L02s01

September 13, 2018

1 Basic Probability Review

Ref

[1] https://github.com/AM207/2016

```
In [1]: ## do all neccessary imports
        import numpy as np
        import matplotlib
        %matplotlib inline
        import matplotlib.pyplot as plt
        import seaborn as sns
        sns.set_style("white")
        sns.set_context("notebook")
        import pandas as pd
        from  scipy import stats

        import math
```

2 Short review of probability theory, distributions etc

Probabilities are numbers that tell us how often things happen (frequentist) or our believe in a different outcome or notion (Bayesian). These two views of probabilities (and which one is preferable) are a very disputed topic. In this course we take a very simple approach and we use whatever is convenient at the time, or whatever is better to demonstrate the methodologies we are learning.

There are a few basic rules that form the basis for all probabilistic methodology. In this lecture, we just briefly review these rules. For an in depth introduction into probability theory, you can for example access the stat110 material online, read Joe Blitzstein's Introduction to Probability, or also Larrry Wasserman's All of Statistics.

Let us define some terminology: X and Y are two events and $p(X)$ is the probability of the event X to happen. $X^c $ is the complement of X, the event which is all the occurrences which are not in X. $X \cup Y$ is the union of X and Y; $X \cap Y$ is the intersection of X and Y. (Both $X \cup Y$ and $X \cap Y$ are also events.)

ex

$$X = \{ \text{ students who has his/her cell phone} \}$$

2.0.1 The very fundamental rules of probability:

1. $p(X) = 1$ X has to happen almost surely
2. $p(X) = 0$ X will certainly not happen almost surely
3. $0 p(X) 1$ X has probability range from low to high
4. $p(X) + p(X^c) = 1$ X must either happen or not happen
5. $p(X \cup Y) = p(X) + p(Y) p(X \cap Y)$ X can happen and Y can happen but we must subtract the cases that are happening together so we do not over-count.

For two random variables x, y the $p(x, y)$ is called the joint distribution, and $p(x|y)$ the conditional distribution.

2.0.2 Sum rule (marginal distribution)

We can write the marginal probability of x as a sum over the joint distribution of x and y where we sum over all possibilities of y,

$$p(x) = \sum_y p(x, y)$$

for continuous random variables this becomes:

$$p(x) = \int_y p(x, y)\, dy$$

2.0.3 Product rule

We can rewrite a joint distribution as a product of a conditional and marginal probability,

$$p(x, y) = p(x|y) p(y)$$

2.0.4 Chain rule

The product rule is applied repeatedly to give expressions for the joint probability involving more than two variables. For example, the joint distribution over three variables can be factorized into a product of conditional probabilities:

$$p(x, y, z) = p(x|y, z)\, p(y, z) = p(x|y, z)\, p(y|z) p(z)$$

2.0.5 Bayes rule

Given the product rule one can derive the Bayes rule, which plays a central role in a lot of the things we will be talking about:

$$p(y|x) = \frac{p(x|y)\, p(y)}{p(x)} = \frac{p(x|y)\, p(y)}{\sum_{y'} p(x, y')} = \frac{p(x|y)\, p(y)}{\sum_{y'} p(x|y') p(y')}$$

2.0.6 Independence

Two variables are said to be independent if their joint distribution factorizes into a product of two marginal probabilities:

$$p(x, y) = p(x)\, p(y)$$

Note that if two variables are uncorrelated, that does not mean they are statistically independent. There are many ways to measure statistical association between variables and correlation is just one of them. However, if two variables are independent, this will ensure there is no correlation between them. Another consequence of independence is that if x and y are independent, the conditional probability of x given y is just the probability of x:

$$p(x|y) = p(x)$$

In other words, by conditioning on a particular y, we have learned nothing about x because of independence. Two variables x and y and said to be conditionally independent of z if the following holds:

$$p(x, y|z) = p(x|z)p(y|z)$$

Therefore, if we learn about z, x and y become independent. Another way to write that x and y are conditionally independent of z is

$$p(x|z, y) = p(x|z)$$

In other words, if we condition on z, and now also learn about y, this is not going to change the probability of x. It is important to realize that conditional independence between x and y does not imply independence between x and y.

2.1 Distributions

A probability distribution (aka probability measure) is a function that takes an event and gives its probability. Sometimes this is not well defined in the whole probability space but we will not worry about this for now. There are two classes of statistical distributions, discete and continous.

2.1.1 Discrete distributions

2.1.2 Bernoulli distribution

The probability for a yes/no outcome of an experiment, is given by the Bernoulli distribution. The Bernoulli distribution is the mother of all distributions. Every experiment, can always be expressed in terms of success/failure. If you do not know which distribution to use, you can think of any problem as a yes/no problem and starting from there you work your way to all other distributions.

Let k be the outcome of our experiment. Then p is the probability of a success ($k = 1$), which means we expect the value $k = 1$, pn times out of n. The probability for a failure ($k = 0$) is $1 - p$. For example if we have $p = 0.6$ and we repeat our experiment 10 times, we expect 6 experiments to be successful and 4 to be not successful.