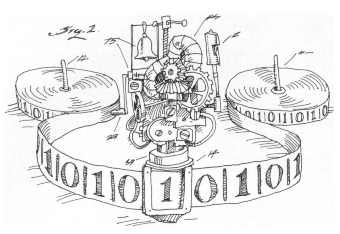
Logic and Probability
Probabilistic Grammars and Programs

Thomas Icard & Krzysztof Mierzewski

August 11, 2022

Minds as (Probabilistic) Machines

Minds as (Probabilistic) Machines



$\lambda x.x(y)$ $\quad$ $acd \rightarrow abbd$

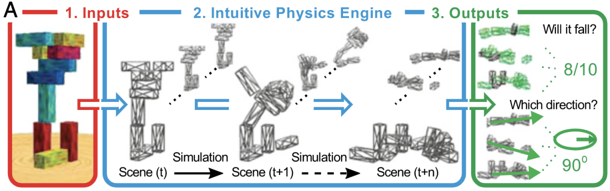
$\mathsf{S} \rightarrow \mathsf{NP\ VP}$

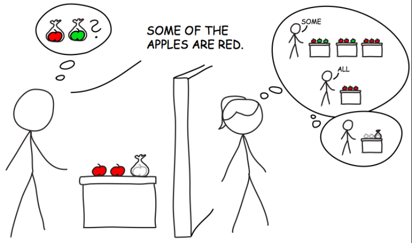
Why Probabilistic?

- Many processes are (well-modeled as) random.
- Randomized procedures can be more efficient.
- Probabilistic generative processes could play the functional role of 'subjective probabilities'.

⋮

Motivation


A
1. Inputs
2. Intuitive Physics Engine
3. Outputs
Will it fall?
8/10
Which direction?
90°
Simulation
Simulation
Scene (t)
Scene (t+1)
Scene (t+n)


SOME OF THE APPLES ARE RED.
SOME
ALL

Probabilistic generative models

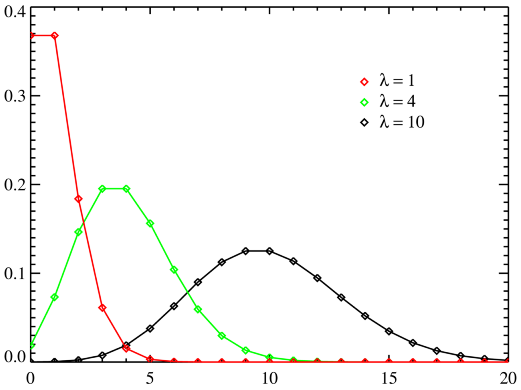
- Hidden Markov models
- Boltzmann machines
- Bayesian networks
- Probabilistic context-free grammars
- Probabilistic programs

  ⋮

These modeling tools typically define distributions on behaviors (or outputs) only implicitly.

Poisson distribution

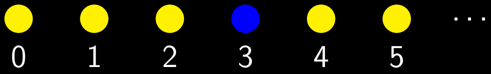
$$\mu(k) = e^{-\lambda}\frac{\lambda^k}{k!}$$



Random Walk Hitting Time

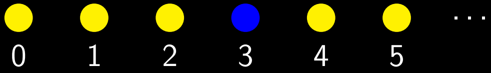
Random Walk Hitting Time

Random Walk Hitting Time



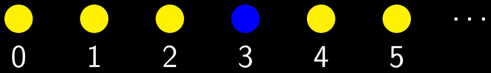
Random Walk Hitting Time

Random Walk Hitting Time



Random Walk Hitting Time

Random Walk Hitting Time



Random Walk Hitting Time

Random Walk Hitting Time

Random Walk Hitting Time

Random Walk Hitting Time

Random Walk Hitting Time

Random Walk Hitting Time

$$\mu(2k+1) = c_k 2^{-2k+1}$$

$$c_k = \binom{2k}{k}\frac{1}{k+1}$$

Beta-Binomial (or Dirichlet-Multinomial)

$$\begin{aligned} p &\sim \text{Beta}(\alpha, \beta) \\ k &\sim \text{Binomial}(n, p) \end{aligned}$$

$$\mu(k) \quad = \quad \binom{n}{k} \frac{B(\alpha + k, \beta + n - k)}{B(\alpha, \beta)}$$

Which generative models are capable of encoding distributions like these?

