

COS 424: Interacting with Data

Lecturer: Dave Blei
Scribe: Ellen Kim

Lecture #2
February 7, 2008

1 Probability Review

1.1 Random Variables

A **random variable** can be used to model any probabilistic outcome.

Examples

- A coin toss is a random variable because the coin lands heads up ('heads') with probability of $1/2$ and lands tails up ('tails') with probability of $1/2$.
- The number of visitors to a store on a given day is not exactly a random variable but can be treated as one because the number of visitors is unknown.
- The temperature on 2/7/2013 is also a random variable. Again, it is not determined randomly, but it is unknown so it can be modeled as a random variable.
- The temperature on 3/4/1905 can be a random variable in the context of looking up the average temperature on that day, out of the temperatures of all other days.

1.1.1 Types of Random Variables

Random variables take on values in a sample space. These values depend on the type of random variable: discrete or continuous. **Discrete random variables** take on values from a finite sample space. **Continuous random variables** take on values from an infinite sample space.

Examples revisited

- A coin toss is a discrete random variable because the sample space is $\{H, T\}$, a finite set of values.
- The number of visitors to a store on a given day is also a discrete random variable because the sample space is $\{0, 1, 2, \dots\}$, a [n infinitely] finite set of values.
- The actual temperature on a given day or at a given time is a continuous random variable because the temperature's sample space is the set of real numbers Re .

1.1.2 Notation for Random Variables

A random variable is denoted by a capital letter (ex. X). A realization of a random variable is lowercase (ex. x).

Also, for probability notation using random variables, for this course, at least, just use lowercase 'p' for discrete random variables (ex. $p(X=x)$).

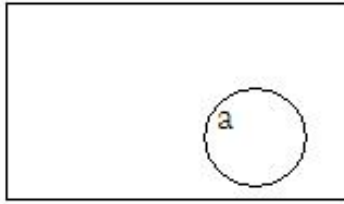


Figure 1: Diagram of the probability of an event a .

1.1.3 Discrete Distributions

For discrete distributions, the probabilities of all possible outcomes sums to 1.

$$\sum_x p(X = x) = 1$$

Example revisited An unfair coin toss:

$$p(X = h) = 0.7$$

$$p(X = t) = 0.3$$

1.2 Using Diagrams

It is convenient and easy to use visual diagrams to make working with probabilities more intuitive. In one representation, below, all atoms are in the box and areas can be drawn in the box to represent an event. An **event** is then a subset of atoms; events are drawn so that the proportion of the area of the event to the whole box approximates the probability of that event. The probability of an event is calculated by

$$\sum_x p(X = x) = p(a).$$

(More examples of diagrams will be used in the rest of this lecture.)

1.3 Distributions of Random Variables

While discrete and continuous random variables are interesting, useful, yay by themselves, typically, we consider **collections** of random variables, such as the following.

1.3.1 Joint Distribution

A **joint distribution** is a distribution over the configuration of all the random variables in an ensemble.

Example revisited again The tossing of four fair coins has 16 different possible outcomes (hhhh, hhht, ..., tttt). Each of these outcomes has a probability of occurring. In whole, the distribution of these probabilities is the joint distribution:

$$p(hhhh) = 0.0625$$

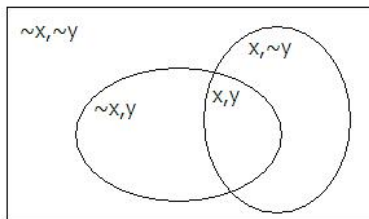


Figure 2: L2-2 multiple events.JPGL2-2 Events x, y

$$p(hhht) = 0.0625$$

...

$$p(tttt) = 0.0625.$$

1.3.2 Conditional Distribution

The **conditional distribution** is the distribution of a random variable given some evidence. The following notation is, for example, the probability that X will be x given that Y is y.

$$p(X = x|Y = y).$$

Example (a new one) David really likes the band Steely Dan [and Bread], but David's wife really does not. The probability that David listens to Steely Dan is

$$p(\text{David listens to Steely Dan}) = 0.5$$

but the probability that David listens to Steely Dan given that his wife Toni is home is

$$p(\text{David listens to Steely Dan}|\text{Toni is home}) = 0.1$$

and the probability that David listens to Steely Dan given that Toni is not at home is

$$p(\text{David listens to Steely Dan}|\text{Toni is not home}) = 0.7.$$

Note that .1 plus .7 does not equal 1, and this is fine, as long as there is a complete distribution for X for each possible outcome of Y. That is, it is fine that

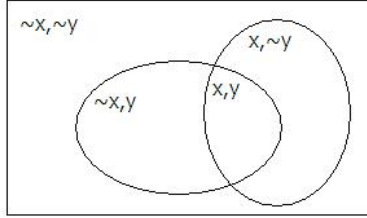
$$\sum_y p(X = x|Y = y) \neq 1$$

as long as

$$\sum_x p(X = x|Y = y) = 1$$

because the latter is necessarily true.

