



University of Illinois Department of Computer Science

Bayesian Learning, Randomness and Logic

Marc Snir

Background

- 25 years old work, far from my current research
 - why present now?
- Because it was done when I was Eli's student
- Because it is about the foundations of epistemology
- Because few people are likely to read a 50 page logic paper
 - Haim Gaifman and Marc Snir "Probabilities over Rich Languages, Testing and Randomness", JSL 47(3) 1982
- Because I never presented this paper

Bayesian learning

- Prior probability $\Pr(s)$ reflect a priori belief
- Tests provides facts a_1, \dots, a_n
- Beliefs change to conditional probability

$$\Pr(s | a_1, \dots, a_n)$$

- Does it work (do we learn about the true state of the world)?
- How dependent it is of the a priori beliefs?

Formalism

- Use first order language L_0 for arithmetic
 - integers, arithmetic operators, logical operators, quantifiers
- L is L_0 augmented with *empirical* predicates (statements about the world)
- Simple case: one empirical predicate $P(.)$
 - world (model) $w = w_1, w_2, w_3, \dots$ is a sequence of coin tosses
 - $w \sim P(i)$ iff $w_i = 1$

Examples of statements

- *Elementary* statements: $P(17) \wedge \neg P(35)$

- (this is a Δ_0 (Σ_0, Π_0) statement)

- Limit theorem:

$$\forall n \exists m \forall k \quad k > m \rightarrow \left| \frac{1}{2} - \frac{1}{k} \sum_{i=1}^k [P(i)] \right| < \frac{1}{n}$$

- (this is a Π_3 statement)

- Statements are classified according to number of quantifier alternations. Further subclasses can be defined

Probabilities defined on formal languages

- Nonnegative real-valued function $\Pr(\cdot)$ that has the following properties
 - If $\sim \varphi \leftrightarrow \psi$ then $\Pr(\varphi) = \Pr(\psi)$
 - If $\sim \varphi$ then $\Pr(\varphi) = 1$
 - If $\sim \neg(\varphi \wedge \psi)$ then $\Pr(\varphi \vee \psi) = \Pr(\varphi) + \Pr(\psi)$
 - $\Pr(\exists i \varphi(i)) = \lim_{n \rightarrow \infty} \Pr(\varphi(1) \vee \dots \vee \varphi(n))$
- $\Pr(\cdot)$ is uniquely defined by the values it gets on elementary statements

Back to induction

$$s^{(w)} = s \quad \text{if } w \models s$$

$$s^{(w)} = \neg s \quad \text{if } w \models \neg s$$

Def: Induction “works” in world w if

$$\lim_{n \rightarrow \infty} \Pr(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = [w \sim s]$$

■ Induction may not always work: assume that

$w \models s$ but $\Pr(s) = 0$ Then

$$\Pr(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = 0$$

■ **absolute beliefs are not affected by evidence!**

■ Thus, induction fails on $\bigcup \{ \text{Mod}(s) : \Pr(s) = 0 \}$

This is a zero probability set

Induction theorem

- **Thm: Induction works with probability one**

$$\lim_{n \rightarrow \infty} \Pr(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = [w \sim s] \quad a.e.$$

- (proof uses Doob's martingale convergence theorem)
- Note that "a.e." is defined wrt prior $\Pr(\cdot)$
 - really shows internal consistency, rather than external validity

Extreme example

- Assume $\Pr(P(i)) = 1, i = 1, 2, \dots$
 - prior reflects absolute belief that all coin tosses will yield heads, i.e., 1's; all probability is concentrated on the unique sequence $(1, 1, 1, \dots)$

$$\Pr(s) = 1 \quad \text{if} \quad (1, 1, 1, \dots) \models s$$

$$\Pr(s) = 0 \quad \text{otherwise}$$

- Nothing is learned from experience!
 - induction works only in world $(1, 1, 1, \dots)$

Moderate example

$$\Pr_p (P(1)^{(w)} \wedge \dots \wedge \pm P(k)^{(w)}) = p^i (1-p)^{k-i} \quad \text{where}$$

$$i = \sum_{j=1}^k w_j \quad (\text{Bernoulli prior})$$

$$\Pr(\cdot) = \lambda \Pr_p + (1-\lambda) \Pr_q, \quad 0 < \lambda < 1$$

(combination of two Bernoulli priors)

$$s = \forall n \exists m \forall k \quad k > m \rightarrow \left| p - \frac{1}{k} \sum_{i=1}^k [P(i)] \right| < \frac{1}{n}$$

$$\Pr(s) = \lambda$$

$$w \sim s \rightarrow \lim_{n \rightarrow \infty} \Pr(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = 1$$

