CmpE 343 Lecture Notes
3: Random Variables and Probability Distributions

Ethem Alpaydın

September 28, 2014

1 Random Variable

There are many different type of random experiments and each sample space may be a set of different type of outcomes. For examples, in a coin toss, it is heads or tails; in drawing a card from a deck, it is a suit (clubs, diamonds, hearts, or spades) and a rank (from Ace to King); in rolling a die, it is the face between one and six; or it can be a numeric value, for example, the weight in kilograms of a randomly chosen college student.

For uniform treatment, we use the concept of a *random variable* where each element of the sample space is represented by a real number. For example, the random variable X can represent heads with 0 and tails with 1. Then for example if this is a fair coin, we write $P\{X = 0\} = 0.5$.

Let us say we have a random experiment where we toss the coin three times and we are interested in the number of heads in three tosses, then the random variable Y stands for the number of tosses and is already a number. Then for each value of Y between zero and three, we can write the corresponding pairs of value and probability as a table:

y	$P(Y = y)$	Corresponding outcomes
0	$1/8$	TTT
1	$3/8$	HTT, THT, TTH
2	$3/8$	HHT, HTH, THH
3	$1/8$	HHH

Here, Y is a *discrete* random variable and we define the *probability distribution* (also called the *probability mass function* when the random variable is discrete) as $f(y) = P\{Y = y\}$. Note that $f(y) \geq 0$ for all y and that $\sum_y f(y) = 1$.

In certain cases, it is easier to work with the *cumulative distribution* $F(y)$ defined as

$$F(y) = \sum_{v \leq y} f(v) \tag{1}$$

For the example above, we have

$$F(y) = \begin{cases} 0 & \text{if } y < 0 \\ 1/8 & \text{if } 0 \leq y < 1 \\ 4/8 & \text{if } 1 \leq y < 2 \\ 7/8 & \text{if } 2 \leq y < 3 \\ 1 & \text{if } y \geq 3 \end{cases}$$

Note that although $f(y)$ is defined only for integer values between zero and three, $F(y)$ is defined for all $y \in \Re$.

For a continuous random variable X, $f(x)$ is the *probability density function* which satisfies $f(x) \geq 0, x \in \Re$ and $\int_{-\infty}^{\infty} f(x)dx = 1$. A continuous random variable X never takes a value but is queried for an interval to get a probability:

$$P\{a < X < b\} = \int_a^b f(x)dx \tag{2}$$

Similar to the discrete case, $F(x) = PX \leq x$, and here we integrate instead of the discrete summation:

$$F(x) = \int_{-\infty}^{x} f(v)dv \tag{3}$$

which implies that $P\{a < X < b\} = F(b) - F(a)$ and $f(x) = \frac{dF(x)}{dx}$ if the derivative exists.

2 Joint Probability Distributions for Two or More Random Variables

In certain random experiments, we are interested in two outcomes defining two random variables. For example, let us say we have a bag containing two red balls, three black balls, and three blue balls, we draw two balls at random and we are interested in the number of red and blue balls. In this case, we have the *joint probability distribution* which again can be discrete or continuous depending on the type of the random variables.

For the discrete case, we have $P\{X = x, Y = y\} = f(x, y)$ which satisfies $f(x, y) \geq 0, \forall x, y$ and $\sum_x \sum_y f(x, y) = 1$—for the continuous case, replace $\sum$ by $\int$ here and below.

The *marginal distribution* of one random variable can be found by summing up over the other:

$$f(x) = \sum_y f(x, y) \text{ and } f(y) = \sum_x f(x, y) \tag{4}$$

The *conditional distribution* of one random variable given the other can be found using the joint and marginal distributions:

$$f(x|y) = \frac{f(x, y)}{f(y)} \text{ and } f(y|x) = \frac{f(x, y)}{f(x)} \tag{5}$$

If X and Y are independent, $f(x|y) = f(x)$ and $f(y|x) = f(y)$, so

$$f(x, y) = f(x)f(y) \tag{6}$$

These formula can be generalized to more than two random variables. In the simplest *naive* case, all are independent:

$$f(x_1, x_2, \ldots, x_d) = f(x_1)f(x_2)\cdots f(x_d)$$

This is a case that is frequently used in many applications when d is large and it is complicated to estimate the different dependencies. But of course when there are dependencies, they should be used. For example, we can have:

$$f(x_1, x_2, x_3, x_4) = f(x_4)f(x_3|x_4)f(x_2|x_4)f(x_1|x_2, x_3)$$

Then given this joint, we can do all sorts of calculations, such as

$$f(x_1, x_3) = \sum_{x_2, x_4} f(x_1, x_2, x_3, x_4)$$

$$f(x_3|x_4, x_2) = \frac{f(x_2, x_3, x_4)}{f(x_2, x_4)}$$

When the number of variables is large and each has many possible values, these calculations may take too much time and researchers have come up with smart data structures and algorithms to represent and manipulate such dependencies in an effective manner—they are known as *graphical models*.