@AndrzejWasowski@scholar.social
Andrzej Wąsowski

Advanced Programming

Probabilistic Programming in a Nuthshell

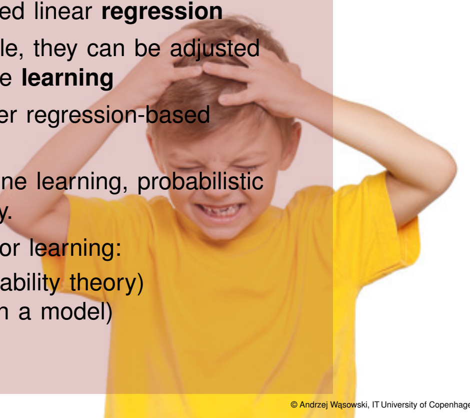
IT UNIVERSITY OF COPENHAGEN

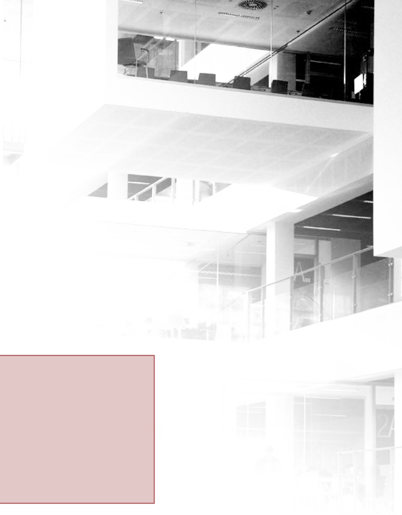
SOFTWARE QUALITY RESEARCH

Probabilistic Programming

What and Why.

- API/language to build a **probabilistic model** — so a probabilistic relation between some observed values and some conclusions/diagnosis
- You can use it to (among others): build **classifiers** (Y/N), **assess risks** (probability), predict values (expectation calculation), perform generalized linear **regression**
- Bayesian probabilistic models (most popular) are adaptable, they can be adjusted easily to new evidence showing up. This allows to capture **learning**
- So probabilistic programming is also an alternative to other regression-based machine learning methods.
- But unlike single-point-estimate models common in machine learning, probabilistic programming incorporates **uncertainty** in a principled way.
- Three nice things about using probabilistic programming for learning:
  - Based on sound and nice mathematical model (probability theory)
  - Gives a systematic way to capture what you know (in a model)
  - Allows learning the model parameters
  - Bonus: it is monadic and functional!





- Probability
- Conditional probability
- Bayes theorem

AGENDA

General definition of probability function

Definition (Dekking et al. p. 16)

A **probability function** p on a **finite sample space** S assigns to each event E in S **a number** $p(E)$ **in** $[0, 1]$ such that

i. $p(S) = 1$, and

ii. $P(E \cup F) = P(E) + P(F)$ if E and F are **disjoint**.

The number $P(E)$ is called the probability that E occurs.

General definition of probability function

Definition (Dekking et al. p. 16)

A **probability function** p on a **finite sample space** S assigns to each event E in S **a number** $p(E)$ **in** $[0, 1]$ such that

i. $p(S) = 1$, and

ii. $P(E \cup F) = P(E) + P(F)$ if E and F are **disjoint**.

The number $P(E)$ is called the probability that E occurs.

The additive property (ii) implies the following theorem.

Theorem

For a finite sample space S we have that

$$p(E) = \sum_{s \in E} p(\{s\})$$

Note: Rosen uses the shorthand notation $p(s) = p(\{s\})$ for $s \in S$.

Conditional probability

Definition (Rosen p. 442)

Let E and F be events with $p(F) > 0$. The conditional probability of E given F, denoted by $p(E|F)$, is defined as

$$p(E|F) = \frac{p(E \cap F)}{p(F)}$$

Example

What is the conditional probability of an odd number given that I rolled a prime number with a fair die?

Let $O = \{1, 3, 5\}$ and $P = \{2, 3, 5\}$. Since $O \cap P = \{3, 5\}$ we have

$$p(O|P) = \frac{2/6}{3/6} = \frac{2}{3}$$

Conditional probability

Example

What is the conditional probability that a family with two children has two boys, given they have at least one boy?

