Codifying Probabilities with The Last Full Measure

Albert Lai

November 1, 2023

Introduction

Motivating Example

Laplace's Rule of Succession

Let $n, k \in \mathbb{N}$, $0 \leq k \leq n$, be given/fixed.

1. pick r from $\mathrm{Uniform}(0, 1)$
2. pick c from $\mathrm{Binomial}(n, r)$
3. on condition / restrict to $c = k$

Expected value of $r = (k + 1)/(n + 2)$

Motivating Example

Laplace's Rule of Succession

Let $n, k \in \mathbb{N}$, $0 \leq k \leq n$, be given/fixed.

1. pick r from $\mathrm{Uniform}(0, 1)$
2. pick c from $\mathrm{Binomial}(n, r)$
3. on condition / restrict to $c = k$

Expected value of $r = (k + 1)/(n + 2)$

How do I work that out? Technical challenges:

- Hybrid continuous-discrete.
- Randomly choose a binomial distribution. 😱

Motivating Example

Laplace's Rule of Succession

Let $n, k \in \mathbb{N}$, $0 \leq k \leq n$, be given/fixed.

1. pick r from $\mathrm{Uniform}(0, 1)$
2. pick c from $\mathrm{Binomial}(n, r)$
3. on condition / restrict to $c = k$

Expected value of $r = (k + 1)/(n + 2)$

How do I work that out? Technical challenges:

- Hybrid continuous-discrete.
- Randomly choose a binomial distribution. 😱

(Finishing what 3Blue1Brown started: probabilities of probabilities part 1, part 2.)

Programming Probabilities

Random Number Generators

$\mathrm{RNG}\,X$ is set/type of programs that output a random X element (probability depends on program)

$\mathrm{Uniform} : \mathbb{R} \times \mathbb{R} \to \mathrm{RNG}\,\mathbb{R}$
$\mathrm{Uniform}(a, b)$ picks real number from $[a, b]$ evenly.

$\mathrm{Binomial} : \mathbb{N} \times [0, 1] \to \mathrm{RNG}\,\mathbb{N}$
$\mathrm{Binomial}(n, r)$ tosses coin n times, head probability r, count heads.

Random Number Generators

$\mathrm{RNG}\, X$ is set/type of programs that output a random X element (probability depends on program)

$\mathrm{Uniform} : \mathbb{R} \times \mathbb{R} \to \mathrm{RNG}\, \mathbb{R}$
$\mathrm{Uniform}(a, b)$ picks real number from $[a, b]$ evenly.

$\mathrm{Binomial} : \mathbb{N} \times [0, 1] \to \mathrm{RNG}\, \mathbb{N}$
$\mathrm{Binomial}(n, r)$ tosses coin n times, head probability r, count heads.

$\mathrm{Unit}_X : X \to \mathrm{RNG}\, X$
$\mathrm{Unit}_X(x)$ Always outputs x. "Rare desert of determinism in vast oasis of randomization." I omit subscript X if inferrable.

Chaining

Chaining operator “bind”, “then”, “flat-map”
$\gg= \,:\, \mathrm{RNG}\,X \times (X \to \mathrm{RNG}\,Y) \to \mathrm{RNG}\,Y$

$g \gg= k$ passes output of g to parametrized RNG k

Chaining

Chaining operator “bind”, “then”, “flat-map”
$\gg= : \mathrm{RNG}\,X \times (X \to \mathrm{RNG}\,Y) \to \mathrm{RNG}\,Y$

$g \gg= k$ passes output of g to parametrized RNG k

Example:

1. pick r from $\mathrm{Uniform}(0,1)$
2. pick c from $\mathrm{Binomial}(n,r)$
3. output (r,c)

$\mathrm{Uniform}(0,1) \gg= (\lambda r \cdot \mathrm{Binomial}(n,r) \gg= (\lambda c \cdot \mathrm{Unit}(r,c)))$
or
$\mathrm{Uniform}(0,1) \gg= \lambda r \cdot \mathrm{Binomial}(n,r) \gg= \lambda c \cdot \mathrm{Unit}(r,c)$

Restriction/Conditional

Restriction operator
$| : \mathrm{RNG}\, X \times (X \to \mathbb{B}) \to \mathrm{RNG}\, X$
($\mathbb{B}$ = {false, true})

$g \mid pred$ = restrict g to when $pred$ is true.