Can learn which is the right parameter p

Variants of induction theorem

- **Need not test in strict order**
 - assume some Gödel numbering of sentences (elementary sentences)
- *Testing procedure*: $t : \{0,1\}^* \rightarrow \mathbb{N}$
 - the following sequence of tests is performed

$$s_1 = t(\Lambda), s_2 = t(s_1^{(w)}), s_3 = t(s_1^{(w)}, s_2^{(w)}), \dots$$
- Def: A testing procedure is *complete* if

$$w_1 \neq w_2 \rightarrow \exists k s_k^{(w_1)} \neq s_k^{(w_2)}$$

Variants

- Thm: If t is complete, $s_1 = t(\Lambda)$, $s_k = t(s_1^{(w)}, \dots, s_{k-1}^{(w)})$ then $\lim_{n \rightarrow \infty} \Pr(s \mid s_1^{(w)}, \dots, s_n^{(w)}) = [s^{(w)}] \quad a.e.$
- It is sufficient for t to be a.e. complete
- Only care about distinctions that are relevant to the truth of s
- Def: t is complete wrt s if

$$w_1 \sim s \wedge w_2 \sim \neg s \rightarrow \exists k \ s_k^{(w_1)} \neq s_k^{(w_2)}$$
- Thm: If t is (a.e.) complete wrt s then

$$\lim_{n \rightarrow \infty} \Pr(s \mid s_1^{(w)}, \dots, s_n^{(w)}) = [s^{(w)}] \quad a.e.$$

Randomness

- Definitions provided by Martin-Löf, Schnorr and others for random binary sequences can be expressed in the following form
 - A family Φ of statements (statistical tests) so that

$$\Pr_{1/2}(\varphi) = 0, \varphi \in \Phi$$
- Def: w is *random* (w.r.t. Φ) if $\forall \varphi \in \Phi \quad w \sim \neg \varphi$
- Different families of statements define different notions of randomness
- Can generalize to other priors

Randomness

- (absolute) randomness: Φ is set of all statements
- Def: w is *random* wrt $\text{Pr}(\cdot)$ if $\text{Pr}(s) = 1 \Rightarrow w \models s$
for any statement
- Def: w is $\sum_n (\Pi_n, \Delta_n)$ *random* wrt $\text{Pr}(\cdot)$ if
 $\text{Pr}(s) = 1 \Rightarrow w \models s$ for any statement s in $\sum_n (\Pi_n, \Delta_n)$

Definable probabilities

- Definable integer function $f : \mathbb{N} \rightarrow \mathbb{N}$ expressed by statement s in L_0

$$f(i) = j \leftrightarrow \sim s(i, j)$$

- Definable rational valued function $f : \mathbb{N} \rightarrow \mathbb{Q}$

$$f(i) = j/k \leftrightarrow \sim s(i, j, k)$$

- Definable real valued function $f : \mathbb{N} \rightarrow \mathbb{R}$: function has a definable approximation

$$g : \mathbb{N} \times \mathbb{N} \rightarrow \mathbb{Q} \text{ such that } |f(i) - g(i, j)| < 1/j$$

- Definable probability $\text{Pr}(\cdot)$: definable as function of index of elementary statement

Induction theorem redux

- Thm: If $\Pr(\cdot)$ is definable then

$$\lim_{n \rightarrow \infty} \Pr(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = [s^{(w)}]$$

for any random w

- Proof outline: can define statement s s.t.

$$\text{Mod}(s) = \{w : \lim_{n \rightarrow \infty} \Pr(s \mid P(0)^{(w)}, \dots, P(n)^{(w)}) = 1\}$$

$$\Pr(s) = 1 \quad \text{hence} \quad w \sim s$$

Induction works in any world that was not ruled out a priori

- strongest possible result!

Relative version of previous

- If $\Pr(\cdot)$ and t are Δ_m definable and t is complete for statements in Σ_m then

$$\lim_{n \rightarrow \infty} \Pr(s \mid s_1^{(w)}, \dots, s_n^{(w)}) = [s^{(w)}]$$

for any Σ_{2m+1} random w

level of randomness required is relative to strength of "tool" used for induction

Can we avoid absolute beliefs?

- Can we assign nonzero probability to every consistent statement?

