Esssentials of Applied Data Analysis IPSA-USP Summer School 2017

Joint Distributions

Leonardo Sangali Barone
leonardo.barone@usp.br

jan/17

Joint Distribution

Causal theories of politics necessarily involve expected relationships among concepts or variables. As such, we want to study joint distributions; marginal distributions are a natural extension as we will see in a moment.

For discrete variables, we can use contingency tables to represent the joint frequency distribution for two random variables.

Joint Distribution - candidates

Example: two variables, gender (Y) and political party (X)
If we take the relative frequence of the cells we get:

Y/X	Party A (A)	Party B (B)	Party C (C)
Women (W)	$P(W \cap A)$	$P(W \cap B)$	$P(W \cap C)$
Man (M)	$P(M \cap A)$	$P(M \cap B)$	$P(M \cap C)$

Note: joint distributions are represented, guess what, by joint probabilities!

Joint Distribution

We can write the joint distributions as

$$P(Women \cap \text{Party B}) \text{ or } P(W \cup B)$$

or, if we have named the variables

$$P(Gender = Women, Party = \text{Party B})$$

$$P(X = W, Y = B)$$

These notations are equivalent if everything is well named.

Joint Distribution - dice

Joint distributions can be build from the process that generate the data (dice) or from a sample.

Example: roll a dice. Prime *vs* not prime (Y); and even *vs* odd (X). If we take the relative frequencies we get:

Y/X	Even (E)	Odd (O)
Prime (I)	$P(I \cap E) = 1/6$	$P(I \cap O) = 2/6$
Not Prime (N)	$P(N \cap E) = 2/6$	$P(N \cap O) = 1/6$

Joint Distribution - candidates

Example: (made-up) survey - 1200 women and 800 men. Gender (Y) and candidate of choice (X). If we take the relative frequencies we get:

Y/X	Dilma (D)	Aécio (A)
Women (W)	$P(W \cap D) = 700/2000$	$P(W \cap A) = 500/2000$
Men (M)	$P(M \cap D) = 400/2000$	$P(M \cap A) = 400/2000$

Joint Distribution - conditional probability

Let's use the same example to see how conditional probabilities can be represented in a table

Y/X	Dilma (D)	Aécio (A)
Women (W)	$P(W\|D) = 700/1100$	$P(W\|A) = 500/900$
Men (M)	$P(M\|E) = 400/1100$	$P(M\|A) = 400/900$
Total (M)	$P(W \cup M\|D) = 1$	$P(W \cup M\|A) = 1$

Joint Distribution - conditional probability

More on notation.

$$P(X = x_i|Y = 1) = P(X|Y = 1) =$$

$$= \frac{P(X = x, Y = 1)}{P(Y = 1)} = \frac{P(X \cap (Y = 1))}{P(Y = 1)}$$

Joint Distribution - conditionak probability

Let's use the same example to see how conditional probabilities can be represented in a table

Y/X	Dilma (D)	Aécio (A)	Total
(W)	$P(D\|W) = 700/1200$	$P(A\|W) = 500/1200$	$P(D \cup A\|W) = 1$
(M)	$P(D\|M) = 400/800$	$P(A\|M) = 400/800$	$P(D \cup A\|M) = 1$

Joint Distribution - Marginal Probabilities

The marginal probability of an event A is the probability that A will occur unconditional on all the other events on which A may depend. It is very easy to comprehend that in our example. If we take the relative frequencies we get:

Example:

Gender/Party	Party A (A)	Party B (B)	Party C (C)	Marginal
Women (W)	$P(W \cap A)$	$P(W \cap B)$	$P(W \cap C)$	P(W)
Man (M)	$P(M \cap A)$	$P(M \cap B)$	$P(M \cap C)$	P(M)
Marginal	$P(A)$	$P(B)$	$P(C)$	1

Joint Distribution - Marginal Probabilities

We can calculate the Marginal Probability by simplying summing the probability of A hapenning conditional on all other events on which A depend (partitions o B):

$$P(A) = P(A \cap B_1) + P(A \cap B_2) + ... + P(A \cap B_n) = \sum_{i=1}^{n} P(A \cap B_i)$$

or

$$P(A) = P(B_1) * P(A|B_1) + P(B_2) * P(A|B_2) + ... + P(B_n) * P(A|B_n) =$$

$$= \sum_{i=1}^{n} P(B_i) * P(A|B_i)$$

This means that one averages over other events and focuses on the one event, A, of interest.

Joint Distribution - Independence

What does happen with the joint distribution of two random variables if they are independent of each other?

Joint Distribution - Independence - Cards

Example: choose a card from a deck. Calculate the relative frequencies:

Y/X	Hearts	Spades	Clubs	Diamonds	Marginal
King	1/52	1/52	1/52	1/52	4/52
Queen	1/52	1/52	1/52	1/52	4/52
Other	11/52	11/52	11/52	11/52	44/52
Marginal	13/52	13/52	13/52	13/52	52/52

Joint Distribution - Independence - Cards

We got a king. What is the probability that it is the king of hearts? $P(H|K) = 1/4$

We got a queen. What is the probability that it is the queen of hearts? $P(H|Q) = 1/4$

We got any other card. What is the probability that it is a card of hearts? $P(H|O) = 1/4$

Joint Distribution - Independence - Cards

If the marginals probabilities are equal to the conditional probabilities, than the two variables are independent from each other.

$$P(H) = P(H|K) = P(H|Q) = P(H|O) = 1/4$$

Under independence:

$$P(H \cap K) = P(H) * P(K) = 1/4 * 1/13 = 1/52$$

or

$$P(Y = Hearts, X = King) = P(Y = Hearts) * P(X = King)$$