Hyper-distributions:
What are they?
And why should you care?

Carroll Morgan
Annabelle McIver
Tahiry Rabehaja

University of New South Wales / NICTA
Macquarie University

Quantitative Information Flow Day
Princess Workshop
16 December 2010

Starting point

xs is a two-bit sequence, thus with four possible values.

$xs := \text{two bits chosen uniformly}$
$xs := xs \oplus -xs$

$\oplus$ means 50/50 probabilistic choice

If you had to guess xs's value *after* the above program had been run, what would your best strategy be?

Conventional approach

"It's just a Markov chain."

$X = \{00, 01, 10, 11\}$

program has type $\mathbb{D}X \to \mathbb{D}X$

$(;)$ is matrix multiplication

$$
\begin{array}{cccc} 00 & 01 & 10 & 11 \\ \left(\frac{1}{4}\right. & \frac{1}{4} & \frac{1}{4} & \left.\frac{1}{4}\right) \end{array}
\quad
\begin{array}{c|cccc} & 00 & 01 & 10 & 11 \\ \hline 00 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \\ 01 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 10 & 0 & \frac{1}{2} & \frac{1}{2} & 0 \\ 11 & \frac{1}{2} & 0 & 0 & \frac{1}{2} \end{array}
=
\begin{array}{cccc} 00 & 01 & 10 & 11 \\ \left(\frac{1}{4}\right. & \frac{1}{4} & \frac{1}{4} & \left.\frac{1}{4}\right) \end{array}
$$

xs := uniform

$xs := xs \oplus -xs$

still uniform

Monadic approach

$X = \{00, 01, 10, 11\}$

Program has type $X \to \mathbb{D}X$

$(;)$ is Kleisli composition

$00 \longmapsto \frac{1}{2} \quad 0 \quad 0 \quad \frac{1}{2}$

$01 \longmapsto 0 \quad \frac{1}{2} \quad \frac{1}{2} \quad 0$

$10 \longmapsto 0 \quad \frac{1}{2} \quad \frac{1}{2} \quad 0$

$11 \longmapsto \frac{1}{2} \quad 0 \quad 0 \quad \frac{1}{2}$

What's the connection between the two approaches?

RE – Starting point

```
xs := two bits chosen uniformly
| i := $0 \oplus 1$
| print $xs_i$
```

i is a local variable, initialised randomly

If you had to guess xs's value after the above program had been run, what would your best strategy be Now?

Maximum a.posteriori probability (MAP)

Information leak is a channel

$00 \longmapsto$	1	0
$01 \longmapsto$	$\frac{1}{2}$	$\frac{1}{2}$
$10 \longmapsto$	$\frac{1}{2}$	$\frac{1}{2}$
$11 \longmapsto$	0	1

print 0

print 1

$\frac{1}{2}$	$\frac{1}{2}$
$\frac{1}{2}$	0
$\frac{1}{4}$	$\frac{1}{4}$
$\frac{1}{4}$	$\frac{1}{4}$
0	$\frac{1}{2}$

If you see 0, guess 00

If you see 1, guess 11.

RE-RE- Starting point

```
$xs :=$ two bits chosen uniformly
|   $i := 0 \oplus 1$
|   print $xs_i$
$xs := xs \oplus -xs$
```

If you had to guess xs's value <u>after</u> the above program had been run, what would your best strategy be

THIS TIME?!

Conventional approach

Monadic approach

"It's just a Hidden Markov Model (HMM)."

It's the same monad, used one level up.

Conventional

Markov matrix $M_{x,x'}$

HMM matrix $H_{x,y,x'}$

Monadic

$X \longrightarrow \mathbb{D}X$

$\mathbb{D}X \longrightarrow \mathbb{D}^2 X$

HMM's combine Markovs and channels.

