The Semantics of Evidential Reasoning

Riccardo Pucella
Northeastern University

Harvard PL Seminar

April 18, 2007

The Problem

Inference on the basis of observations

Suppose Alice has two coins

- A fair coin (Pr heads = 1/2)
- A biased coin (Pr heads = 99/100)

She chooses one of the coins and tosses it 100 times

- It lands heads every time

What is the probability that the coin is biased?

- No probability on choice!

The Problem

Inference on the basis of observations

Suppose Alice has two coins

- A fair coin (Pr heads = 1/2)
- A biased coin (Pr heads =

She chooses one of the coins a

- It lands heads every time

But it still seems more "likely" that the coin is biased

What is the probability that the coin is biased?

- No probability on choice!

Evidence

The outcome of an experiment (Heads, Tails) provides evidence for an hypothesis (Fair, Biased)

Many approaches to encoding evidence are based on comparing likelihood functions

$\mu_h(ob)$ = probability of observing *ob* if *h* holds

- μ_{Biased} (Heads) = 99/100
- μ_{Fair} (Heads) = 1/2

Shafer's Weight of Evidence

An evidence space $E = (H, O, \boldsymbol{\mu})$ captures all the relevant information:

- H : hypotheses $\{h_1, ..., h_n\}$
- O : observations
- $\boldsymbol{\mu}$: H-indexed likelihood functions

Weight of evidence of observation *ob* for hypothesis *h*

$$w_E(ob, h) = \frac{\mu_h(ob)}{\mu_{h_1}(ob) + \cdots + \mu_{h_n}(ob)}$$

Revisiting The Puzzles

E = ({ Fair, Biased}, {100Heads,< 100Heads}, $\boldsymbol{\mu}$)

μ	Fair	Biased
100 Heads	2^{-100}	0.37
< 100 Heads	$1-2^{-100}$	0.63

w_E	Fair	Biased
100 Heads	~ 0	~ 1
< 100 Heads	~ 0.61	~ 0.39

Justifying Evidence

Consider the original example

Suppose we have a probability on Alice's choice

- Then probability of coin being biased after seeing Heads can be computed by Bayes' Theorem:

$$\Pr(Biased \mid Heads) = \frac{\Pr(Heads \mid Biased)\Pr(Biased)}{\Pr(Heads)}$$

Evidence captures the information used in this update...

Updating Via Evidence

Can use weight of evidence function $\lambda x.w_E(ob,x)$ to update prior μ_0 to a posterior μ_{ob}:

$$\mu_{ob} = \mu_o \oplus (\lambda x.w_E(ob, x))$$

$$(f_1 \oplus f_2)(x) = \frac{f_1(x)f_2(x)}{\sum_{y \in X} f_1(y)f_2(y)}$$

Theorem: If Pr(h) = $\mu_0(h)$, then $\mu_{ob}(h)$ = Pr (h | ob)

Application:

Rabin's primality test for a number n

- Polynomial-time computable predicate $P(n,a)$:
  - n composite $\Rightarrow$ $P(n,a)=1$ for $\geq n/2$ choices of a
  - n prime $\Rightarrow$ $P(n,a)=0$ for all a

Pick a number a at random between 0 and n
If $P(n,a) = 1$ return "composite"; otherwise return "prime"

w	prime	composite
"prime"	≥2/3	≤1/3
"composite"	0	1

A More Complex Experiment

Suppose Alice has two coins

- Biased (Pr heads = 99/100)
- Slightly biased (Pr heads = 3/4)

Suppose Bob has a fair coin (Pr heads = 1/2)

- Alice and Bob each hand a coin to Zoe
- Zoe tosses one of the coins
- Outcome is heads

What is the evidence that Zoe chose Alice's coin?