Diagramming multiple events If we know a condition y is true, the area of y “becomes the box” – you are not considering “all atoms” to be just the atoms in event y.



Conditional Probability The **conditional probability** of an event x given that an event y has occurred is given by

$$p(X = x|Y = y) = \frac{p(X = x, Y = y)}{p(Y = y)}$$

and holds when $p(Y=y)$ is greater than 0.

1.4 Manipulating probabilities

1.4.1 Chain Rule

The **chain rule** is useful for many things. For instance, in diagnosing a patient in medicine, if Y is a disease and X is a symptom, knowing the frequency (probability) at which X and Y occur, once can find the joint distribution. The chain rule is

$$p(X = x, Y = y) = p(X, Y) = \frac{p(X, Y) * p(Y)}{p(Y)} = p(X|Y) * p(Y)$$

so that in general,

$$p(X_1, X_2, \dots, X_N) = p(X_1) \prod_{n=2}^N p(X_n|X_1, X_2, \dots, X_{n-1}).$$

1.4.2 Marginalization

Given random variables, we are often only interested in a subset, so we use **marginalization**. We will verify that this is mathematically correct below:

$$\begin{aligned} p(X) &= \sum_y \sum_z p(X, Y = y, Z = z) \\ &= \sum_y \sum_z p(X, y, z) \\ &= p(X) \sum_y \sum_z p(Y = y, Z = z|X) \\ &= p(X) * p(Y = y, Z = z|X) \end{aligned}$$

1.4.3 Bayes' Rule

$$\begin{aligned} p(Y|X) &= \frac{p(X|Y) * p(Y)}{\sum_y p(X|Y=y) * p(Y=y)} \\ &= \frac{p(X, Y)}{p(X)} \end{aligned}$$

Bayes' Rule is from chain rule, marginalizing out Y in the denominator, and the definition of conditional probability.

Example revisited Going back to the disease example, now, knowing the symptom and the probability of the symptom when the disease is present, we can find the probability that the patient has the disease given the presence of the symptom.

An important application that will not be discussed here is Bayesian statistics.

1.4.4 Independence

Random variables are **independent** if knowing one doesn't tell us anything about the other.

$$p(X|Y=y) = p(X) \text{ for all } y.$$

This means that if X and Y are independent, the joint probability factorizes:

$$p(X, Y) = p(X|Y) * p(Y) = p(X) * p(Y).$$

An implication of this is that we can use 2 5-vectors rather than a 5x5 grid. Another is that we can collect data for 2 factors independently and still find joint probabilities.

Examples The following are examples of independent variables:

- whether it rains and who the president will be
- the outcome of 2 tossed dice

The following examples are of variables that are not independent:

- whether it rains and whether we go to the beach
- height and sex of a person
- the results of drawing without replacement from a deck of cards
- whether the alarm clock goes off and getting to class on time

1.4.5 Conditional Independence by Example

Suppose there are two coins, one fair and the other unfair, so that

$$p(C_1 = H) = 0.5 \text{ and } p(C_2 = H) = 0.7.$$

Now choose one coin Z at random, so

$$Z \in \{1, 2\}.$$

Then flip C_Z twice to get two outcomes X, Y .

Question: Are X and Y independent?

Answer: Yes, if it is known whether Z is coin 1 or coin 2, but we don't know which Z is. So the answer is no, because if the first flip comes up heads, then it is more likely to be $Z=2$ so on the next flip heads is expected more than tail; if the first flip is tails, then it is more likely to be $Z=1$ so on the next flip tails is expected again with probability $1/2$.

X and Y are **conditionally independent** given Z .

$$p(Y|X, Z = z) = p(Y|Z = z)$$

Again, this implies a factorization,

$$p(Y, X|Z = z) = p(Y|Z = z) * p(X|Z = z).$$

1.5 Monty Hall problem for thought

Monty Hall shows you 3 doors with 2 goats and 1 prize behind them. You, the contestant, pick one. Then he shows you one of the others which is a goat; you can stay with the door you chose initially or pick the other remaining door (the door you neither picked nor he revealed as with goat). What should you do? (answer in next lecture...)

1.6 Announcement

Using R help session on Monday at 5:30 PM in the small auditorium in the Computer Science building.

Also, please sit in the front of the hall during Tuesday and Thursday lectures.