(Rejection sampling: Keep retrying g until output satisfies $pred$.)

Restriction/Conditional

Restriction operator
$| : \mathrm{RNG}\,X \times (X \to \mathbb{B}) \to \mathrm{RNG}\,X$
($\mathbb{B}$ = {false, true})

$g \mid pred$ = restrict g to when $pred$ is true.

(Rejection sampling: Keep retrying g until output satisfies $pred$.)

Example:

1. pick x from $\mathrm{Uniform}(0, 1)$
2. on condition $x > 0.3$
3. (output x)

$\mathrm{Uniform}(0, 1) \mid (\lambda x \cdot x > 0.3)$

Express as Program

Define

$$
\begin{aligned}
g &= \text{Uniform}(0,1) \gg= \lambda r \cdot \text{Binomial}(n,r) \gg= \lambda c \cdot \text{Unit}(r,c) \\
g' &= g \mid (\lambda(r,c) \cdot c = k) \\
\mu &= g' \gg= \lambda(r,c) \cdot \text{Unit}(r)
\end{aligned}
$$

Then μ generates an r as prescribed.

Re-run many times to approximate distribution and expected value.

(Actual Haskell code modulo syntax.)

Re-read Program as Probability Measure

There is also a measure-theory reading of

$$\begin{aligned} g &= \mathrm{Uniform}(0,1) \gg\!= \lambda r \cdot \mathrm{Binomial}(n,r) \gg\!= \lambda c \cdot \mathrm{Unit}(r,c) \\ g' &= g \,|\, (\lambda(r,c) \cdot c = k) \\ \mu &= g' \gg\!= \lambda(r,c) \cdot \mathrm{Unit}(r) \end{aligned}$$

Then μ is the probability distribution/measure of the r in question.

Can find expected value and pdf.

Re-read Program as Probability Measure

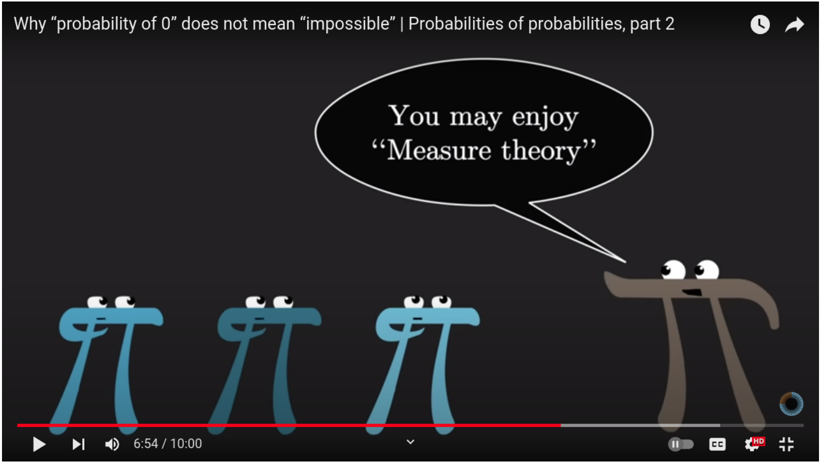
There is also a measure-theory reading of

$$\begin{aligned} g &= \text{Uniform}(0,1) \gg= \lambda r \cdot \text{Binomial}(n,r) \gg= \lambda c \cdot \text{Unit}(r,c) \\ g' &= g \mid (\lambda(r,c) \cdot c = k) \\ \mu &= g' \gg= \lambda(r,c) \cdot \text{Unit}(r) \end{aligned}$$

Then μ is the probability distribution/measure of the r in question.

Can find expected value and pdf.

"**Codifying probability** with the last full **measure**."


Why “probability of 0” does not mean “impossible” | Probabilities of probabilities, part 2
You may enjoy
“Measure theory”
6:54 / 10:00

Probability Measures

Measure Theory: Motivation 1/2

Integration by Lateral Thinking. Literally.

Riemann integral: Pixelate the x-axis.

$$\sum_i f(x_i) \times \mathrm{length}[x_i, x_i + \epsilon)$$

Measure Theory: Motivation 1/2

Integration by Lateral Thinking. Literally.

Riemann integral: Pixelate the x-axis.

$$\sum_i f(x_i) \times \text{length}[x_i, x_i + \epsilon)$$

Lebesgue integral: Pixelate the y-axis.

$$\sum_i y_i \times \text{length}\left(f^{-1}[y_i, y_i + \epsilon)\right)$$

Measure Theory: Motivation 1/2

Integration by Lateral Thinking. Literally.

