Constructive Artificial Intelligence

Probability

Daniel Polani

School of Computer Science
University of Hertfordshire

November 10, 2009

Probabilities

Motivation

History:
- theory of gambling
- dealing with uncertainty
- Blaise Pascal and Pierre Fermat

Original Problem:
1. A game of luck is interrupted prematurely, pot contains money.
2. Question: what is a fair way to split the pot?

Scenario

Dice Game

Assume:

- A die is thrown.
- Possible outcomes are 1,2. . . 6.
- How much should a player get before throwing the dice if he wins
  1. on a 6?
  2. on a 2 or a 5?
  3. on a 3,4, or a 5?

Observations I

Elementary Outcomes

- $\mathcal{D} = \{1, 2, \ldots, 6\}$ are **mutually exclusive** outcomes and
- they are **complete** in the sense that the outcome of the dice throw must be one of them.
- If
  - the die is symmetric
  - and there is no reason to assume that one outcome is favoured over the other (fair dice)

  all outcomes should be treated equally
  (**Laplace's Principle of Insufficient Reason**)

Observations II

It Follows

- if the player wins on the 6, that win should not be different from a win on 1,2, . . . or 5.
- If the dice throw is repeated k times, with k large, $k^{(1)}, k^{(2)}, \ldots, k^{(6)}$ of the throws will show a $1, 2, \ldots, 6$, respectively.
- Note that $k^{(1)} + k^{(2)} + \cdots + k^{(6)} = k$. Since all outcomes should be treated equally, expect that $k^{(d)} \approx k/6$ for all outcomes $d \in \mathcal{D}$.

Fair Payouts

Make it Invariant

- to get rid of k which is an arbitrary parameter of the experiment, divide $k^{(d)}$ by k;
- for large k the number $p(d) := k^{(d)}/k$ becomes an invariant of the scenario

Fair Payouts

Make it Invariant

- to get rid of k which is an arbitrary parameter of the experiment, divide $k^{(d)}$ by k;
- for large k the number $p(d) := k^{(d)}/k$ becomes an invariant of the scenario

  Actually, this is an assumption, however we won't discuss it here;
- in our case of equivalent outcomes and large k, $p(d) = 1/6$ for all $d \in \mathcal{D}$.

Fair Payouts II

Fair Payout

Can now solve the fair payout problem:

- a player that wins exactly on $d = 6$ will win in $1/6$ of the experiments.
- for 1 unit in the pot, a fair payout will therefore consist of $1/6$ units.

Fair Payouts II

Fair Payout

Can now solve the fair payout problem:

- a player that wins exactly on $d = 6$ will win in $1/6$ of the experiments.
- for 1 unit in the pot, a fair payout will therefore consist of $1/6$ units.
- if the player wins on $d = 2$ or $d = 5$, same idea holds for $(k^{(2)} + k^{(5)})/k$ which is approximately $1/6 + 1/6 = 1/3$ — payout $1/3$ units per unit in the pot.
- if the player wins on $d = 3$, $d = 4$ or $d = 5$, player gets $1/2$ units per unit in the pot.

Probabilities

From the previous discussions, we come to the following definition:

Definition (Probabilities)

Consider a set $\mathcal{D} = \{d_1, d_2, \dots\}$ of elementary outcomes d_k. Then a **probability (distribution)** p assigns to each $d \in \mathcal{D}$ a real number $p(d)$ such that:

1. $p(d) \geq 0$ for all $d \in \mathcal{D}$.

   This corresponds to the count $k^{(d)}$ in above example being 0 or a positive integer.

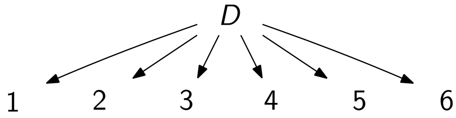
2. $\sum_{d \in \mathcal{D}} p(d) = 1$.

   This corresponds to the counts of the elementary events adding up to k: $k^{(d_1)} + k^{(d_2)} + \dots = k$.

Remark

1. When one has to distinguish different probabilities and there is danger of confusion, one attaches an index to the probability: p_D.
2. The index D is called **random variable** and identifies p_D and its domain $\mathcal{D}$.
3. We will use the index notation only if there is danger of confusion (and in programs). Else we write $p(d)$ instead of $p_D(d)$, $p(x)$ instead of $p_X(x)$ etc.

Elementary Outcomes of a Random Variable



Composite Outcomes

Definition (Composite Outcomes and their Probability)

If $\mathcal{D}$ is a set of elementary outcomes, then a **composite outcome** is a subset $U \subseteq \mathcal{D}$ of the set of elementary outcomes.

As special case, we can identify an elementary outcome d with the set $\{d\}$.

If a probability $p(d)$ is defined over the elementary outcomes $d \in \mathcal{D}$, we define the **probability of a composite outcome** U as $p(U) := \sum_{d \in U} p(d)$.

Remarks

1. In the gambler's example, the player winning on $d = 2$ or $d = 5$ can be considered the composite outcome $U = \{2, 5\} \subseteq \mathcal{D}$. One has $p(\{2, 5\}) = p(2) + p(5)$.

   The last observation follows directly from considering the dice outcome counts.

2. In general, from the definition above it is clear that for composite outcomes A and B, $p(A \cup B) = p(A) + p(B)$ if $A \cap B = \emptyset$.

