Econometrics: Review of Basic Probability

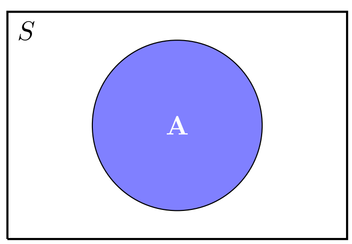
Ryan Safner

Fall 2018

1 Basic Probability Rules

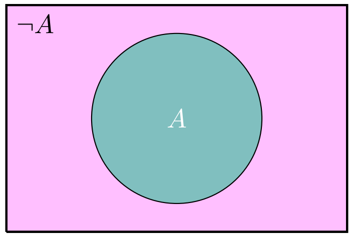
- Probability is the study of randomness
  - Random phenomena produce outcomes that are individually unknown, but we can describe overall, long-run tendencies
- Definitions
  - Event: a single outcome of a random phenomenon
  - Trial: a single attempt (as in an experiment) that produces an outcome
  - Sample Space (S or Ω): the set of all possible events
- Theoretical probability

$$P(\mathbf{A}) = \frac{\#\text{ of outcomes in }\mathbf{A}}{\text{Total }\#\text{ of outcomes}}$$



The sample space S and event A. $P(S) = 1$

- Probability Rules
  1. $0 \leq P \leq 1$
  2. $P(S) = 1$
  3. The probability of an event $\mathbf{A}$ *not* occurring, $P(\neg\mathbf{A}) = 1 - P(\mathbf{A})$ is the complement of $\mathbf{A}$
     - e.g. if the probability of picking a blue M&M is $P(Blue) = 0.47$, the probability of picking a *not* blue M&M is $P(\neg Blue) = 0.53$

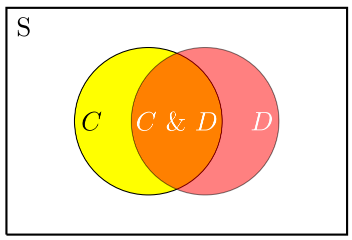


$S = A + \neg A$

The event **A** in teal and its complement ¬**A** in red

4. **Generalized Addition Rule**:

$$P(\mathbf{A} \text{ or } \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \text{ and } \mathbf{B})$$



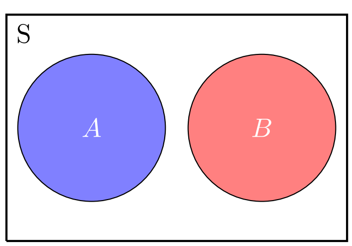
Non-disjoint events, e.g. The probability of picking a face card (**C**) OR a heart card (**D**) is: the probability of a heart plus the probability of a face minus the probability of cards that are both

- The symbol for "OR" is $\cup$, the union of two disjoint events. The symbol for "AND" is $\cap$, the intersection (overlap) of two events. So the generalized addition rule is:

$$P(\mathbf{A} \cup \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B}) - P(\mathbf{A} \cap \mathbf{B})$$

- If $P(\mathbf{A} \text{ and } \mathbf{B}) = 0$, then **A** and **B** are **disjoint**
  * Disjoint events cannot occur simultaneously, e.g. picking one M&M that is BOTH red AND blue
- If two events are disjoint–the ***simple*** **addition rule** (because $P(\mathbf{A} \text{ and } \mathbf{B}) = 0$):

$$P(\mathbf{A} \text{ or } \mathbf{B}) = P(\mathbf{A}) + P(\mathbf{B})$$



Disjoint events **A** and **B**

5. **Conditional Probability:**

$$P(\mathbf{B}|\mathbf{A}) = \frac{P(\mathbf{A} \text{ and } \mathbf{B})}{P(\mathbf{A})}$$

   - "The probability of event **B** *given* (|) event **A**"
   - e.g. the probability of someone watching the Super Bowl, *given* that they are male