One Approach

Consider evidence spaces arising from all possible (sequences of) choices

- Agent ascribing weight of evidence does not know which is the right one evidence space
- Compute weights of evidence with respect to each evidence space
- Get a range of weights of evidence
  - One per possible sequence of choices

Generalized Weight of Evidence

Generalized evidence space $G = (H,O,\Delta)$

- H : hypotheses
- O : observations
- Δ : set of H-indexed likelihood functions

$$S(G) = \{ (H,O,\boldsymbol{\mu}) : \boldsymbol{\mu} \in \Delta \}$$

$$w_G(ob,h) = \{ w_E(ob,h) : E \in S(G) \}$$

The minimal and maximal elements of $w_G(ob,h)$ are often useful

Updating Priors

Can update prior μ_0 by natural generalization:

$$\mathcal{P}_{ob} = \{\mu_0 \oplus w_E(ob, -) : E \in S(G)\}$$

This gives the set of all possible posteriors based on choices

Theorem: Bounds on posteriors agree with taking all possible results of Bayes' Theorem applied to the prior

What about Repeated Tosses?

Consider the Alice/Bob/Zoe scenario with 100 reps

- Zoe always makes same choice
- Difference is whether Alice gets to choose again
  - Same coin repeated?
  - Different coin chosen at every toss?

How do we derive evidence spaces here?

The First Experiment

Given hyp ∈ { AliceCoin, BobCoin }

Alice chooses [Biased, SlightlyBiased]
Bob chooses Fair
if hyp is AliceCoin then
  Zoe chooses Alice's coin
else
  Zoe chooses Bob's coin
Zoe tosses coin 100 times

The First Experiment

Given hyp ∈ { AliceCoin, BobCoin }

Alice ← **choose** [{ H:0.99, T:0.01 }, { H:0.75, T:0.25 }];
Bob chooses Fair
if hyp is AliceCoin then
  Zoe chooses Alice's coin
else
  Zoe chooses Bob's coin
Zoe tosses coin 100 times

The First Experiment

Given hyp ∈ { AliceCoin, BobCoin }

Alice ← **choose** [{ H:0.99, T:0.01 }, { H:0.75, T:0.25 }];
Bob ← { H:0.5, T:0.5 };
if hyp is AliceCoin then
  Zoe chooses Alice's coin
else
  Zoe chooses Bob's coin
Zoe tosses coin 100 times

The First Experiment

Given hyp ∈ { 1, 2 }

Alice ← **choose** [{ H:0.99, T:0.01 }, { H:0.75, T:0.25 }];
Bob ← { H:0.5, T:0.5 };
if hyp = 1 **then**
  Zoe chooses Alice's coin
else
  Zoe chooses Bob's coin
Zoe tosses coin 100 times

The First Experiment

Given hyp ∈ { 1, 2 }

Alice ← **choose** [{ H:0.99, T:0.01 }, { H:0.75, T:0.25 }];
Bob ← { H:0.5, T:0.5 };
if hyp = 1 **then**
  Zoe ← Alice
else
  Zoe ← Bob;
Zoe tosses coin 100 times

The First Experiment

Given hyp ∈ { 1, 2 }

Alice ← **choose** [{ H:0.99, T:0.01 }, { H:0.75, T:0.25 }];
Bob ← { H:0.5, T:0.5 };
if hyp = 1 **then**
  Zoe ← Alice
else
  Zoe ← Bob;
loop 100 **do observe** Zoe

The Second Experiment

Given hyp ∈ { 1, 2 }

loop 100 **do**
  Alice ← **choose** [{H:0.99,T:0.01}, {H:0.75,T:0.25}];
  Bob ← {H:0.5, T:0.5};
  if hyp = 1 **then**
    Zoe ← Alice
  else
    Zoe ← Bob;
  observe Zoe

A Language for Experiments

S ::=

$X \leftarrow P$

$x := E$

observe X

$S_1;S_2$

if B **then** S_1 **else** S_2

loop E **do** S

P ::=

$\{ob_1:r_1, \ldots, ob_n:r_n\}$

X

choose $[P_1, \ldots, P_n]$

A Language for Experiments

S ::=

X ← *P*

x := *E*

observe

S$_1$;*S*$_2$

if *B* **ther**

loop *E* **d**

E ::= some language of arithmetic expressions

P ::=

{*ob*$_1$:*r*$_1$, ..., *ob*$_n$:*r*$_n$}

X

choose [*P*$_1$, ..., *P*$_n$]