Same Markov as before

$00 \longmapsto \frac{1}{2} \quad 0 \quad 0 \quad \frac{1}{2}$

$01 \longmapsto 0 \quad \frac{1}{2} \quad \frac{1}{2} \quad 0$

$10 \longmapsto 0 \quad \frac{1}{2} \quad \frac{1}{2} \quad 0$

$11 \longmapsto \frac{1}{2} \quad 0 \quad 0 \quad \frac{1}{2}$

Information leak is a channel

$00 \longmapsto 1 \quad 0$

$01 \longmapsto \frac{1}{2} \quad \frac{1}{2}$

$10 \longmapsto \frac{1}{2} \quad \frac{1}{2}$

$11 \longmapsto 0 \quad 1$

print 0

print 1

```
xs := two bits chosen uniformly
i := 0 ⊕ 1
print $xs_i$
xs := xs ⊕ -xs
```

i is a local variable

|[var i · · ·]|

What does the print statement reveal about the final value of xs?

The mathematical structure underlying this is the HMM (hidden Markov model).

HMM-reasoning is conventionally done by matrix-wrangling.

HMM's — a summary

just xs (no i)

$\mathcal{X}$ — state space

$\mathcal{Y}$ — observation space

anything with at least 2 values

↑ is matrix type

$H : \mathcal{X} \to \mathcal{Y} \times \mathcal{X}$

output state space

observation space

input state space

x, x′ ∈ $\mathcal{X}$

y ∈ $\mathcal{Y}$

$H_{x,y,x'}$ — the probability that input x prints y and goes to new state x′.

Pure Markov transition is $H_{x,y,x'} = M_{x,x'}$

Write it $(:M)$

Pure channel transmission is

Write it $(C:)$

$$H_{x,y,x'} = \begin{cases} C_{x,y} & \text{if } x'=x \\ 0 & \text{otherwise} \end{cases}$$

Elementary HMM is

Write it $(C:M)$

$$H_{x,y,x'} = C_{x,y} * M_{x,x'}$$

Composition

compound observation

$$(H^1 ; H^2)_{x,\overbrace{y_1, y_2},x'} = \sum_{x''} H^1_{x,y_1,x''} * H^2_{x'',y_2,x'}$$

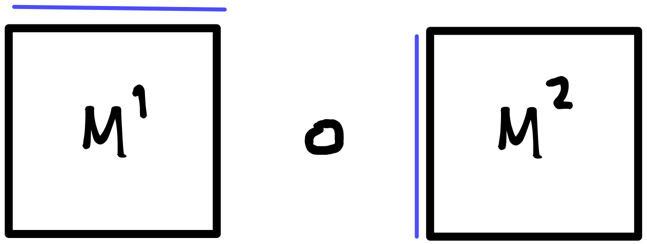
Then $(C:M) = (C:)\,;\,(:M)$

and $(:M^1)\,;\,(:M^2) = (:M^1 \circ M^2)$

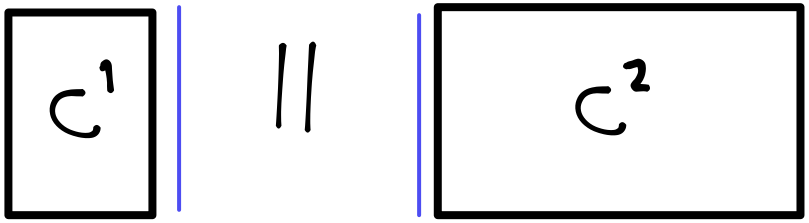
and $(C^1:)\,;\,(C^2:) = (C^1 \,||\, C^2 :)$

nice special cases!

matrix multiplication



matrix concatenation



Transition seen monadically (well known)

X	state
$\mathbb{D}$	type constructor
$\mathbb{D}X$	distributions on X
$X \to \mathbb{D}X$	programs
$\mathbb{D}X \to \mathbb{D}X$	Kleisli extension Markov chain

sequential composition respected

; between matrices maps to Kleisli composition in the monad

Transmission seen monadically (not well known)