Riemann integral: Pixelate the x-axis.

$$\sum_i f(x_i) \times \text{length}[x_i, x_i + \epsilon)$$

Lebesgue integral: Pixelate the y-axis.

$$\sum_i y_i \times \text{length}\left(f^{-1}[y_i, y_i + \epsilon)\right)$$

Need “length” for fairly general subsets. Require:

$$\text{length}(\bigcup_i A_i) = \sum_i \text{length}(A_i) \quad \text{(countable disjoint union)}$$

Turns out problematic for all subsets. Settle for large enough family closed under complement, countable union.

Measure Theory: Motivation 2/2

Probability Theory

Sample space Ω.

$\Pr$: subsets of $\Omega \to [0, 1]$

Require:

$$\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i) \quad \text{(countable disjoint union)}$$

Hmm déjà vu...

Measure Theory: Motivation 2/2

Probability Theory

Sample space Ω.

$\Pr$: subsets of $\Omega \to [0,1]$

Require:

$$\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i) \quad \text{(countable disjoint union)}$$

Hmm déjà vu…

Moreover, expected value of $f : \Omega \to \mathbb{R}$ = use $\Pr$ for "length" in Lebesgue integration!

$$\sum_i y_i \times \Pr\left(f^{-1}[y_i, y_i + \epsilon)\right)$$

Measure Theory: Motivation 2/2

Probability Theory

Sample space Ω.

$\Pr$: subsets of $\Omega \to [0, 1]$

Require:

$$\Pr(\bigcup_i A_i) = \sum_i \Pr(A_i) \quad \text{(countable disjoint union)}$$

Hmm déjà vu…

Moreover, expected value of $f : \Omega \to \mathbb{R}$ = use $\Pr$ for "length" in Lebesgue integration!

$$\sum_i y_i \times \Pr\left(f^{-1}[y_i, y_i + \epsilon)\right)$$

Riemann may not work: Ω may not even have "intervals".

Measurable Space; σ-Algebra

Data of a measurable space:

- Set of points X.
- σ-algebra $\mathfrak{M}$ or $\mathfrak{M}_X$, $\subseteq \wp(X)$, closed under:
  - owning X, owning $\emptyset$
  - complement, countable union
  - (corollary: also subtraction, countable intersection)

  Members “measurable subsets”.

Measurable Space; σ-Algebra

Data of a measurable space:

- Set of points X.
- σ-algebra $\mathfrak{M}$ or $\mathfrak{M}_X$, $\subseteq \wp(X)$, closed under:
  - owning X, owning $\emptyset$
  - complement, countable union
  - (corollary: also subtraction, countable intersection)

  Members "measurable subsets".

Examples:

- X countable set, $\mathfrak{M} = \wp(X)$

Measurable Space; σ-Algebra

Data of a measurable space:

- Set of points X.
- σ-algebra $\mathfrak{M}$ or $\mathfrak{M}_X$, $\subseteq \wp(X)$, closed under:
  - owning X, owning $\emptyset$
  - complement, countable union
  - (corollary: also subtraction, countable intersection)

  Members “measurable subsets”.

Examples:

- X countable set, $\mathfrak{M} = \wp(X)$
- $X = \mathbb{R}$, $\mathfrak{M}$ Borel algebra: smallest σ-algebra owning open sets.

Measurable Space; σ-Algebra

Data of a measurable space:

- Set of points X.
- σ-algebra $\mathfrak{M}$ or $\mathfrak{M}_X$, $\subseteq \wp(X)$, closed under:
  - owning X, owning $\emptyset$
  - complement, countable union
  - (corollary: also subtraction, countable intersection)

  Members "measurable subsets".

Examples:

- X countable set, $\mathfrak{M} = \wp(X)$
- $X = \mathbb{R}$, $\mathfrak{M}$ Borel algebra: smallest σ-algebra owning open sets.
- $X = \mathbb{R}$, $\mathfrak{M}$ Lebesgue-measurable subsets (appendix). Larger than Borel algebra. Default.

