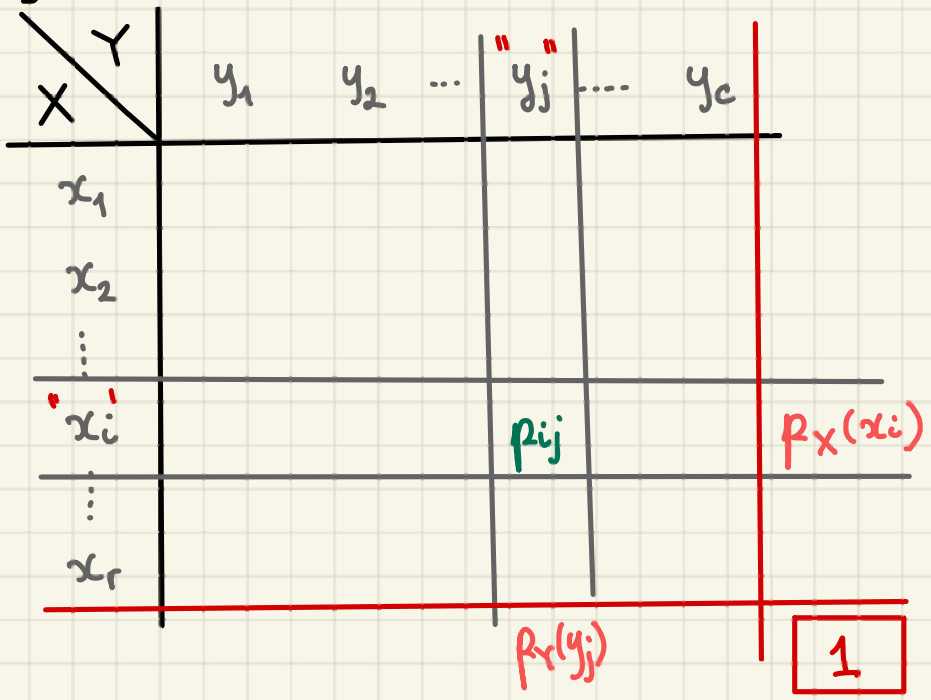
M358K : November 16th, 2020.

χ^2 test of independence

X, Y } categorical random variables



dimension: $r \times c$

Joint probability mass function

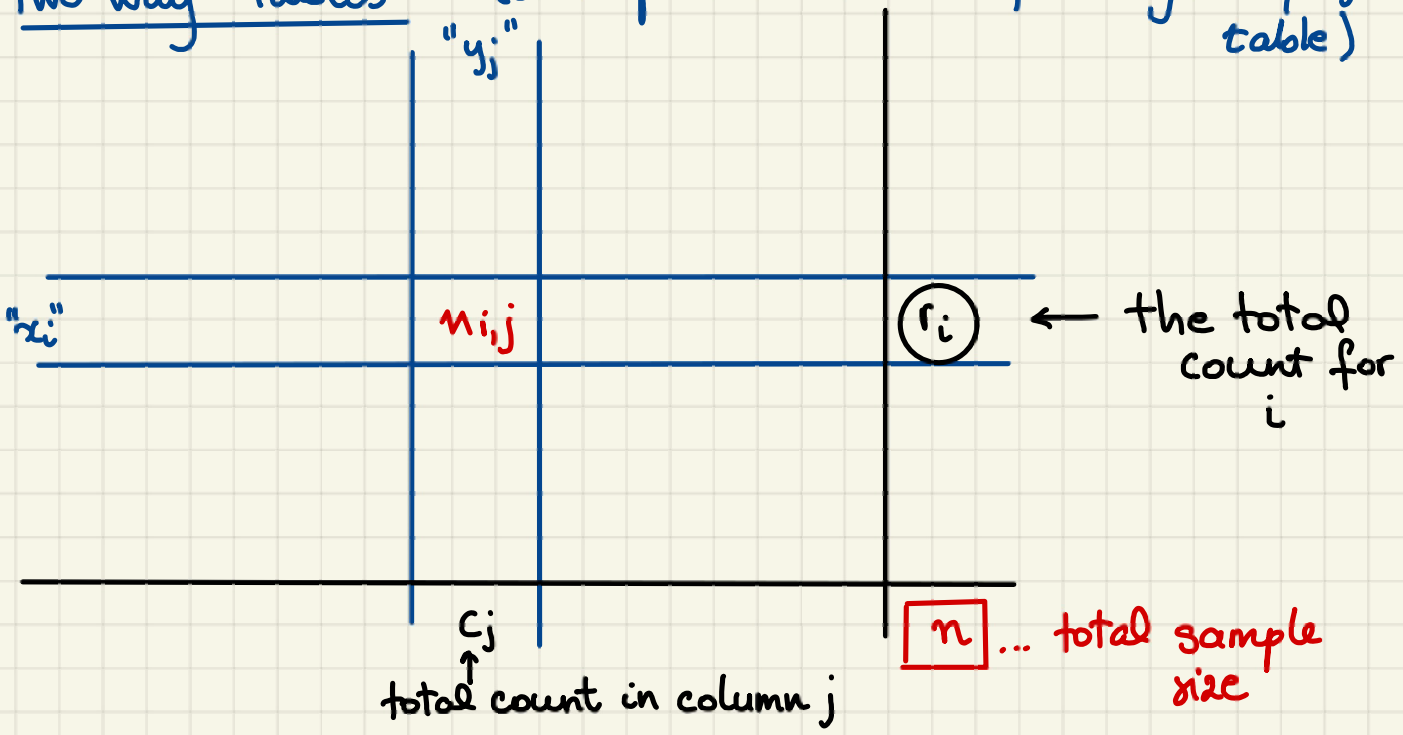
$$\mathbb{P}[X = x_i, Y = y_j] =: p_{ij} \quad \text{for all } i,j$$

If X and Y are INDEPENDENT:

$$p_{ij} = \underbrace{\mathbb{P}[X = x_i]}_{p_X(x_i)} \cdot \underbrace{\mathbb{P}[Y = y_j]}_{p_Y(y_j)} \quad \text{for all } i,j$$

marginal dist'ns

Two-way tables (an empirical version of the joint pmf table)



n_{ij}... # of observed cases w/ the combination "x_i", "y_j"

Empirically:

Q: What is the probability that you land in row i?

$$\frac{r_i}{n}$$

Q: What is the probability that you land in column j?

$$\frac{c_j}{n}$$

If the row and column effects (random variables) are independent, then the probability of landing in cell (i,j) is $\frac{r_i}{n} \cdot \frac{c_j}{n} = \frac{r_i \cdot c_j}{n^2}$.

⇒ The expected count in cell (i,j) if the two effects are independent is

$$n \cdot \frac{r_i \cdot c_j}{n^2} = \boxed{\frac{r_i \cdot c_j}{n}}$$