Given Σ ("terminal symbols") and $\mathcal{N}$ ("nonterminal symbols"), we consider **productions** of the form:

$$(\alpha \rightarrow \beta)$$

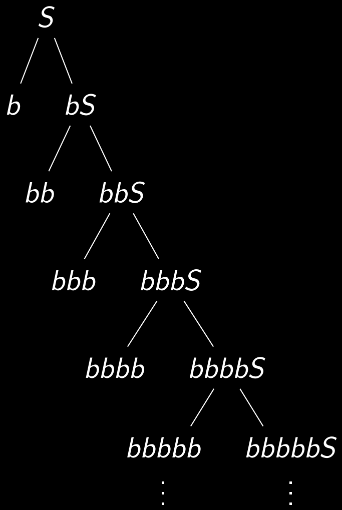
with $\alpha, \beta \in (\Sigma \cup \mathcal{N})^*$ strings over Σ and $\mathcal{N}$.

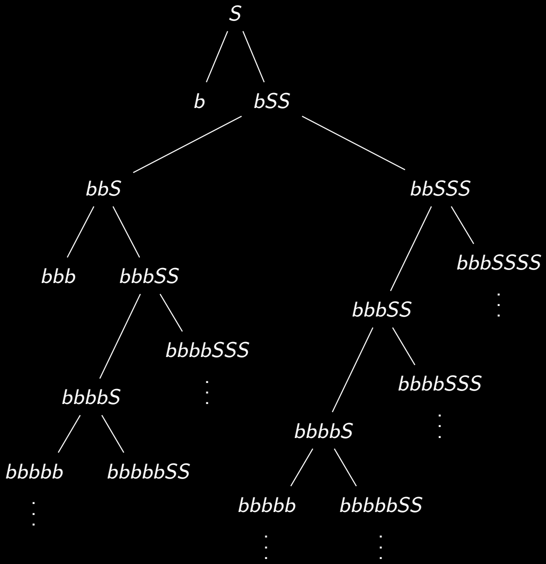
A grammar is a quadruple $(\Sigma, \mathcal{N}, \Pi, S)$.

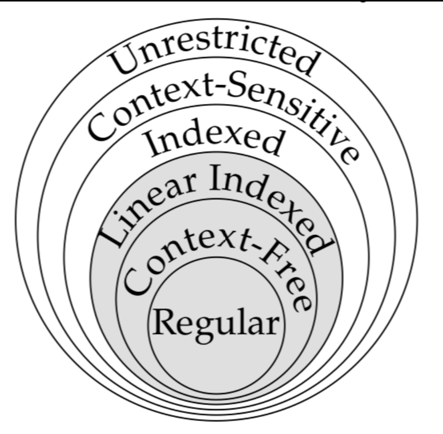
- **Regular (Type 3) Grammars**:
  - $(X \to \sigma Y)$
  - $(X \to \sigma)$
- **Context-Free (Type 2) Grammars**:
  - $(X \to \alpha)$
- **Context-Sensitive (Type 1) Grammars**:
  - $(\alpha X \beta \to \alpha \gamma \beta)$
  - $(S \to \epsilon)$
- **Unrestricted (Type 0) Grammars**:
  - $(\alpha \to \beta)$

Regular vs. Context-Free

$$
\begin{aligned}
S &\rightarrow bS \\
S &\rightarrow b
\end{aligned}
\qquad\qquad
\begin{aligned}
S &\rightarrow bSS \\
S &\rightarrow b
\end{aligned}
$$


S
b
bS
bb
bbS
bbb
bbbS
bbbb
bbbbS
bbbbb
bbbbbS


S
b
bSS
bbS
bbSSS
bbb
bbbSS
bbbSSSS
bbbSS
bbbbSSS
bbbbS
bbbbSSS
bbbbS
bbbbb
bbbbbSS
bbbbb
bbbbbSS


Unrestricted
Context-Sensitive
Indexed
Linear Indexed
Context-Free
Regular

Context-Free but not Regular

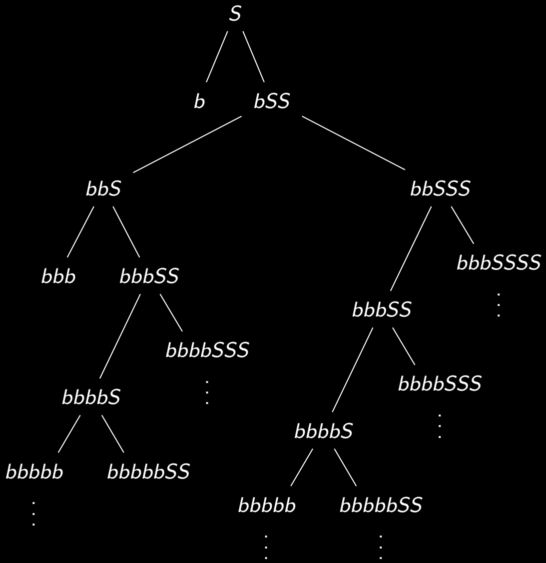
$$
\begin{aligned}
S &\rightarrow aSb \\
S &\rightarrow \epsilon
\end{aligned}
$$

Probabilistic Grammars

For each α we assume there are at most two β such that Π includes production

$$(\alpha \to \beta).$$

If Π includes $(\alpha \to \beta_1)$ and $(\alpha \to \beta_2)$, we think of the grammar as flipping a fair coin.


