M358K: November 12th, 2021.

χ^2-test of Independence.

Independence of categorical random variables.

Let X and Y be two categorical r.v.s.

Their joint probability mass function can be represented as:

$X \backslash Y$	y_1	y_2	$\cdots$	y_j	$\cdots$	y_c	
x_1							
x_2							
$\vdots$							
x_i				p_{ij}			$p_X(x_i) := \sum_{j=1}^{c} p_{ij}$
$\vdots$							
x_r							
				$p_Y(y_j) = \sum_{i=1}^{r} p_{ij}$			$\boxed{1}$

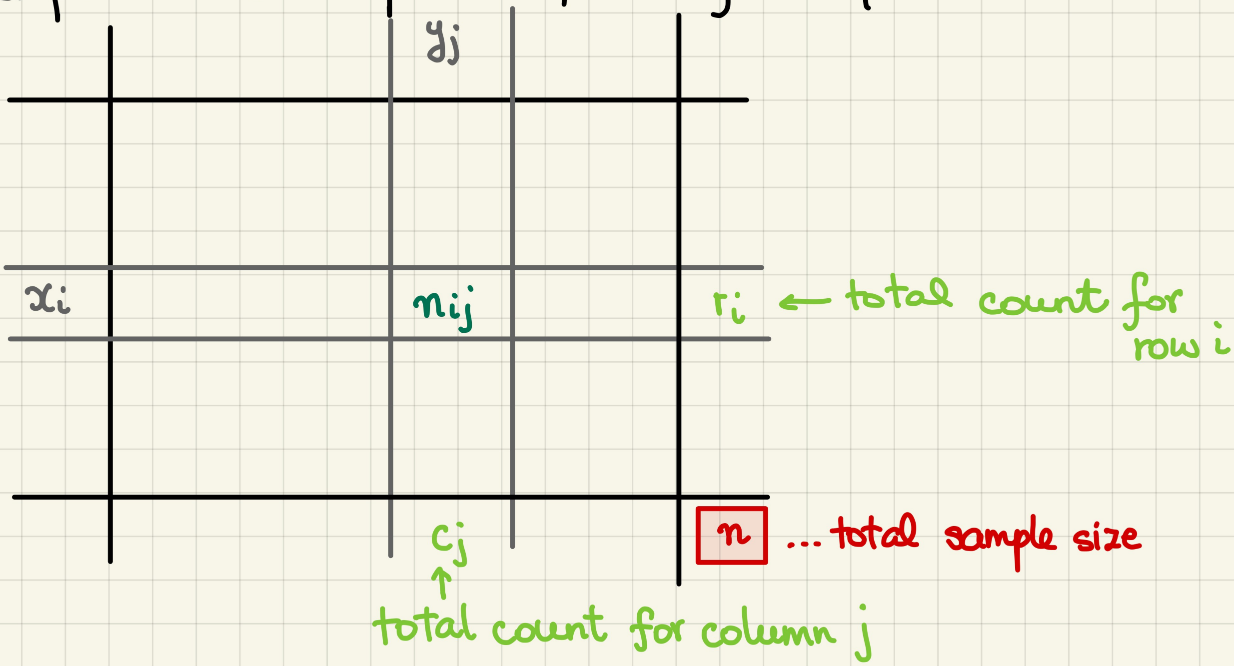
marginal distributions

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

$$\boxed{\begin{array}{c} X \text{ and } Y \text{ are independent} \\ \Longleftrightarrow \\ p_{ij} = p_X(x_i) \cdot p_Y(y_j) \\ \text{for all } i,j \end{array}}$$

Two-way Tables

... empirical counterparts of the joint pmf table



n_{ij} ... the # 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?

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

Q: If the row and column r.v.s were independent, what would be the probability of landing in the cell (i,j)?

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

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

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