Joint, Marginal, and Conditional Distributions
An Example

Ricard Torres

ITAM

Microeconomía Aplicada II, 2nd Semester 2015

The US Presidential Elections of 2012

The following data about the US 2012 Presidential Elections is from wikipedia:

Total votes for Obama	65915796
Total votes for Romney and others	63170128
Total votes	129085924

Classification by genre

From survey data, it was estimated that:

- Out of the total number of Obama voters, 55% were women, and 45% were men.
- Analogously, out of the total number of voters for Romney and others, 45% were women, and 55% were men.

Joint Distribution in Absolute Terms

Therefore, taking into account the two previous pieces of information, we may represent the distribution of votes as:

		Genre	
		Female	Male
Candidate	Obama	36253688	29662108
	Romney	32161852	31008276

($*$) Note that, for simplicity, we associate with Romney all votes that were not for Obama.

Joint and Marginal Distributions in Absolute Terms

Adding up both by rows and by columns, we obtain the numbers for each of the two classification criteria separately:

		Genre		
		Female	Male	
Candidate	Obama	36253688	29662108	65915796
	Romney	32161852	31008276	63170128
		68415540	60670384	129085924

- We refer to the totals for the different classification criteria as the marginal distributions, whereas the results of simultaneously considering all classification criteria are the joint distribution.

A Little Bit of Notation

- Let us denote each candidate with its initial (O, R)
- Let us also denote genre by the initial (F, M)
- Since we are dealing with absolute numbers, let us denote the number of individuals that fall into a particular category by $\mathbb{N}(\cdot)$.
- The joint distribution of absolute numbers gives us the result of simultaneously applying the two criteria:
  - $\mathbb{N}(\text{O\& F}) = 36253688$, $\mathbb{N}(\text{R\& M}) = 31008276$, etc.
- The marginal distributions of absolute numbers are the result of applying a single criterion:
  - $\mathbb{N}(\text{O}) = 65915796$, $\mathbb{N}(\text{M}) = 60670384$, etc.

Joint and Marginal Distributions in Relative Terms

It is much easier to compare different vote distributions if we represent the numbers in relative, rather than absolute, terms (ie, as percentages of the grand total).

		Genre		
		Female	Male	
Candidate	Obama	.28	.23	.51
	Romney	.25	.24	.49
		.53	.47	1

- In this case, we have joint and marginal proportions, or probability distributions

More Notation

- Since now we are dealing with proportions (which we may view as probabilites), we will now use $\mathbb{P}(\cdot)$.
- The joint probability distribution gives us the result of simultaneously applying the two criteria:
    - $\mathbb{P}(\text{O\& F}) = .28$, $\mathbb{P}(\text{R\& M}) = .24$, etc.
- The marginal probability distributions are the result of applying a single criterion:
    - $\mathbb{P}(\text{O}) = .51$, $\mathbb{P}(\text{M}) = .47$, etc.

Conditional Distributions

- In many cases, we are interested in questions like: out of the total votes for Obama, what is the proportion of female voters?
- That is,

$$\mathbb{P}(\mathsf{F}|\mathsf{O}) = \frac{\mathbb{N}(\mathsf{F}\ \&\ \mathsf{O})}{\mathbb{N}(\mathsf{O})} = \frac{36253688}{65915796} = .55$$

- The above quotient is not altered if we divide both numerator and denominator by the grand total, that is, we may use proportions instead of absolute numbers:

$$\mathbb{P}(\mathsf{F}|\mathsf{O}) = \frac{\mathbb{P}(\mathsf{F}\ \&\ \mathsf{O})}{\mathbb{P}(\mathsf{O})} = \frac{.28}{.51} = .55$$

- These proportions that are obtained by fixing one of the classification criteria are named conditional distributions.

More on Conditional Distributions

- First of all, we must note that, once we fix a particular criterion, a conditional distribution is a genuine probability distribution:
  1. $\mathbb{P}(\text{F}|\text{O}) \geq 0$ and $\mathbb{P}(\text{M}|\text{O}) \geq 0$.
  2. $\mathbb{P}(\text{F}|\text{O}) + \mathbb{P}(\text{M}|\text{O}) = 1$.
- The same is true when we condition on R, F, or M.
- As we mention above, we may also condition on genre, rather than on candidate. For example, the probability of voting for Obama among females was:

$$\mathbb{P}(\text{O}|\text{F}) = \frac{\mathbb{P}(\text{F} \; \& \; \text{O})}{\mathbb{P}(\text{F})} = \frac{.28}{.53} = .53$$

Reversing Conditional Distributions

- The numerator is the same no matter which is the criterion we condition on, so there is a relationship between the two conditional probabilities:

$$\mathbb{P}(\mathsf{F}|\mathsf{O})\,\mathbb{P}(\mathsf{O}) = \mathbb{P}(\mathsf{F}\&\ \mathsf{O}) = \mathbb{P}(\mathsf{O}|\mathsf{F})\,\mathbb{P}(\mathsf{F})$$

- We may use this to express one conditional probability in terms of the other:

$$\mathbb{P}(\mathsf{O}|\mathsf{F}) = \frac{\mathbb{P}(\mathsf{F}|\mathsf{O})\,\mathbb{P}(\mathsf{O})}{\mathbb{P}(\mathsf{F})} = \frac{.55 \times .51}{.53} = .53$$

- This idea of expressing a conditional probability in terms of the "reverse" ones is often used, and is the basis of the so-called Bayes' Formula.