S
b
bSS
bbS
bbSSS
bbb
bbbSS
bbbSSSS
bbbSS
bbbbSSS
bbbbS
bbbbSSS
bbbbS
bbbbb
bbbbbSS
bbbbb
bbbbbSS

$$\begin{array}{rcl} S & \to & WYaZ \\ Ya & \to & aaY \\ YZ & \to & U \\ YZ & \to & VZ \\ aV & \to & Va \\ WV & \to & WY \\ WU & \to & \epsilon \\ aU & \to & Ua \end{array}$$

This defines $\mu(2^k) = 2^{-k}$.

Probabilistic Grammars and Machines

Theorem

Probabilistic grammars and probabilistic Turing machines define the same class of distributions.

Example (Flajolet et al. 2011)

```
x1, x2 := Geom(1/4)
t := x1 + x2
if flip(5/9) then t := t + 1
for j = 1, 2, 3
    draw 2t fair coin flips
    if #Heads ≠ #Tails then return 0
return 1
```

$\mu(1) = 1/\pi$.

$\mu : \mathbb{N} \to [0, 1]$ is a **semi-measure** if $\sum_k \mu(k) \leq 1$.

μ is **semi-computable** if for each k there is a computably enumerable sequence of rationals $q_1 \leq q_2 \leq q_3 \ldots$ with $\lim_{i \to \infty} q_i = \mu(k)$.

Theorem

Probabilistic grammars define exactly the semi-computable semi-measures.


```
t := 1; h := 0
while (h < t)
    t := t + 1
    if flip(1/4) then h := h + Unif(1,7)
return t-1
```


```
t := 1; h := 0
while (h < t)
    t := t + 1.0000000000001
    if flip(1/4) then h := h + Unif(1,7)
return t-1
```

The following are equally expressive:

- Probabilistic regular grammars
- Probabilistic finite-state automata
- Discrete hidden Markov models

Suppose we only had a q-biased coin. To reproduce

$$\begin{array}{rcl} X & \to & Y_1 \\ X & \to & Y_2 \end{array}$$

introduce nonterminal Z_1, Z_2 and write

$$\begin{array}{lll} X \stackrel{q}{\to} Z_1 & Z_1 \stackrel{q}{\to} X & Z_2 \stackrel{1-q}{\to} X \\ X \stackrel{1-q}{\to} Z_2 & Z_1 \stackrel{1-q}{\to} Y_1 & Z_2 \stackrel{q}{\to} Y_2. \end{array}$$

Nondyadic rationals

$$X \xrightarrow{1/3} Y_1$$
$$X \xrightarrow{1/3} Y_2$$
$$X \xrightarrow{1/3} Y_3$$

$$\begin{array}{lll} X \to Z_1 & Z_1 \to X & Z_2 \to Y_2 \\ X \to Z_2 & Z_1 \to Y_1 & Z_2 \to Y_3 \end{array}$$

Theorem

Probabilistic regular grammars can express every rational-valued distribution with finite support.

- Beta-Binomial (parameters in $\mathbb{N}$)
- Dirichlet-Multinomial
- Bayesian networks
- Arbitrarily good approximation to any Borel probability measure whatsoever!

Proposition

PRGs can *only* define rational-valued distributions.

Probability generating functions

Given μ we define the pgf $\mathfrak{G}_\mu(z)$ so that:

$$\mathfrak{G}_\mu(z) \;=\; \sum_{k=0}^{\infty} \mu(k) z^k$$

- **Rational** if $\mathfrak{G}_\mu(z) = \frac{Q_0(z)}{Q_1(z)}$
- **Algebraic** if $y = \mathfrak{G}_\mu(z)$ is a solution to a polynomial equation $0 = Q(y, z)$
- **Transcendental** otherwise

Example (Geometric Distribution)

The probability generating function for $\mu(k) = 2^{-k}$ is $\frac{1}{2-z}$.

$$\begin{aligned} S &\rightarrow bS \\ S &\rightarrow b \end{aligned}$$

Theorem (Schützenberger)

The probability generating function for any probabilistic regular grammar will be rational.

Example

The random walk hitting time distribution has pgf $(1 - \sqrt{1 - z^2})/z$, algebraic but not rational.

Example (Random Walk Hitting Time)

$$S \rightarrow bSS$$
$$S \rightarrow b$$

Example (Olmedo et al. 2016)

$$\begin{aligned} S &\rightarrow SSS \\ S &\rightarrow \epsilon \end{aligned}$$

The probability of returning ϵ is the solution to $x = \frac{1}{2}x^3 + \frac{1}{2}$, i.e., the reciprocal of the golden ratio!