Measures; Probability Measures

[Positive] Measure μ over measurable space $(X, \mathfrak{M})$:

- $\mu : \mathfrak{M} \to [0, \infty]$
- $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$ for countable disjoint union
  "countably additive"
- (corollary: $\mu(\emptyset) = 0$)
- (corollary: if $A \subseteq B$ then $\mu(A) \leq \mu(B)$)

Measures; Probability Measures

[Positive] Measure μ over measurable space $(X, \mathfrak{M})$:

- $\mu : \mathfrak{M} \to [0, \infty]$
- $\mu(\bigcup_i A_i) = \sum_i \mu(A_i)$ for countable disjoint union
  "countably additive"
- (corollary: $\mu(\emptyset) = 0$)
- (corollary: if $A \subseteq B$ then $\mu(A) \leq \mu(B)$)

Probability measure a.k.a. distribution: Furthermore:

- $\mu(X) = 1$
- (corollary: $\mu : \mathfrak{M} \to [0, 1]$)

Define ΠX = set of probability measures over $(X, \mathfrak{M}_X)$.

Discrete Measures

Counting measure: X countable set, $\mathfrak{M} = \wp(X)$
$\# : \wp(X) \to \mathbb{N} \cup \{\infty\} \subseteq [0, \infty]$

$$\#(A) = \begin{cases} |A| & A \text{ finite} \\ \infty & A \text{ infinite} \end{cases}$$

Discrete Measures

Counting measure: X countable set, $\mathfrak{M} = \wp(X)$
$\# : \wp(X) \to \mathbb{N} \cup \{\infty\} \subseteq [0, \infty]$

$$\#(A) = \begin{cases} |A| & A \text{ finite} \\ \infty & A \text{ infinite} \end{cases}$$

Binomial distribution: $X = \mathbb{N}$ (for simplicity), $\mathfrak{M} = \wp(X)$
$\mathrm{Binomial} : \mathbb{N} \times [0, 1] \to \Pi\, \mathbb{N}$

$$\mathrm{Binomial}(n, r)(A) = \sum_{\substack{i \in A \\ 0 \leq i \leq n}} \binom{n}{i} r^i (1 - r)^{n-i}$$

Continuous Measures

Lebesgue measure: $X = \mathbb{R}$, $\mathfrak{M}$ Lebegue-measurable subsets.
$m : \mathfrak{M} \to [0, \infty]$
$m([a, b]) = b - a$, same for $(a, b]$ etc. Full defn in appendix.
Default.

Continuous Measures

Lebesgue measure: $X = \mathbb{R}$, $\mathfrak{M}$ Lebegue-measurable subsets.
$m : \mathfrak{M} \to [0, \infty]$
$m([a, b]) = b - a$, same for $(a, b]$ etc. Full defn in appendix.
Default.

Uniform distribution: Lebesgue measure with rescaling:
$\mathrm{Uniform} : \mathbb{R} \times \mathbb{R} \to \Pi\, \mathbb{R}$

$$\mathrm{Uniform}(a, b)(A) = \frac{1}{b - a} m(A \cap [a, b])$$

Continuous Measures

Lebesgue measure: $X = \mathbb{R}$, $\mathfrak{M}$ Lebegue-measurable subsets.
$m : \mathfrak{M} \to [0, \infty]$
$m([a, b]) = b - a$, same for $(a, b]$ etc. Full defn in appendix.
Default.

Uniform distribution: Lebesgue measure with rescaling:
$\text{Uniform} : \mathbb{R} \times \mathbb{R} \to \Pi\, \mathbb{R}$

$$\text{Uniform}(a, b)(A) = \frac{1}{b - a} m(A \cap [a, b])$$

Other continuous distributions mentioned later after integration.

The Unit Measure

$\mathrm{Unit}_X : X \to \Pi X$

$$\mathrm{Unit}_X(x)(A) = \chi_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}$$

(χ_A characteristic function of set A)

Deterministic corner case.

X can be discrete or continuous or any measurable space.

The Unit Measure

$\mathrm{Unit}_X : X \to \Pi X$

$$\mathrm{Unit}_X(x)(A) = \chi_A(x) = \begin{cases} 0 & x \notin A \\ 1 & x \in A \end{cases}$$

(χ_A characteristic function of set A)

Deterministic corner case.

X can be discrete or continuous or any measurable space.

In the continuous case, no probability density function.