$\mathcal{X} = \{0, 1, 2\}$

$\mathcal{Y} = \{0, 1\}$

Choose x uniformly
print x $\underline{\text{mod}}$ 2

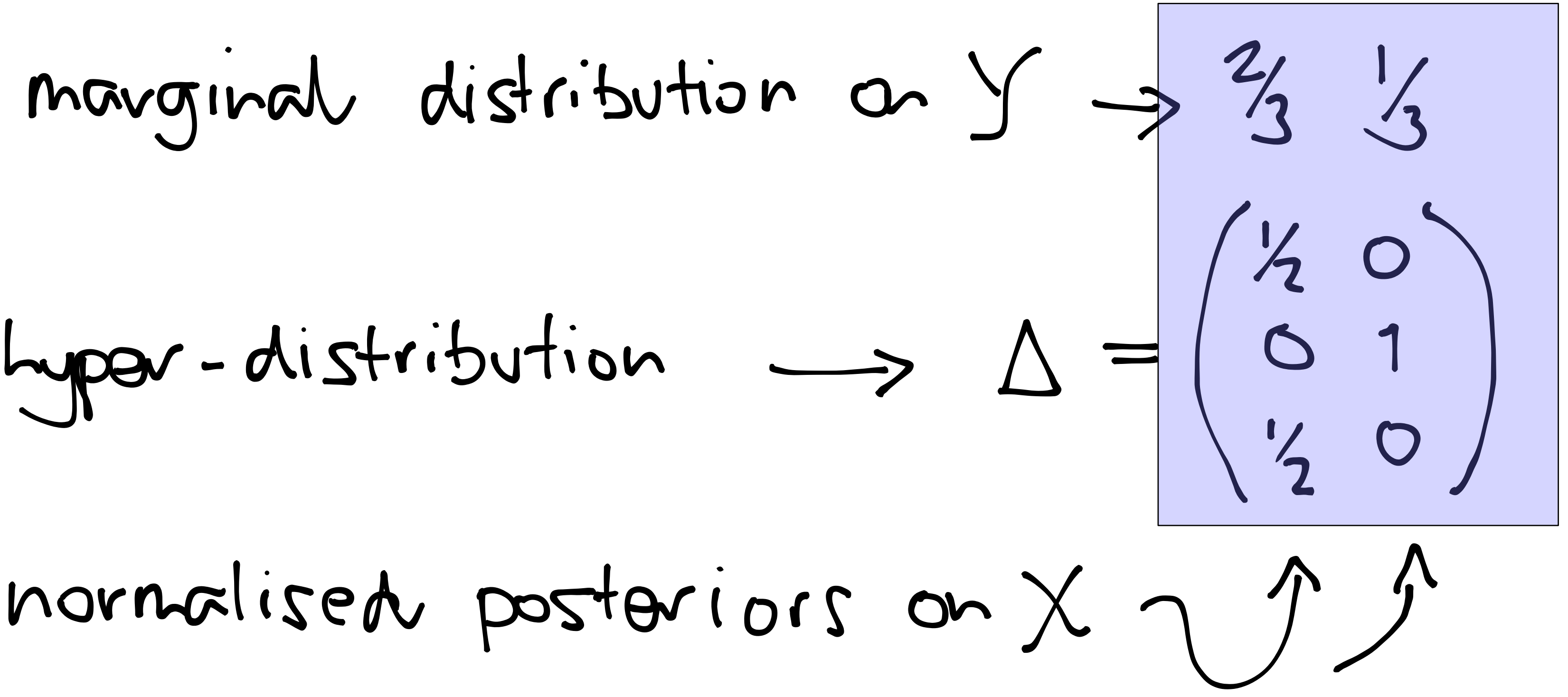
$$C = \begin{pmatrix} 1 & 0 \\ 0 & 1 \\ 1 & 0 \end{pmatrix}$$

deterministic channel

$$J = \begin{array}{cc} & y \\ x & \begin{pmatrix} 1/3 & 0 \\ 0 & 1/3 \\ 1/3 & 0 \end{pmatrix} \end{array}$$

HMM omitting x'