Conditional probability

Example

What is the conditional probability that a family with two children has two boys, given they have at least one boy?

Let the sample space be $S = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}, \mathrm{GG}\}$ and assume that each possible outcome is equally likely.

Exercise

What is the probability of having two boys?

Conditional probability

Example

What is the conditional probability that a family with two children has two boys, given they have at least one boy?

Let the sample space be $S = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}, \mathrm{GG}\}$ and assume that each possible outcome is equally likely.

Let E be the event that they have two boys, i.e. $E = \{\mathrm{BB}\}$.

Let F be the event that they have at least one boy, $F = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}\}$

Conditional probability

Example

What is the conditional probability that a family with two children has two boys, given they have at least one boy?

Let the sample space be $S = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}, \mathrm{GG}\}$ and assume that each possible outcome is equally likely.

Let E be the event that they have two boys, i.e. $E = \{\mathrm{BB}\}$.

Let F be the event that they have at least one boy, $F = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}\}$

Since the four possibilities are equally likely, we have that $p(E \cap F) = 1/4$ and $p(F) = 3/4$.

Conditional probability

Example

What is the conditional probability that a family with two children has two boys, given they have at least one boy?

Let the sample space be $S = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}, \mathrm{GG}\}$ and assume that each possible outcome is equally likely.

Let E be the event that they have two boys, i.e. $E = \{\mathrm{BB}\}$.

Let F be the event that they have at least one boy, $F = \{\mathrm{BB}, \mathrm{BG}, \mathrm{GB}\}$

Since the four possibilities are equally likely, we have that $p(E \cap F) = 1/4$ and $p(F) = 3/4$. Therefore we conclude that

$$p(E|F) = \frac{p(E \cap F)}{p(F)} = \frac{1/4}{3/4} = 1/3.$$

Independence

Definition (Rosen p. 443)

The events E and F are independent if and only if

$$p(E \cap F) = p(E)p(F).$$

Independence

Definition (Rosen p. 443)

The events E and F are independent if and only if

$$p(E \cap F) = p(E)p(F).$$

Example

Assume that the four ways a family can have two children is equally likely. Are the events E, that a family with two children has two boys, and F, that they have at least one boy, independent?

Independence

Definition (Rosen p. 443)

The events E and F are independent if and only if

$$p(E \cap F) = p(E)p(F).$$

Example

Assume that the four ways a family can have two children is equally likely. Are the events E, that a family with two children has two boys, and F, that they have at least one boy, independent?

Remember, we have that $p(E) = \frac{1}{4}$, $p(F) = \frac{3}{4}$ and $p(E \cap F) = \frac{1}{4}$.

Independence

Definition (Rosen p. 443)

The events E and F are independent if and only if

$$p(E \cap F) = p(E)p(F).$$

Example

Assume that the four ways a family can have two children is equally likely. Are the events E, that a family with two children has two boys, and F, that they have at least one boy, independent?

Remember, we have that $p(E) = \frac{1}{4}$, $p(F) = \frac{3}{4}$ and $p(E \cap F) = \frac{1}{4}$.

Since $p(E)p(F) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$ we have that $p(E \cap F) \neq p(E)p(F)$ and therefore E and F are **not independent**.

Independence

Definition (Rosen p. 443)

The events E and F are independent if and only if

$$p(E \cap F) = p(E)p(F).$$

Example

Assume that the four ways a family can have two children is equally likely. Are the events E, that a family with two children has two boys, and F, that they have at least one boy, independent?

Remember, we have that $p(E) = \frac{1}{4}$, $p(F) = \frac{3}{4}$ and $p(E \cap F) = \frac{1}{4}$.

Since $p(E)p(F) = \frac{1}{4} \cdot \frac{3}{4} = \frac{3}{16}$ we have that $p(E \cap F) \neq p(E)p(F)$ and therefore E and F are **not independent**.

Exercise

For independent events E and F show that $p(E|F) = p(E)$.