Measurable Functions

We will only integrate functions that satisfy:

$f : X \to [-\infty, \infty]$ measurable function iff any of:

- for all open B, $f^{-1}(B) \in \mathfrak{M}_X$
- for all $y \in \mathbb{R}$, $\{x \mid f(x) > y\} \in \mathfrak{M}_X$
- or $\geq$, or $<$, or $\leq$

Motivation: Need $\mu\left(f^{-1}[y, y+\epsilon)\right)$, makes sense for $\mathfrak{M}_X$ only.

Measurable Functions

We will only integrate functions that satisfy:

$f : X \to [-\infty, \infty]$ measurable function iff any of:

- for all open B, $f^{-1}(B) \in \mathfrak{M}_X$
- for all $y \in \mathbb{R}$, $\{x \mid f(x) > y\} \in \mathfrak{M}_X$
- or $\geq$, or $<$, or $\leq$

Motivation: Need $\mu\left(f^{-1}[y, y + \epsilon)\right)$, makes sense for $\mathfrak{M}_X$ only.

Easy: $f : \mathbb{N} \to [-\infty, \infty]$ is measurable using $\mathfrak{M} = \wp(\mathbb{N})$.

Theorem: Piecewise continuous $f : \mathbb{R} \to \mathbb{R}$ is measurable using Lebesgue-measurable subsets.

Lebesgue Integration: Idea

Notation: $\int f\, d\mu$ integrates f over all of X using measure μ.

Assume $f \geq 0$ for now. Approximate from below by simple functions (finite range).

Example: $X = [-4, 4]$, $f(x) = x^2$, an approximation is

$$s(x) = 4 \cdot \chi_{(-3,-2]\cup[2,3)}(x) + 9 \cdot \chi_{[-4,-3]\cup[3,4]}(x)$$

"Clearly"

$$\int s\, d\mu = 4 \cdot \mu((-3, -2] \cup [2, 3)) + 9 \cdot \mu([-4, -3] \cup [3, 4])$$

Take supremum over all approximations.

For general f, split into positive and negative parts, both treatable as above.

Lebesgue Integration

Simple functions (finite range):

$$\int \sum_{i=1}^{n} a_i \, \chi_{E_i} \, d\mu = \sum_{i=1}^{n} a_i \, \mu(E_i) \quad (a_i \in \mathbb{R}, \, E_i \in \mathfrak{M})$$

Extend to non-negative $f : X \to [0, \infty]$

$$\int f \, d\mu = \sup\{\int s \, d\mu \mid s \text{ simple}, 0 \leq s \leq f\}$$

Extend to full range $f : X \to [-\infty, \infty]$

$$\int f \, d\mu = \left(\int \max(0, f) \, d\mu\right) - \left(\int -\min(f, 0) \, d\mu\right)$$

assuming not $\infty - \infty$.

Lebesgue Integration

Notation: $\int_A f\, d\mu$ is over arbitrary $A \in \mathfrak{M}$, instead of all X.

Two equivalent treatments:

- Revise the definitions, change $\mu(E_i)$ to $\mu(E_i \cap A)$.
- Just use $\int f \chi_A\, d\mu$

Discrete Integration

X countable set, $\mathfrak{M} = \wp(X)$, counting measure:

$$\int f \, d\# = \sum_{i \in X} f(i)$$

(assume $f \geq 0$ or else absolute convergence or other conditions)

Discrete Integration

X countable set, $\mathfrak{M} = \wp(X)$, counting measure:

$$\int f \, d\# = \sum_{i \in X} f(i)$$

(assume $f \geq 0$ or else absolute convergence or other conditions)

Binomial expressible as integral $d\#$:

$$\mathrm{Binomial}(n, r)(A) = \int_A \lambda i \cdot \binom{n}{i} r^i (1-r)^{n-i} \, d\#$$

Hence $\lambda i \cdot \binom{n}{i} r^i (1-r)^{n-i}$ is probability mass function.

$$\int f \, d(\mathrm{Binomial}(n, r)) = \int \lambda i \cdot f(i) \binom{n}{i} r^i (1-r)^{n-i} \, d\#$$

Continuous Integration (Sorry!)

$\int f\,dm$ default for $X = \mathbb{R}$ or subspace. Riemann toolbox reusable because:

Theorem: For Riemann-integrable $f : [a, b] \to \mathbb{R}$,

$$\int_{[a,b]} f\,dm = \text{Riemann} \int_a^b f(x)\,dx$$

Continuous Integration (Sorry!)

