

Table 1: Relation between hair color and eye color.

	Hair color		
	Fair/Red	Medium	Dark/black
Eye color			
Light	1168	825	305
Dark	573	1312	1200

## 1 Joint distributions and independence

A joint probability is as the name implies the probability of ‘joining’ the probability of two events i.e.,  $P(A \cap B)$  (the intersection of  $A$  and  $B$ ). Take for example this table below:

To investigate the relations between hair color and eye colour, the hair color and eye color of 5383 was recorded. The data are given in Table 3. Eye color is encoded by the values 1 (Light) and 2 (Dark), and hair color by 1 (Fair/red), 2 (Medium), and 3 (Dark/black). By dividing the numbers in the table by 5383, the table is turned into a joint probability distribution for random variables  $X$  (hair color) taking values 1 to 3 and  $Y$  (eye color) taking values 1 and 2.

### 1.1 Joint probability distribution

We determine the joint probability density distribution by dividing the eye and hair color combinations by the total amount of records (5383) giving us this table (Table 4):

```
pmf <- matrix(round(c(1168, 825, 305, 573, 1312, 1200) / 5383, 3),
  byrow = TRUE, nrow = 2)
rownames(pmf) <- c("1 ", "2 ")
colnames(pmf) <- c("1", "2", "3")
dimnames(pmf) <- list("Y" = rownames(pmf), "X" = colnames(pmf))

kable(pmf,
  caption = "Joint probability distributions of X and Y", booktabs = T) %>%
  kable_styling(latex_options = "HOLD_position")
```

Table 2: Joint probability distributions of  $X$  and  $Y$

	1	2	3
1	0.217	0.153	0.057
2	0.106	0.244	0.223

### 1.2 Marginal probability distribution

In order to determine the marginal probability distribution of  $X$  and  $Y$ , we have to marginalize for  $Y$  and  $X$  respectively which we do by finding  $P(X = x)$  and  $P(Y = y)$  which is done by summing all the values that  $Y$  and  $X$  can take. See table 5

```
marg <- rbind(pmf, colSums(pmf))
marg2 <- cbind(marg, rowSums(pmf))
```

```
## Warning in cbind(marg, rowSums(pmf)): number of rows of result is not a multiple
## of vector length (arg 2)
```

```

rownames(marg2) <- c("1 ", "2 ", "Px ")
colnames(marg2) <- c("1", "2", "3", "Py ")
marg2[3, 4] <- 1

kable(marg2, format = "latex",
      caption = "Marginal probability distributions of X and Y", booktabs = T) %>%
  row_spec(2, extra_latex_after = "\\cline{1-5}") %>%
  column_spec(5, border_left = T) %>%
  kable_styling(latex_options = "HOLD_position")

```

Table 3: Marginal probability distributions of X and Y

	1	2	3	Py
1	0.217	0.153	0.057	0.427
2	0.106	0.244	0.223	0.573
Px	0.323	0.397	0.280	1.000

### 1.3 Independence between two variables

In order to find out whether  $X$  and  $Y$  are independent, we have to check that  $P(X = x, Y = y) = P(X = x)P(Y = y)$  for all  $x$  and  $y$ .

We'll start by checking for  $x = 1$  and  $y = 1$ :

$$P(X = 1, Y = 1) = 0.217$$

$$P(X = 1)P(Y = 1) = 0.057 \cdot 0.323 = 0.018411$$

Because  $P(X = x, Y = y) \neq P(X = x)P(Y = y)$ . We can say that  $X$  and  $Y$  are dependent