Bayes' Theorem

Theorem (Rosen p. 455)

Let E and F be events from a sample space S such that $p(E) \neq 0$ and $p(F) \neq 0$. Then

$$p(F|E) = \frac{p(E|F)p(F)}{p(E)}$$

By showing that

$$p(E) = p(E|F)p(F) + p(E|\bar{F})p(\bar{F})$$

we can also express Bayes' theorem as

$$p(F|E) = \frac{p(E|F)p(F)}{p(E|F)p(F) + p(E|\bar{F})p(\bar{F})}$$

Proof: Black board...

Example with Bayes' Theorem (I)

Example (Rosen p.455)

We have **two boxes A and B**:

- Box A contains **2 green balls** and **7 red balls**.
- Box B contains **4 green balls** and **3 red balls**.

Bob selects a ball by

- first choosing one of the two **boxes** at random, and
- then selects one of the **balls** in this box at random.

If Bob has selected a red ball, what is the probability that he selected a ball from the first box?

Example with Bayes' Theorem (I)

Example (Rosen p.455)

We have **two boxes A and B**:

- Box A contains **2 green balls** and **7 red balls**.
- Box B contains **4 green balls** and **3 red balls**.

Bob selects a ball by

- first choosing one of the two **boxes** at random, and
- then selects one of the **balls** in this box at random.

If Bob has selected a red ball, what is the probability that he selected a ball from the first box?

Let R be the event that Bob has chosen a red ball and $\bar{R}$ is the event that Bob has chosen a green ball.

Let A be the event that Bob has chosen a ball from box A and $\bar{A}$ is the event that Bob has chosen a ball from box B.

Example with Bayes' Theorem (I)

Example (Rosen p.455)

We have **two boxes A and B**:

- Box A contains **2 green balls** and **7 red balls**.
- Box B contains **4 green balls** and **3 red balls**.

Bob selects a ball by

- first choosing one of the two **boxes** at random, and
- then selects one of the **balls** in this box at random.

If Bob has selected a red ball, what is the probability that he selected a ball from the first box?

Let R be the event that Bob has chosen a red ball and $\bar{R}$ is the event that Bob has chosen a green ball.

Let A be the event that Bob has chosen a ball from box A and $\bar{A}$ is the event that Bob has chosen a ball from box B.

We then want to find $p(A|R)$.

Example with Bayes' Theorem (II)

Example (Rosen p.455)

We then want to find $p(A|R)$ and have that

$$\begin{aligned} p(A) = p(\bar{A}) &= 1/2 \\ p(R|A) &= 7/9 \\ p(R|\bar{A}) &= 3/7. \end{aligned}$$

This means that

$$P(R) = p(R|A)p(A) + p(R|\bar{A})p(\bar{A}) = \frac{7}{9} \cdot \frac{1}{2} + \frac{3}{7} \cdot \frac{1}{2} = \frac{38}{63}.$$

Using Bayes' theorem we then get

$$p(A|R) = \frac{p(R|A)p(A)}{p(R)} = \frac{7/9 \cdot 1/2}{38/63} = \frac{49}{76} \approx 0.645$$

This means that the probability that Bob selected a ball from box A given that the selected ball was red is approximately 0.645*.*

Random variables

Definition (Rosen p. 446)

A random variable is a function $X : S \to \mathbb{R}$ from the sample space of an experiment to the set of real numbers. That is, a random variable assigns a real number to each possible outcome.

Note that a random variable is a function.
It is not a variable, and it is not random!

Definition

$p(X = r)$ is the probability that X takes the value r, that is

$$p(X = r) = p(\{s \in S : X(s) = r\}).$$

Bernoulli trial

Definition

A Bernoulli trial is a experiment that can only have two possible outcomes: **success** and **failure**.

Exercise

If $\theta \in [0, 1]$ is the probability of **success** in a Bernoulli trial, what is the probability of **failure**?

Bernoulli trial

Definition

A Bernoulli trial is a experiment that can only have two possible outcomes: **success** and **failure**.

Exercise

If $\theta \in [0, 1]$ is the probability of **success** in a Bernoulli trial, what is the probability of **failure**?