$\int f\,dm$ default for $X = \mathbb{R}$ or subspace. Riemann toolbox reusable because:

Theorem: For Riemann-integrable $f : [a, b] \to \mathbb{R}$,

$$\int_{[a,b]} f\,dm = \text{Riemann} \int_a^b f(x)\,dx$$

Uniform distribution: Reuse m with rescaling:

$$\int f\,d(\mathrm{Uniform}(a, b)) = \frac{1}{b - a} \int_{[a,b]} f\,dm$$

Continuous Integration (Sorry!)

$\int f\, dm$ default for $X = \mathbb{R}$ or subspace. Riemann toolbox reusable because:

Theorem: For Riemann-integrable $f : [a, b] \to \mathbb{R}$,

$$\int_{[a,b]} f\, dm = \text{Riemann} \int_a^b f(x)\, dx$$

Uniform distribution: Reuse m with rescaling:

$$\int f\, d(\text{Uniform}(a, b)) = \frac{1}{b-a} \int_{[a,b]} f\, dm$$

Most continuous distributions have probability density functions:

$$\mu(A) = \int_A \text{pdf}\, dm$$

$$\int f\, d\mu = \int f\text{pdf}\, dm$$

Unit Integration

Sifting property:

$$\int f \, d(\mathrm{Unit}(x)) = f(x)$$

Integral well defined (and beautiful), even in continuous case, even if no pdf. The power of measure theory!

Chaining Probability Measures

Chaining operator “bind”, “then”, “flat-map”
$\gg= \, : \Pi X \times (X \to \Pi Y) \to \Pi Y$

Intuition for $(\mu \gg= k)(B)$: For each $x \in X$, $k(x) \in \Pi Y$, $k(x)(B)$ is probability of B. Average over X according to μ. (Total probability).

Likewise for expected values.

$$(\mu \gg= k)(B) = \int \lambda x \cdot k(x)(B) \, d\mu$$

$$\int f \, d(\mu \gg= k) = \int \left(\lambda x \cdot \int f \, d(k(x)) \right) d\mu$$

(Require $k : X \to \Pi Y$ measurable function. Boils down to $(\lambda x \cdot k(x)(B)) : X \to [0, 1]$ measurable function for all $B \in \mathfrak{M}_Y$.)

Interlude

Happy Belated Halloween! 👻

$(\Pi, \text{Unit}, \gg=)$ is a programmer-friendly version of…

The Giry monad! Due to Michèle Giry. More in appendix.

Restricting Probability Measures

Restriction operator
$| : \Pi X \times (X \to \mathbb{B}) \to \Pi X$

Conditional probability.

$$(\mu \,|\, pred)(A) = \mu(A \cap B)/\mu(B)$$
$$\int f \, d(\mu \,|\, pred) = \frac{1}{\mu(B)} \int f \chi_B \; d\mu$$

where $B = \{x \mid pred(x)\}$, assuming $\in \mathfrak{M}$.

(One may say: $pred$ is a measurable predicate.)

Re-read Program as Probability Measure

Define

$$
\begin{aligned}
g &= \text{Uniform}(0,1) \gg\!= \lambda r\cdot \text{Binomial}(n,r) \gg\!= \lambda c\cdot \text{Unit}(r,c) \\
g' &= g \,|\, (\lambda(r,c)\cdot c = k) \\
\mu &= g' \gg\!= \lambda(r,c)\cdot \text{Unit}(r)
\end{aligned}
$$

Then μ is the probability measure of the r in question.

$$
\begin{aligned}
\text{expected} &= \int \lambda r\cdot r \, d\mu \\
&= (n+1) \int_0^1 r\binom{n}{k} r^k (1-r)^{n-k} \, dr \qquad \text{appendix} \\
&= \frac{k+1}{n+2} \\
\text{pdf}(r) &= (n+1)\binom{n}{k} r^k (1-r)^{n-k}
\end{aligned}
$$

Conclusion

Conclusion

Formal constructs for expressing probabilistic models.

Program interpretation gives executable samplers.

Meaure-theory interpretation gives resulting distributions.

Works the same way for discrete, continuous, hybrid, nested.

Bibliography

Bibliography

H. L. Royden. Real Analysis.

Walter Rudin. Real And Complex Analysis.

Philip Wadler. Monads for functional programming. In Marktoberdorf Summer School on Program Design Calculi, 1993. (copy)

Michèle Giry. A categorical approach to probability theory. In Categorical Aspects of Topology and Analysis, LNM 915, 1982.