A Language for Experiments

S ::=

$X \leftarrow P$

$x := E$

observe X

$S_1;S_2$

if B **then** S_1 **else** S_2

loop E **do** S

P ::=

$\{ob_1:r_1, \ldots, ob_n:r_n\}$

X

choose $[P_1, \ldots, P_n]$

A Language for Experiments

S ::=

X ← *P*

x := *E*

observe *X*

$S_1;S_2$

if *B* **then** S_1 **else** S_2

loop

P ::=

{ob_1:

X

choose [P_1, ...,

B ::= some language of Boolean expressions

A Language for Experiments

S ::=

$X \leftarrow P$

$x := E$

observe X

$S_1;S_2$

if B **then** S_1 **else** S_2

loop E **do** S

P ::=

$\{ob_1:r_1, \ldots, ob_n:r_n\}$

X

choose $[P_1, \ldots, P_n]$

A Language for Experiments

S ::=

$X \leftarrow P$

$x := E$

observe X

$S_1;S_2$

if B **then** S_1 **else** S_2

loop E **do** S

P ::=

$\{ob_1:r_1, \ldots, ob_n:r_n\}$

X

choose $[P_1, \ldots, P_n]$

An experiment is a tuple

(H,O,S)

An Evidence-Based Semantics

Notation

- Prob(A) : probability distributions on set A
- Σ : maps from variables to integers
- Γ : maps from observables to Prob(O)

$[\![B]\!] : \Sigma \rightarrow$ {true,false}

$[\![E]\!] : \Sigma \rightarrow$ Integer

$[\![P]\!] : \Gamma \rightarrow \wp(\text{Prob}(O))$

$[\![S]\!] : \mathit{State} \rightarrow \wp(\mathit{State})$

where $\mathit{State} = \Sigma \times \Gamma \times \text{Prob}(O^*)$

Definitions of $[\![P]\!]$ and $[\![S]\!]$

$[\![\{ob_1{:}r_1,\ldots,ob_n{:}r_n\}]\!]\ (\gamma) = \{ \mu \}$ where $\mu(ob_i)=r_i$

$[\![X]\!]\ (\gamma) = \{ \gamma(X) \}$

$[\![$ **choose** $[P_1,\ldots,P_n]]\!]\ (\gamma) = \bigcup\ [\![P_i]\!]\ (\gamma)$

Definitions of $[\![P]\!]$ and $[\![S]\!]$

$[\![\{ob_1{:}r_1,\ldots,ob_n{:}r_n\}]\!]\ (\gamma) = \{\ \mu\ \}$ where $\mu(ob_i)=r_i$

$[\![X]\!]\ (\gamma) = \{\ \gamma(X)\ \}$

$[\![$ **choose** $[P_1,\ldots,P_n]]\!]\ (\gamma) = \bigcup\ [\![P_i]\!]\ (\gamma)$

This is where we get *sets* of probabilities

Definitions of $[\![P]\!]$ and $[\![S]\!]$

$[\![\{ob_1:r_1,\ldots,ob_n:r_n\}]\!]\ (\gamma) = \{ \mu \}$ where $\mu(ob_i)=r_i$

$[\![X]\!]\ (\gamma) = \{ \gamma(X) \}$

$[\![$ **choose** $[P_1,\ldots,P_n]]\!]\ (\gamma) = \bigcup\ [\![P_i]\!]\ (\gamma)$

$[\![X \leftarrow P]\!]\ (\sigma,\gamma,\mu) = \{(\sigma,\gamma[X \mapsto v],\mu) : v \in [\![P]\!]\ (\gamma)\}$

$[\![x := E]\!]\ (\sigma,\gamma,\mu) = \{(\sigma[x \mapsto [\![E]\!]\ (\sigma)],\gamma,\mu)\}$

$[\![$ **observe** $X]\!]\ (\sigma,\gamma,\mu) = \{(\sigma,\gamma,\mu') : \mu'(obs;ob)=\mu(obs)\gamma(X)(ob)\}$