6. **Generalized Multiplication Rule**:

$$P(\mathbf{C} \text{ and } \mathbf{D}) = P(\mathbf{C}) * P(\mathbf{D}|C)$$

   - If $P(\mathbf{D}|\mathbf{C}) = P(\mathbf{D})$, then **C** and **D** are **independent** (**C**'s occurrence does not change P(**D**))
     * Independent events: if one event occurring gives *no* information about the probability of another event
   - If **C** and **D** are independent, the ***Simple*** **Multiplication Rule**:

$$P(\mathbf{C} \text{ and } \mathbf{D}) = P(\mathbf{C}) * P(\mathbf{D})$$

   - Independent $\neq$ Disjoint! Disjoint events *cannot* be independent!
     * If events **A** and **B** are disjoint ($P(\mathbf{A} \cap \mathbf{B}) = 0$), this implies that if **A** occurs, **B** cannot possibly occur (mutually exclusive!) so $P(\mathbf{B}|\mathbf{A}) = 0 \neq P(\mathbf{B})$
     * e.g. If you get an A in this course, that means the probability of you getting a B, C, D, or F= 0!

2 Contingency Tables, Joint & Marginal Probability

- Contingency tables display the joint and marginal distributions of two variables

	# of Bedrooms		
Price	1	2	**Total**
Low	0.30	0.20	**0.50**
High	0.10	0.40	**0.50**
Total	**0.40**	**0.60**	**1.00**

  - Each cell is a disjoint union of events with a **joint probability**
    * e.g. $P(\text{Low Price} \cap 1 \text{ Bedroom}) = 0.30$, given by the table
  - **Marginal probabilities** are the probability of each category occurring overall, in margin of table
    * e.g. P(1 Bedroom)=0.40; P(Low)=0.50
  - **Conditional distribution** (e.g. of price) can be calculated with conditional probabilities for one condition (e.g. for an apartment having 2 Bedrooms)

$$P(\text{Low Price}|\text{2 Bedrooms}) = \frac{P(\text{Low Price} \cup 2 \text{ Bedrooms})}{P(\text{2 Bedrooms})} = \frac{0.20}{0.60} = 0.33$$

$$P(\text{High Price}|\text{2 Bedrooms}) = \frac{P(\text{High Price} \cup 2 \text{ Bedrooms})}{P(\text{2 Bedrooms})} = \frac{0.40}{0.60} = 0.67$$

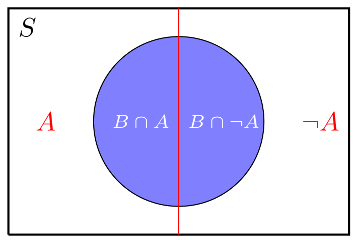
3 Bayes' Rule

- We know $P(B|A)$ but may want to find $P(A|B)$, they are not the same!
- **Bayes' Rule**:

$$P(A|B) = \frac{P(B|A)P(A)}{P(B)}$$

- We often don't know $P(B)$ (the denominator). We use the **law of total probability**, if we know B can occur simultaneously with either A or $\neg A$:

$$P(B) = P(B|A)P(A) + P(B|\neg A)P(\neg A)$$



- Example: Suppose 1% of the population has a rare disease. A test that can diagnose the disease is 95% accurate. What is the probability that a person who takes the test and comes back positive has the disease? [Stop and try to think through this before proceeding. Your intuitions will fail you!]

$$P(Disease|Positive) = \frac{P(Positive|Disease)P(Disease)}{P(Positive)}$$

  - We know $P(Positive|Disease) = 0.95$, $P(Disease) = 0.01$, what is $P(Positive)$?
  - Find using law of total probability:

$$P(Positive) = P(Positive|Disease)P(Disease) + P(Positive|NoDisease)P(NoDisease)$$

$$P(Positive) = 0.95 * 0.01 + 0.05 * 0.99 = 0.0095 + 0.0495 = 0.0590$$

  - So finally:

$$P(Disease|Positive) = \frac{0.95 * 0.01}{0.059} = 0.16 \implies 16\%!$$

  - The magic of Bayes' rule is that everyone forgets the base rate (the disease itself is so rare).

- In Bayesian updating:
  - $P(A)$ is the "base rate" or "prior"
  - $P(A|B)$ is the "posterior," having accounted for **B**
  - $P(B|A)$ is the "likelihood" of A and B being compatible
  - $P(A)$ is the "marginal likelihood" of (all possible events of) A irrespective of B

- Most important to remember $P(A|B) \neq P(B|A)$!