Example

Coin flipping is an example of a Bernoulli trial.

For instance H could be **success** and T could be **failure**.

Expected value

Definition (Rosen p. 463)

The **expected value**, also called the *expectation* or *mean*, of the random variable X on the sample space S is equal to

$$E(X) = \sum_{s \in S} p(s) X(s)$$

Theorem

Suppose that X is a random variable with range $X(S)$, and let $p(X = r)$ be the probability that the random variable X takes the value r, then

$$E(X) = \sum_{r \in X(S)} p(X = r) r$$

You can think of $E(X)$ as the mean value of X if you perform the experiment many times.

Expected value

Example (Rosen p. 463)

Let X be the number that comes up when a fair die is rolled. What is the expected value of X?

As X takes values in $\{1, 2, 3, 4, 5, 6\}$ with equal probability $1/6$, we get

$$E(X) = \frac{1}{6} \cdot 1 + \frac{1}{6} \cdot 2 + \frac{1}{6} \cdot 3 + \frac{1}{6} \cdot 4 + \frac{1}{6} \cdot 5 + \frac{1}{6} \cdot 6 = \frac{7}{2}$$

Expected value

Theorem

The expected number of successes when n mutually independent Bernoulli trials are performed, where θ is the probability of success on each trial, is $n\theta$.

The core of Pigaro implementation (simplified)

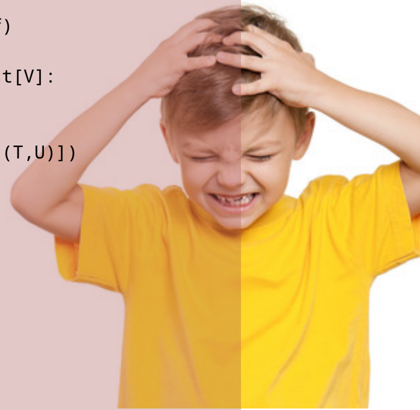
```scala
trait Dist[+T]: self =>
  protected def sample[U >: T](using RNG): IData[U] // IData is LazyList
  def sample[U >: T](n: SampleSize)(using RNG): IData[U] =
    this.sample(using summon[RNG]).take(n)

  def map[U](f: T => U): Dist[U] = new Dist[U]:
    def sample[V >: U](using rng: RNG): IData[V] = self.sample.map(f)

  def map2[U, V](other: Dist[U]) (f: (T,U) => V): Dist[V] =  new Dist[V]:
    def sample[W >: V](using rng: RNG): IData[W] =
      self.sample.map2(other.sample)(f)
  def zip[U](that: Dist[U]): Dist[(T,U)] = this.map2(that)(identity[(T,U)])
  infix def -> [U](that: Dist[U]): Dist[(T,U)] = this.zip(that)

  def flatMap[U](f: T => Dist[U]): Dist[U] = new Dist:
    def sample[S >: U](using RNG): IData[S] =
      self.sample.flatMap(t => f(t).sample(1))

  def filter(p: T => Boolean): Dist[T] = new Dist[T]:
    override def sample[S >: T](using rng: RNG): IData[S] =
      self.sample.filter(p)
  def condition(p: T => Boolean): Dist[T] = self.filter(p)
  def matching[A](f: PartialFunction[T, A]): Dist[T] = self.filter(f.isDefinedAt)
```



An example distribution: Uniform

```scala
case class Uniform[+T] (values: T*)
  extends Dist[T]:

  val rangeMin: Int = 0
  val rangeMax: Int = values.length - 1

  private lazy val indices = rangeMin to rangeMax
  lazy val valueMap: Map[Int, T] = (indices zip values).toMap

  // use an existing uniform generator from Spire
  // the Java builtin is statistically very weak!
  private lazy val gen = spire.random.Uniform(rangeMin, rangeMax)

  def sample[S >: T](using rng: RNG): IData[S] =
    gen.toLazyList(rng).map(valueMap)
```

Next time

- Interpreters of programming languages as computation in monads
- Reading: Phil Wadler's paper
- See you next week ...