Hyper-distributions $\mathbb{D}^2\mathcal{X}$



marginal distribution on $\mathcal{Y}$ → $\frac{2}{3} \quad \frac{1}{3}$

hyper-distribution → $\Delta = \begin{pmatrix} \frac{1}{2} & 0 \\ 0 & 1 \\ \frac{1}{2} & 0 \end{pmatrix}$

normalised posteriors on $\mathcal{X}$

An abstract channel has type $\mathbb{D}\mathcal{X} \longrightarrow \boxed{\mathbb{D}^2\mathcal{X}}$

Distribution → Hyper

HMM seen monadically

```
xs := two bits chosen uniformly
$i := 0 \oplus 1$
print $xs_i$
$xs := xs \oplus -xs$
```

uniform initially

1	½	½	½	½
¼	½	0	¼	¼
¼	¼	¼	¼	¼
¼	¼	¼	¼	¼
¼	0	½	¼	¼

1	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	$\frac{1}{2}$	1
$\frac{1}{4}$	$\frac{1}{2}$	0	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$
$\frac{1}{4}$	0	$\frac{1}{2}$	$\frac{1}{4}$	$\frac{1}{4}$	$\frac{1}{4}$

the merge is automatic

π

input prior

Δ

output hyper

1
0
1/3
1/3
1/3

1/3	2/3
0	0
1/2	1/4
1/2	1/4
0	1/2

1/3	2/3
0	1/4
1/2	1/4
1/2	1/4
0	1/4
	Δ

An abstract HMM has type $\mathbb{D}X \longrightarrow \boxed{\mathbb{D}^2 X}$

A brief guide to hyper-space

1. Based on a finite set X
2. It is $\mathbb{D}^2 X$, i.e. distributions of dist's.
3. It has a partial order $\sqsubseteq$ of refinement, where $S \sqsubseteq I$ means

   I is functionally equivalent to S, but reveals less.

The refinement order of security (an aside)

For two hypers $\Delta_S, \Delta_I : \mathbb{D}^2 X$

$$\Delta_S \sqsubseteq \Delta_I$$

means you can reach Δ_I from Δ_S by merging posteriors.

This preserves the overall functionality, but "forgets" which posterior(s) were responsible.

A brief guide (continued)

4. That order generalises Landauer and Redmond's lattice of information — but is not a lattice.

5. $\mathbb{D}^2 X$ is not chain-complete under $\sqsubseteq$, but can be completed using measures.

6. $\mathbb{D}^2 X$ admits the Kantorovich metric, which as a distance between hypers is related to the difference in how much they reveal.

A brief guide (continued)

7. The analogue of predicates, on hypers, is "uncertainty measures" that generalise entropies (like Shannon's).

8. Uncertainties are concave, continuous functions in $\mathbb{D}X \to \mathbb{R}^{\geq}$.

9. An uncertainty u is applied to Δ by taking the expected value $\mathcal{E}_{\Delta} u$.

Uncertainty orders (an aside)

Shannon entropy is an example

$$\pi \longmapsto \sum_x \pi_x \cdot \mathrm{l}\bar{\mathrm{g}}\, \pi_x$$

is continuous and concave

Another example is Bayes Risk: $\pi \mapsto 1 - \bigsqcup_x \pi_x$

Another example is "g co-vulnerability", for any $g \geq 0$:

$$\pi \mapsto \bigsqcap_\omega \sum_x \pi_x g_\omega(x) .$$

A brief guide (continued)

10. Fundamental refinement theorem ("Coriaceous"):

$$\Delta_1 \sqsubseteq \Delta_2 \text{ iff } \mathcal{E}_{\Delta_1} u \leq \mathcal{E}_{\Delta_2} u$$

for all uncertainties u

A brief guide (continued)

11. Given an abstract hyper

$$h : \mathbb{D}X \to \mathbb{D}^2 X$$

define

$$wp.h.u.\pi := \mathcal{E}_{h.\pi}\, u$$

Thus $wp.h$ is an

"uncertainty transformer".

A brief guide (continued)

12. There are "healthiness conditions ...".

For example, it must be shown that if u is continuous and concave, then so is wp.h.u.

Not true for individual entropies!

But that is enough for now.