentially an evolutionary method:

- acceptance/rejection of the samples corresponds to selection
- production of new samples via modifications of the accepted ones corresponds to mutations to produce offspring from the survivors.

Probabilistic sampling and evolutionary programming

Bayesian Optimization Algorithm changes the procedure of producing the next generation in genetic algorithms from pairwise crossover to the resampling from the estimated distribution of the individuums selected for fitness.

Martin Pelikan. Bayesian Optimization Algorithm: from Single Level to Hierarchy, PhD Thesis 2002.
`http://www.medal-lab.org/files/2002023.pdf`

Used by the seminal
Moshe Looks. Competent Program Evolution, PhD Thesis 2006.
`http://metacog.org/doc.html`

Probabilistic sampling and evolutionary programming

A weakness of the known genetic programming schemes seems to be that none of them seems to implement the regulation of gene expression.

Variability in the regulation of gene expression, rather than in genes themselves, seems to be an important factor making fast biological evolution feasible.

If we associate a protein with a computational process, then one might want an architecture where proteins correspond to parallel computational processes, and the degree of expression of a given protein corresponds to the share of computational resources the computational process in question gets.

Linear models are especially nice in this sense: the degree of expression of a gene can be simply modelled via the corresponding coefficient in the linear combination of computational processes corresponding to a parallel program (set of genes).

Hybrid systems

Instead of implementing everything in terms of linear systems one can use a hybrid approach, mixng linear sytems and traditional software.

Inspiration: hybrid hardware connecting live neural tissue and electronic circuits.

One can decide to use large existing software components and try to automate the process of connecting them together using flexible probabilistic pieces. This is potentially very important.

One can try to use small inflexible components inside the flexible "tissue" of linear models.

MCMC sampling in deep learning

Oliver Woodford. **Notes on Contrastive Divergence.**

http://www.robots.ox.ac.uk/∼ojw/files/NotesOnCD.pdf

A conference dedicated to MCMC sampling

Fifth IMS-ISBA joint meeting

MCMSki IV

6 - 8 January 2014

http://www.pages.drexel.edu/∼mwl25/mcmski/program.html

Electronic textbook on probabilistic programming

https://probmods.org

N. D. Goodman and J. B. Tenenbaum (electronic).
Probabilistic Models of Cognition.
Retrieved on Oct 28, 2014 from http://probmods.org

Contains code samples in WebChurch which can be edited and executed in a Web browser.

Sampling the samplers

http://arxiv.org/abs/1407.2646

Yura N. Perov, Frank D. Wood.
Learning Probabilistic Programs. July 9, 2014.

- A notion of compilation for probabilistic program (more similar to partial evaluation).
- **Anglican** engine (PMCMC, Clojure)
- Maddison-Tarlow paper

Sampling the samplers

http://cims.nyu.edu/∼brenden/LakePhDThesis.pdf

Brenden M. Lake.
Towards more human-like concept learning in machines: Compositionality, causality, and learning-to-learn.
MIT PhD Thesis, September 2014.

- Learning from one or a few examples
- Learning rich conceptual representations

A bit more about animations

Probabilistic programming is better if the goal is well-defined, animations are better if one wants to explore emergent behavior.

Very short programs can express complex dynamics.

A large and very active **creative coding community**.
Students love this subject.

I am going to show a few short demos using Processing (`processing.org`).

Electronic coordinates

These slides are linked from my page on partial inconsistency and vector semantics of programming languages:

http://www.cs.brandeis.edu/~bukatin/partial_inconsistency.html

E-mail:

bukatin@cs.brandeis.edu