Given $\Pr(\cdot)$ define w as follows

$$w_1 = 1 \text{ if } \Pr(P(1)) \leq \frac{1}{2}, \\ = 0 \text{ otherwise}$$

$$w_k = 1 \text{ if } \Pr(P(k) \mid P(1)^{(w)}, \dots, P(k-1)^{(w)}) \leq \frac{1}{2}, \\ = 0 \text{ otherwise}$$

If $\Pr(\cdot)$ is definable then w is definable by a statement S so that $\{w\} = \text{Mod}(S)$

Dogmatism

- s asserts "next coin toss is never the outcome predicted to be most likely to occur".
- s is consistent, but $\Pr(s) = 0$
- Any definable prior entails an absolute belief in its own efficiency!
- Thm: if $\Pr(\cdot)$ is approximated by definable f then there exist in L_0 predicate $\lambda(\cdot)$ recursive in f so that

$$\Pr(\forall n \lambda(n) \leftrightarrow P(n)) = 0$$

Can we find, for each consistent statement, a prior that does not rule it out?

Dogmatism

- Thm: There exists a consistent Π_2 statement s so that $\Pr(s) = 0$ for every definable $\Pr()$
- Proof outline: can build statement s that asserts that $P()$ is a truth predicate for L_0
 - s asserts $P(i) \leftrightarrow \text{True}(\varphi_i)$ for a Gödel enumeration of L_0 formulas $\varphi_1, \varphi_2, \dots$
- Statement built using recursive definition of the truth predicate

Construction

s is a conjunction of statements of the form

$$\forall m, n \quad " \varphi_m = \neg \varphi_n " \rightarrow P(m) \leftrightarrow \neg P(n)$$

$$\forall m, n \quad " \varphi_m = \forall(k) \varphi_n(k) " \rightarrow$$

$$P(m) \leftrightarrow \forall k P(\text{Sub}(n, k))$$

where $\text{Sub}(n, k)$ is the number of the formula obtained by substituting the name of the number k for the free variable of φ_n

And so on...

$w \in \text{Mod}(s)$ iff $P(\cdot)$ is the truth predicate for L_0 in w

Let $\{w_t\} = \text{Mod}(s)$

Contradiction

- Assume that $\Pr(s) > 0$

$\{w_t\} = \text{Mod}(s) \cap \text{Mod}(P(i)) = \text{Mod}(s)$ if φ_i is true

$\text{Mod}(s) \cap \text{Mod}(P(i)) = \emptyset$ if φ_i is false

$\Pr(s \wedge P(i)) = \Pr(s) > 0$ if φ_i is true

$\Pr(s \wedge P(i)) = 0$ if φ_i is false

 $\Pr(s \wedge P(i)) > 0$ iff φ_i is true

Have defined in L_0 truth function for L_0 . Contradiction!

- Any definable prior entails an absolute belief that nature does not encode the truth of arithmetic!

Consistent priors

- When do people with different initial priors converge to agreement?
- Def: $\Pr_1(\cdot)$ is *consistent* with $\Pr_2(\cdot)$ if $\Pr_1(s) = 0 \leftrightarrow \Pr_2(s) = 0$ (assume both are definable)
- Consistent priors define same notion of randomness
- Consistency is necessary and sufficient to agreement

Consistent priors

- If $\Pr_1(\cdot)$ is consistent with $\Pr_2(\cdot)$ then for all s and all random w

$$\lim_{n \rightarrow \infty} \Pr_1(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = \lim_{n \rightarrow \infty} \Pr_2(s \mid P(1)^{(w)}, \dots, P(n)^{(w)})$$

- If $\Pr_1(\cdot)$ is not consistent with $\Pr_2(\cdot)$ then for some s and all $\Pr_2(\cdot)$ random w

$$\lim_{n \rightarrow \infty} \Pr_1(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = 0, \lim_{n \rightarrow \infty} \Pr_2(s \mid P(1)^{(w)}, \dots, P(n)^{(w)}) = 1$$

- people with distinct priors agree only if they hold the same absolute beliefs

Strong agreement

- Consistency is sufficient for *uniform agreement*
- Let $\Pr_1(\cdot)$ and $\Pr_2(\cdot)$ be definable, consistent probabilities. Then

$$\lim_{n \rightarrow \infty} \text{Sup}_s (\Pr_1(s | P(1)^{(w)}, \dots, P(n)^{(w)}) - \Pr_2(s | P(1)^{(w)}, \dots, P(n)^{(w)})) = 0$$

- Example:

$$\lim_{n \rightarrow \infty} (\Pr_1(P(n+1) | P(1)^{(w)}, \dots, P(n)^{(w)}) - \Pr_2(P(n+1) | P(1)^{(w)}, \dots, P(n)^{(w)})) = 0$$

- Two people with consistent priors will converge to the same estimate for the probability that next coin toss is head

Future (?) work

- Develop similar framework for lower levels of the hierarchy (replace definable by computable)
- Express modern learning theory in Bayesian framework

Thanks

- To Haim Gaifman
- And to Eli Shamir, of course