Joint Distributions Have More Information Than Marginals

- In general, there are many joint distributions that are compatible with the same marginals.
- For example, let us fix the marginal distributions in the previous example, and look at all joint distributions that are compatible.
- In order to do that, let $\mathbb{P}(\text{O}\&\ \text{F}) = a$, and note that this parameter determines the remaining values.

		Genre		
		F	M	
Candidate	O	a	$.51 - a$	.51
	R	$.53 - a$	$a - .04$	.49
		.53	.47	1

In order for all joint probabilities to be nonnegative, we need $0.04 \leq a \leq 0.51 \rightarrow$ but any such value will do

Two Joint Distributions with the Same Marginals

Let us consider the joint distributions generated in the two extreme cases: $a = 0.04$ and $a = 0.51$:

		Genre		
		F	M	
Candidate	O	.04	.47	.51
	R	.49	0	.49
		.53	.47	1

		Genre		
		F	M	
Candidate	O	.51	0	.51
	R	.02	.47	.49
		.53	.47	1

- Suppose the two criteria were ordered from smaller to larger values, e.g.: $\mathrm{O} = \mathrm{F} = 1$, $\mathrm{R} = \mathrm{M} = 2$.
- In this case, the random variables corresponding to the left joint distribution are negatively correlated: there is a much higher probability for low-high or high-low values, than for low-low or high-high values.
- Analogously, the random variables corresponding to the right joint distribution are positively correlated: there is a much higher probability for low-low or high-high values, than for low-high or high-low values.

Independence

- Another value of a that gives rise to a particular case of singular importance is when $a = .53 \times .51$, the product of the two marginals.
- This will always be feasible because, for example, the fact that $\mathrm{P(O)} + \mathrm{P(R)} = 1$ implies: $[\mathrm{P(F)} \times \mathrm{P(O)}] + [\mathrm{P(F)} \times \mathrm{P(R)}] = \mathrm{P(F)} \times [\mathrm{P(O)} + \mathrm{P(R)}] = \mathrm{P(F)}$.
- The previous reasoning implies that all other joint probabilities are the product of the corresponding marginals.

		Genre		
		F	M	
Candidate	O	.2703	.2397	.51
	R	.2597	.2303	.49
		.53	.47	1

In this case, marginal and conditional probabilities will coincide:

$$\mathrm{P(F|O)} = \frac{\mathrm{P(F\ \&\ O)}}{\mathrm{P(O)}} = \frac{\mathrm{P(F)\,P(O)}}{\mathrm{P(O)}} = \mathrm{P(F)}.$$

Joint and Conditional Distributions

- From the conditionals on a particular criterion (eg, genre) and the corresponding marginals, we can obtain the joint distributions.
- But conditionals on a particular criterion are compatible with different joint distributions (depending on the marginals).

Conditional probabilities and information

- An interpretation of conditional probabilities is as a way to update our estimates once we receive information about the phenomenon we are interested in.
- Suppose that we want to know the probability that a person will vote for Obama. If we know nothing about this person, then our estimate that he or she will vote for Obama is $\mathbb{P}(\mathsf{O}) = .51$.
- Suppose now that we observe that the voter is female, then we may refine our estimate: $\mathbb{P}(\mathsf{O}|\mathsf{F}) = .28/.53 = .53 > .51$.

Interpretation of Independence

- Intuitively, the concept of independence means that the two criteria of classification do not interfere with (are not informative with respect to) each other.
- Formally, we say that the two marginal distributions are independent if the conditional probabilities coincide with the respective marginals.
- This implies that the joint probabilities are the product of the marginals:

$$\mathbb{P}(\mathsf{O}|\mathsf{F}) = \mathbb{P}(\mathsf{O}) \;\rightarrow\; \frac{\mathbb{P}(\mathsf{O}\&\mathsf{F})}{\mathbb{P}(\mathsf{F})} = \mathbb{P}(\mathsf{O}) \;\rightarrow\; \mathbb{P}(\mathsf{O}\&\mathsf{F}) = \mathbb{P}(\mathsf{F})\,\mathbb{P}(\mathsf{O})$$

- Note that $\mathbb{P}(\mathsf{O}|\mathsf{F}) = \mathbb{P}(\mathsf{O})$ implies $\mathbb{P}(\mathsf{F}|\mathsf{O}) = \mathbb{P}(\mathsf{F})$.