

## Probability Recap

Lectures: Edo Liberty and Matthijs Douze

**Warning:** Please do not cite this note as a peer reviewed source. Please submit requests and corrections as issues or pull requests at [github.com/edoliberty/vector-search-class-notes](https://github.com/edoliberty/vector-search-class-notes)

## 1 Preliminaries

A variable  $X$  is a random variable if it assumes different values according to a probability distribution. For example,  $X$  can denote the outcome of a three sided die throw. The variable  $X$  takes the values  $x = 1, 2, 3$  with equal probabilities.

The expectation of  $X$  is the sum over the possible values times the probability of the events.

$$\mathbb{E}[X] = \sum_{x=1}^3 x \Pr(X = x) = 1 \frac{1}{3} + 2 \frac{1}{3} + 3 \frac{1}{3} = 2 \quad (1)$$

Continuous random variable are described by their distribution function  $\varphi$ .

$$\Pr[Y \in [a, b]] = \int_a^b \varphi(t) dt.$$

For a function  $\varphi$  to be a valid distribution we must have:

$$\forall t, \varphi(t) \geq 0 \quad (\text{where it is defined}) \quad (2)$$

$$\int_a^b \varphi(t) dt \quad \text{is well defined for all } a \text{ and } b \quad (3)$$

$$\int_{-\infty}^{\infty} \varphi(t) dt = 1 \quad (4)$$

For example consider the continuous variable  $Y$  taking values in  $[0, 1]$  uniformly. That means  $\varphi(t) = 1$  if  $t \in [0, 1]$  and zero else. This means that the probability of  $Y$  being in the interval  $[t, t + dt]$  is exactly  $dt$ . And so the expectation of  $Y$  is:

$$\mathbb{E}[Y] = \int_{t=0}^1 t \varphi(t) dt = \int_{t=0}^1 t \cdot dt = \frac{1}{2} t^2 \Big|_0^1 = 1/2 \quad (5)$$

**Remark 1.1.** Strictly speaking, distributions are not necessarily continuous or bounded functions. In fact, they can even not be a function at all. For example, the distribution of  $X$  above includes three Dirac- $\delta$  functions which are not valid functions.

### 1.1 Dependence and Independence

A variable  $X$  is said to be *dependent* on  $Y$  if the distribution of  $X$  given  $Y$  is different than the distribution of  $X$ . For example. Assume the variable  $X$  takes the value 1 if  $Y$  takes a value of less than  $1/3$  and the values 2 or 3 with equal probability otherwise ( $1/2$  each). Clearly, the probability of  $X$  assuming each of its

values is still  $1/3$ . however, if we know that  $Y$  is 0.7234 the probability of  $X$  assuming the value 1 is zero. Let us calculate the expectation of  $X$  given  $Y$  as an exercise.

$$\mathbb{E}(X|Y) = \sum_{x=1}^3 x \Pr(X = x|Y \leq 1/3) = 1 \cdot 1 \quad (6)$$

$$\mathbb{E}(X|Y) = \sum_{x=1}^3 x \Pr(X = x|Y > 1/3) = 1 \cdot 0 + 2 \cdot \frac{1}{2} + 3 \cdot \frac{1}{2} = 2.5 \quad (7)$$

$E(X|Y) = 1$  for  $y \in [0, 1/3]$  and  $E(X|Y) = 2.5$  for  $y \in (1/3, 1]$ .

Remember:  $\mathbb{E}(X|Y)$  is a function of  $y$ !

**Definition 1.1** (Independence). *Two variables are said to be Independent if:*

$$\forall y, \Pr[X = x|Y = y] = \Pr[X = x].$$

*They are dependent otherwise.*

**Fact 1.1.** *If two variables  $X$  and  $Y$  are Independent then so are  $f(X)$  and  $g(Y)$  for any functions  $f$  and  $g$ .*

**Fact 1.2** (Linearity of expectation 1). *For any random variable and any constant  $\alpha$ :*

$$\mathbb{E}[\alpha X] = \alpha \mathbb{E}[X] \quad (8)$$

**Fact 1.3** (Linearity of expectation 2). *For any two random variables*

$$\mathbb{E}_{X,Y}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y] \quad (9)$$

*even if they are dependent.*

**Fact 1.4** (Multiplication of random variables). *For any two **independent** random variables*

$$\mathbb{E}_{X,Y}[XY] = \mathbb{E}[X]\mathbb{E}[Y] \quad (10)$$

*This does not necessarily hold if they are dependent.*

**Definition 1.2** (Variance). *For a random variable  $X$  we have*

$$\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2 \quad (11)$$

*The standard deviation  $\sigma$  of  $X$  is defined to be  $\sigma(X) \equiv \sqrt{\text{Var}[X]}$ .*

**Definition 1.3** (Additivity of variances). *For any two **independent** variables  $X$  and  $Y$  we have*

$$\text{Var}[X + Y] = \text{Var}[X] + \text{Var}[Y] \quad (12)$$

**Fact 1.5** (Markov's inequality). *For any positive random variable  $X$ :*

$$\Pr(X > t) \leq \frac{\mathbb{E}[X]}{t} \quad (13)$$

**Fact 1.6** (Chebyshev's inequality). *For any random variable  $X$*

$$\Pr[|X - \mathbb{E}[X]| > t] \leq \frac{\sigma^2(X)}{t^2} \quad (14)$$