Norman Ramsey and Avi Pfeffer. Stochastic lambda calculus and monads of probability distributions. In POPL '02, 2002. (copy)

E. C. R. Hehner. A probability perspective. In Formal Aspects of Computing 23(4), 2011. (copy, video)

Bart Jacobs. Structured Probabilistic Reasoning. WIP. (copy)

Appendix

Lebesgue Measure And Measurable Subsets

The following is well-defined for all $A \in \wp(\mathbb{R})$; denoted $m^*(A)$:

- let $\{I_n\}$ be sequence of disjoint open intervals, $\bigcup_n I_n \supseteq A$
- $\sum_n \mathrm{length}(I_n)$ uncontroversial
- take infinum over all possibilities

Only issue: Not countably additive for some subsets.

Carathéodory's solution: Restrict to

$$\mathfrak{M} = \{A \mid \forall B \cdot m^*(B) = m^*(B \cap A) + m^*(B - A)\}$$

Adopted for Lebesgue-measurable subsets.

Then define Lebesgue measure $m = m^*|_{\mathfrak{M}}$

The Giry Monad 1/2

Prologue: Generalize "measurable functions" from $X \to [-\infty, \infty]$ to $X \to Y$, with Y any measurable space:
$h : X \to Y$ measurable iff $\forall B \in \mathfrak{M}_Y \cdot h^{-1}(B) \in \mathfrak{M}_X$

Π as endofunctor (on category of measurable spaces):

Object map: $\Pi\, X$ = set of probability measures over $(X, \mathfrak{M}_X)$.
σ-algebra: Smallest s.t. for all $A \in \mathfrak{M}_X$, $(\lambda\mu \cdot \mu(A)) : \Pi\, X \to [0, 1]$ measurable function.

Morphism map: For measurable $h : X \to Y$, $\Pi\, h : \Pi\, X \to \Pi\, Y$

$$(\Pi\, h)(\mu)(B) = \mu(h^{-1}(B))$$
$$\int f\, d((\Pi\, h)(\mu)) = \int f \circ h\, d\mu$$

The Giry Monad 2/2

Π as monad:

$\mathrm{Unit}_X : X \to \Pi X$ as defined earlier.

Multiplication $\mathrm{Flat}_X : \Pi(\Pi X) \to \Pi X$

$$\mathrm{Flat}(\mu)(A) = \int \lambda \nu \cdot \nu(A)\, d\mu$$

$$\int f\, d(\mathrm{Flat}(\mu)) = \int \left(\lambda \nu \cdot \int f\, d\nu \right) d\mu$$

Flattens “probability of probabilities” to total probability.

Then define $\mu \gg= k = \mathrm{Flat}((\Pi\, k)(\mu))$.

Detailed Calculation 1/2

Let $B = \{(r, c) \mid c = k\} = [0, 1] \times \{k\}$.

$$\int f\,dg = \int \left(\lambda r \cdot \int \left(\lambda c \cdot \int f\,d(\mathrm{Unit}(r, c))\right) d(\mathrm{Binomial}(n, r))\right)$$
$$d(\mathrm{Uniform}(0, 1))$$
$$= \int_0^1 \sum_{c=0}^{n} f(r, c)\binom{n}{c} r^c (1 - r)^{n-c}\,dr$$
$$g(B) = \int \chi_B\,dg$$
$$= \int_0^1 \binom{n}{k} r^k (1 - r)^{n-k}\,dr$$
$$= \binom{n}{k}\frac{k!(n-k)!}{(n+1)!} \qquad \text{(Beta functions)}$$
$$= \frac{1}{n+1}$$

Detailed Calculation 2/2

$$
\begin{aligned}
\int \lambda r \cdot r \, d\mu &= \int \left(\lambda(r,c) \cdot \int \lambda r \cdot r \, d(\mathrm{Unit}(r)) \right) dg' \\
&= \int (\lambda(r,c) \cdot r) \; dg' \\
&= \frac{1}{g(B)} \int (\lambda(r,c) \cdot r \chi_B(r,c)) \; dg \\
&= \frac{1}{g(B)} \int_0^1 r \binom{n}{k} r^k (1-r)^{n-k} \, dr \\
&= (n+1) \binom{n}{k} \frac{(k+1)!(n-k)!}{(n+2)!} \qquad \text{(Beta functions)} \\
&= \frac{k+1}{n+2}
\end{aligned}
$$