Example (Etessami & Yannakakis 2009)

$$S \xrightarrow{1/6} SSSSS \qquad S \xrightarrow{1/2} b \qquad S \xrightarrow{1/3} \epsilon$$

To find probability of returning ϵ we need to solve $x = \frac{1}{6}x^5 + \frac{1}{3}$, which has no closed form.

Theorem

The pgf for a PCFG is always algebraic.

(Cf. Chomsky-Schützenberger Theorem)
(Cf. also Parikh's Theorem)

Proposition

For distributions with **finite support**, PCFGs define only the rational-valued ones.

Indexed Grammars

Add to $\mathcal{N}$ and Σ a finite set $\mathcal{I}$ of **indices**. Each non-terminal can carry a **stack** of indices.

- $X[l] \to \alpha[l]$
- $X[l] \to \alpha[kl]$
- $X[l] \to \alpha$

Theorem

Probabilistic indexed grammars can define distributions with transcendental pgfs.

Example

$$
\begin{aligned}
S[] &\rightarrow Y[I] \\
Y[I] &\rightarrow Y[II] \\
Y[I] &\rightarrow Z[I] \\
Z[I] &\rightarrow ZZ \\
Z[] &\rightarrow b
\end{aligned}
$$

This defines $\mu(2^k) = 2^{-k}$, with transcendental pgf.

Probabilistic Linear Indexed Grammars

- Allow only one non-terminal on the right.
- Equivalent to Tree-Adjoining Grammar, Combinatory Categorial Grammar, etc.
- Still algebraic, but can define finite-support, irrational-valued measures—thus surpassing expressive power of PCFGs.
- Equivalent to probabilistic pushdown automata.

Example ((Right-)Linear Indexed Grammar)

$$S[] \xrightarrow{1/2} b \qquad Y[I] \xrightarrow{1/4} Y[II] \qquad Y[I] \xrightarrow{1/2} Y$$
$$S[] \xrightarrow{1/2} bY[I] \qquad Y[I] \xrightarrow{1/4} b \qquad Y[] \xrightarrow{1} \epsilon$$

With 1/2 probability S rewrites to $bY[I]$, while $Y[I]$ in turn rewrites to ϵ with irrational probability $2 - \sqrt{2}$. Thus, $\mathbb{P}(b) = \frac{3-\sqrt{2}}{2}$, while $\mathbb{P}(bb) = \frac{\sqrt{2}-1}{2}$

Theorem

Probabilistic context-sensitive grammars define the same distributions as **non-erasing** Turing machines.

Corollary

PCSGs can define transcendental distributions that elude all the grammars considered up to this point.

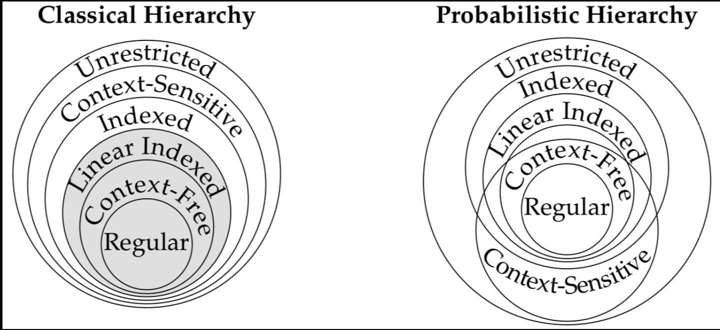
Proposition

Consider any semi-computable semi-measure $\mathbb{P} : \Sigma^* \to [0, 1]$. There is a PCSG on augmented vocabulary $\Sigma \cup \{\lhd\}$ defining a semi-measure $\tilde{\mathbb{P}}$ such that $\mathbb{P}(\sigma) = \sum_n \tilde{\mathbb{P}}(\sigma \lhd^n)$ for all $\sigma \in \Sigma^*$.

Proposition

PCSGs can only define rational probabilities!

Summary


Classical Hierarchy
Unrestricted
Context-Sensitive
Indexed
Linear Indexed
Context-Free
Regular
Probabilistic Hierarchy
Unrestricted
Indexed
Linear Indexed
Context-Free
Regular
Context-Sensitive

Some Open Questions

1. Exact characterization of the classes of distributions defined by PRGs or PCFGs?
2. Probabilistic (right-)linear grammars or PCSGs?
3. What kinds of generative models could naturally define Poisson distributions?
4. Efficient approximation at lower levels of distributions definable at higher levels?
5. Closure under probabilistic conditioning?

Summary of today

- Minds as (probabilistic) machines.
- Target: landscape of grammars and machines and what classes of distributions they can express.
- Many open questions and directions.

Tomorrow: computable measure theory + applications to Bayesian epistemology