COMP 441 — Large Scale Machine Learning	Jan 11, 2016
Lecture 1	
Lecturer: Anshumali Shrivastava	*Scribe By: Anshumali Shrivastava*

This scribe may contain errors, please do not cite. Please email if you find any errors.

1 Some Probability Recap

Definition: A random variable X is a variable whose value is stochastic, i.e. it depends on the outcome (or future) of some experiments. For example, if we toss a fair coin and define a variable X such that

$$X = \begin{cases} 1, & \text{if the coin toss shows heads} \\ 0, & \text{otherwise.} \end{cases},$$

then X is a random variable.

For discrete random variable X, every possible value of X is associated with a chance or a probability value. In the above case, $X = 1$ has associated probability of 0.5 because the coin is fair. Similarly $X = 0$ also has the associated probability value of 0.5. For continuous random variables we have a density function ϕ associated with the random variable. $\phi(x)$ can be thought of as the probability value associated with $X = x$. The definition for continuous case is not formal in the strictest sense, as we need familiarity with measure theory which we will skip.

Since random variables, such as X, do not have a fixed value one important practical quantity of interest is the expectation $\mathbb{E}(X)$. Expectation of X is defined as weighted summation of all possible values of the given random variable X, weighted by its probability.

$$\mathbb{E}(X) = \begin{cases} \sum_{v \in Values} v \times Pr(X = v), & \text{discrete case} \\ \int_{-\infty}^{\infty} \phi(t)dt, & \text{continous case.} \end{cases}$$

1.1 Some Useful Identities about Expectations.

$$\mathbb{E}(X + Y) = \mathbb{E}(X) + \mathbb{E}(Y) \tag{1}$$

$$\mathbb{E}(aX) = a\mathbb{E}(X) \quad \text{a is constant} \tag{2}$$

$$\mathbb{E}(XY) = \mathbb{E}(X)\mathbb{E}(Y) \quad \text{iff X and Y are independent} \tag{3}$$

2 Estimating Turtles

Suppose you are hired to estimate the number of turtles in a giant pond. A reasonable strategy is the following: Capture k turtles (first sample), mark them and release them in the pond. Then after some days again catch k (need not be k, could be a different number) turtles (second sample) and count the number of marked turtles, call it $M \leq k$.

M is a random variable and it is related to the total number of turtles, call it n. Let us compute $\mathbb{E}(M)$. A good way of getting tackle on M, is by making use of indicator variables.

Let us name the marked turtles from 1 to k. Now, define indicator variables $z_{i,j}$ as follows

$$z_{i,j} = \begin{cases} 1, & \text{if the turtle named } i \text{ occupies } j^{th} \text{ position in the second sample} \\ 0, & \text{otherwise.} \end{cases},$$

If we assume that the second sample of k turtles is a perfectly random draw from the total n turtles (Is this a good assumption ? under no extra information this is our best bet anyway). Then the probability that $z_{i,j} = 1$ is $\frac{1}{n}$. (Why ?)

Now, we can write M as

$$M = \sum_{i,k} z_{i,j}.$$

Therefore by linearity of expectation, we have

$$\mathbb{E}(M) = \mathbb{E}(\sum_{i,k} z_{i,j}) = \sum_{i,j} \mathbb{E}(z_{i,j}) = \frac{k^2}{n}.$$

Therefore, once we observe the value of M, a reasonable estimate for n is $\frac{k^2}{M}$. We can show that if k is large enough then this is actually a good estimate.
Note that $\mathbb{E}(\frac{k^2}{M}) \neq n$ because $\mathbb{E}(\frac{1}{M}) \neq \frac{1}{\mathbb{E}(M)}$