3. As a special case, one has $p(\emptyset) = 0$.

Homework

1. Show that for general composite outcomes A and B, one has $p(A \cup B) = p(A) + p(B) - p(A \cap B)$.
2. What is $p(\mathcal{D})$, i.e. the probability of the whole set $\mathcal{D}$ of elementary outcomes?
3. What is $p(\mathcal{D} \setminus A)$?
4. In the gambler example, what is the probability of the (composite) outcome $\{1, 2\}$?
5. And what is the probability that this outcome does not occur?

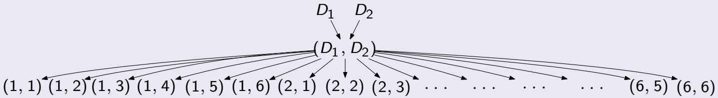
Hint: investigate how to express this situation in terms of composite outcomes.

Multiple Random Variables

Consider the dice problem again. Assume now that not one dice is thrown, but two.

1. Individually, the outcomes look as follows

2. Jointly, the outcomes are the following



3. From Laplace's principle: each of the 36 outcomes has probability $p(d_1, d_2) = 1/36$.

Marginalization

What if we want to ignore, say, the second die?

Filtering out one Variable

Assume we are only interested in a particular outcome d_1 of the first die (i.e. the random variable D_1). In the two-dice scenario, this corresponds to getting the probability of the composite outcome $\{(d_1, d_2) \mid d_2 \in \mathcal{D}\}$, i.e. the outcome where d_1 has a particular value, but d_2 can be anything. From this it follows that

$$p(d_1) = \sum_{d_2 \in \mathcal{D}} p(d_1, d_2)$$

Of course, in the die example $p(d_1) = 1/6$, consistent with the individual throw.

Remark

This is true in general. To get the probability of a subset of a subset of outcomes (i.e. random variables) from a joint probability, one sums the joint probability over the other variables. This is called **marginalization**.

Conditional Probabilities

This is in a way the "opposite" of marginalization.

Conditioning

- Want to know the probability of an outcome d_1 of the first die, with respect only to cases where the second die has particular value d_2.
- $p(d_1, d_2)$: total probability of having d_1 and d_2 occurring at the same time.
- However, now only interested in the probability of d_1 occurring in cases where the second die is d_2.
- For given d_2, consider set of elementary outcomes $\mathcal{D}_{|d_2} = \{(d_1, d_2) | d_1 \in \mathcal{D}_1\}$ as new outcome set.
- Want now the probabilities of d_1 with respect to this new elementary outcome set. Use a counting argument as before and arrive at:

Definition (Conditional Probability)

The **probability of** d_1 **conditioned on** d_2:

$$p(d_1 | d_2) := \frac{p(d_1, d_2)}{p(d_2)}$$

We obviously assume $p(d_2) \neq 0$.

Example

One has

$$p(d_1|d_2) = \frac{p(d_1, d_2)}{p(d_2)} = \frac{1/36}{1/6} = 1/6$$

for all d_1, d_2.

Remark

Obviously, the dice example is "boring" — one has $p(d_1|d_2) = p(d_1)$, or $p(d_1, d_2) = p(d_1|d_2)p(d_2) = p(d_1)p(d_2)$.

Definition (Independence)

If $p(d_1|d_2) = p(d_1)$, then the probability for d_1 and for d_1 conditioned (given) d_2 are the same and we say that the random variables D_1 and D_2 are **independent**.

Bayes' Theorem

Remark

Note that

$$p(d_2|d_1)p(d_1) = p(d_1, d_2) = p(d_1|d_2)p(d_2)\,.$$

From this follows

Theorem (Bayes')

One has:

$$p(d_2|d_1) = \frac{p(d_1|d_2)p(d_2)}{p(d_1)}\,.$$

Note

- Bayes' Theorem highly important: allows one to turn around the direction of a conditional. If conditional in one direction is known, the other can be inferred.
- Sufficient to compute $p(d_1|d_2)p(d_2)$; denominator obtained by normalization.

Example Dice Code (prob.py)

```
import random
from rational import Rational

p = {}

for i in xrange(1000000):
    die = random.randint(1,6)
    if not die in p: p[die] = 0
    p[die] += 1

Z = float(sum(p[k] for k in p))

for k in p: p[k] /= Z

for k in p: print k, Rational(p[k], eps=0.001)
```

Rational Numbers Approximation (rational.py) I

```
class Rational:
    def __init__(self, r, eps = 1e-5, thresh = 0.00):
        """eps is the absolute error, thresh the cutoff threshold. The
        absolute error cuts off the search once the precision is
        reached. The threshold allows to search for good truncation."""
        r_orig = r
        coeff = []
        while True:
            c = r//1.0
            coeff.append(c)
            r -= c                      # take the remainder
            num, den = self.reconstruct(coeff)

            if abs(r) < thresh: break      # break if good threshold reached
            if abs(float(num)/den - r_orig) < eps: break

            r = 1.0/r                   # get next term

        self.num = num
        self.den = den

    def reconstruct(self, coeff):
        coeff = coeff[:]
        num = 0
        den = 1
        while True:
            num = coeff.pop() * den + num
            if not coeff: break
            num, den = den, num
        return num, den
```

Rational Numbers Approximation (rational.py) II

```
def __repr__(self):
    return "%d/%d" % (self.num, self.den)
```

Judea Pearl.
Heuristics: Intelligent Search Strategies for Computer Problem-Solving.
Addison Wesley, 1984.

S. Russell and P. Norvig.
Artificial Intelligence: A Modern Approach.
Prentice Hall Series in Artificial Intelligence. Prentice Hall, 2 edition, 2002.