$[\![S_1;S_2]\!]\ (\sigma,\gamma,\mu) = \bigcup\{ [\![S_2]\!]\ (\sigma',\gamma',\mu') : (\sigma',\gamma',\mu') \in [\![S_1]\!]\ (\sigma,\gamma,\mu)\}$

$[\![$ **if** B **then** S_1 **else** $S_2]\!]\ (\sigma,\gamma,\mu) =$

$[\![S_1]\!]\ (\sigma,\gamma,\mu)$ if $[\![B]\!]\ (\sigma) =$ true

$[\![S_2]\!]\ (\sigma,\gamma,\mu)$ if $[\![B]\!]\ (\sigma) =$ false

$[\![$ **loop** E **do** $S]\!]\ (\sigma,\gamma,\mu) =$ *(exercise...)*

Extracting an Evidence Space

Given an experiment (H,O,S):

$$G_S = (H,O^*,\Delta)$$

where $\boldsymbol{\mu} \in \Delta$ iff for all $h \in H$, there exists σ,γ s.t.

- $\sigma(\text{hyp}) = h$
- $(\sigma,\gamma,\boldsymbol{\mu}_h) \in [\![S]\!] (\sigma_0[\text{hyp}\mapsto h],\gamma_0,\mu_0)$

and

- σ_0 : sets every variable to 0
- γ_0 : sets every observable to uniform probability
- μ_0 : $\mu_0(<>) = 1$

Justifying the Semantics

Recall that evidence can be used to update a prior probability on hypotheses to a posterior probability on hypotheses, based on seeing a sequence of observations

Let S be a deterministic experiment

- No occurrence of **choose** [...]

Standard probabilistic semantics for language:

- $State = \Sigma \times \Gamma \times O^*$
- $P\ [\![S]\!]\ (s,A)$: probability that program S starting in state s terminates in a state in A

The Correspondence

Let (H,O,S) be a deterministic experiment

Let μ be a prior probability distribution on hypotheses

Theorem: The following probabilities are equal:

- $\mu \oplus w_E\,(obs,\text{-})$ where $S(G_S) = \{\, E \,\}$
- $\lambda h.(\mu(h)\ \mathsf{P}\ [\![S]\!]\ (\sigma_0[\mathrm{hyp}\mapsto h],\Sigma\times\Gamma\times\{obs\}))$

Proof is a fairly straightforward induction on S

General Experiments

What if S is a general experiment?

Still somewhat standard probabilistic semantics:

- State = $\Sigma \times \Gamma \times O^*$
- $T[\![S]\!](s,A)$: set of probabilities that program S starting in state s terminates in a state in A
- One probability for every possible nondeterministic sequence of choices

General Experiments

What if S is a general experiment?

Still somewhat standard probabilistic semantics:

- State = $\Sigma \times$ [illegible] Q^*
- $T[\![S]\!](s,A)$: [illegible] S starting in st[illegible]
- One probab[illegible] sequence of [illegible]

Plotkin & Jones
Morgan & McIver
Mislove
...

The Correspondence

Let (H,O,S) be a general experiment

Let μ be a prior probability distribution on hypotheses

Theorem: The following sets of probabilities are equal:

- $\{ \mu \oplus w_E(obs,\text{-}) \ : E \in S(G_S) \}$
- $\{\lambda h.(\mu(h)p : p \in \mathsf{T}\ [\![S]\!]\ (\sigma_0[\mathrm{hyp} \mapsto h], \Sigma \times \Gamma \times \{obs\}))\}$

Proof is a not so straightforward induction on S

The Correspondence

Let (H,O,S) be a general experiment

Let μ be a prior probability [illegible] hypotheses

Theorem: The foll[illegible] equal:

- $\{ \mu \oplus w_E (obs,$ [illegible]
- $\{\lambda h.(\mu(h)p : p \in$ [illegible] $obs\}))\}$

Proof is a not so straightforward induction on S

Not so straightforward
because semantics $T[\![S]\!]$
is *not* compositional!

Conclusion

We have devised an unambiguous well-founded way of describing possibly complex experiments and reason about how much evidence outcomes carry

This is a first step towards a general theory of understanding how much evidence can be gleaned by execution of probabilistic programs

Applications:

- Anonymity
- Quantitative information flow