Today's topics

- Probability
  - Definitions
  - Events
  - Conditional probability
- Reading: Sections 5.1-5.3
- Upcoming
  - Expected value

Why Probability?

- In the real world, we often don't know whether a given proposition is true or false.
- Probability theory gives us a way to reason about propositions whose truth is *uncertain*.
- It is useful in weighing evidence, diagnosing problems, and analyzing situations whose exact details are unknown.

Random Variables

- A *"random variable"* V is any variable whose value is unknown, or whose value depends on the precise situation.
  - *E.g.*, the number of students in class today
  - Whether it will rain tonight (Boolean variable)
- Let the domain of V be **dom**$[V] \equiv \{v_1,\ldots,v_n\}$
  - Infinite domains can also be dealt with if needed.
- The proposition $V=v_i$ may have an uncertain truth value, and may be assigned a *probability*.

Information Capacity

- The *information capacity* **I**$[V]$ of a random variable V with a finite domain can be defined as the logarithm (with indeterminate base) of the size of the domain of V, **I**$[V] :\equiv \log |$**dom**$[V]|$.
  - The log's base determines the associated information unit!
    - Taking the log base 2 yields an information unit of 1 *bit* b = log 2.
      - Related units include the *nybble* N = 4 b = log 16 (1 hexadecimal digit),
      - and more famously, the *byte* B = 8 b = log 256.
    - Other common logarithmic units that can be used as units of information:
      - the *nat*, or *e-fold* n = log e,
        - » widely known in thermodynamics as *Boltzmann's constant k*.
      - the *bel* or *decade* or *order of magnitude* (D = log 10),
      - and the *decibel* or dB = D/10 = (log 10)/10 ≈ log 1.2589
- **Example:** An 8-bit register has $2^8 = 256$ possible values.
  - Its information capacity is thus: log 256 = 8 log 2 = 8 b!
    - Or 2N, or 1B, or $\log_e 256 \approx 5.545$ n, or $\log_{10} 256 = 2.408$ D, or 24.08 dB

Experiments & Sample Spaces

- A (stochastic) *experiment* is any process by which a given random variable V gets assigned some *particular* value, and where this value is not necessarily known in advance.
  - We call it the "actual" value of the variable, as determined by that particular experiment.
- The *sample space* S of the experiment is just the domain of the random variable, $S = \text{dom}[V]$.
- The *outcome* of the experiment is the specific value v_i of the random variable that is selected.

Events

- An *event* E is any set of possible outcomes in S…
  - That is, $E \subseteq S = \textbf{dom}[V]$.
    - E.g., the event that "less than 50 people show up for our next class" is represented as the set $\{1, 2, \ldots, 49\}$ of values of the variable V = (# of people here next class).
- We say that event E *occurs* when the actual value of V is in E, which may be written $V \in E$.
  - Note that $V \in E$ denotes the proposition (of uncertain truth) asserting that the actual outcome (value of V) will be one of the outcomes in the set E.

Probability

- The *probability* $p = \Pr[E] \in [0,1]$ of an event E is a real number representing our degree of certainty that E will occur.
  - If $\Pr[E] = 1$, then E is absolutely certain to occur,
    - thus $V \in E$ has the truth value **True**.
  - If $\Pr[E] = 0$, then E is absolutely certain *not* to occur,
    - thus $V \in E$ has the truth value **False**.
  - If $\Pr[E] = ½$, then we are *maximally uncertain* about whether E will occur; that is,
    - $V \in E$ and $V \notin E$ are considered *equally likely*.
  - How do we interpret other values of p?

Note: We could also define probabilities for more general propositions, as well as events.

Four Definitions of Probability

- Several alternative definitions of probability are commonly encountered:
  - Frequentist, Bayesian, Laplacian, Axiomatic
- They have different strengths & weaknesses, philosophically speaking.
  - But fortunately, they coincide with each other and work well together, in the majority of cases that are typically encountered.

Probability: Frequentist Definition

- The probability of an event E is the limit, as $n\to\infty$, of the fraction of times that we find $V\in E$ over the course of n independent repetitions of (different instances of) the same experiment.

$$\Pr[E] :\equiv \lim_{n\to\infty} \frac{n_{V\in E}}{n}$$

- Some problems with this definition:
  - It is only well-defined for experiments that can be independently repeated, infinitely many times!
    - or at least, if the experiment can be repeated in principle, *e.g.*, over some hypothetical ensemble of (say) alternate universes.
  - It can never be measured exactly in finite time!
- **Advantage:** It's an objective, mathematical definition.

Probability: Bayesian Definition

- Suppose a rational, profit-maximizing entity R is offered a choice between two rewards:
  - Winning $1 if and only if the event E actually occurs.
  - Receiving p dollars (where $p\in[0,1]$) unconditionally.
- If R can honestly state that he is completely indifferent between these two rewards, then we say that R's probability for E is p, that is, $\Pr_R[E] :\equiv p$.
- **Problem:** It's a subjective definition; depends on the reasoner R, and his knowledge, beliefs, & rationality.
  - The version above additionally assumes that the utility of money is linear.
    - This assumption can be avoided by using "utils" (utility units) instead of dollars.

Probability: Laplacian Definition

- First, assume that all individual outcomes in the sample space are *equally likely* to each other...
  - Note that this term still needs an operational definition!
- Then, the probability of any event E is given by, $\Pr[E] = |E|/|S|$. Very simple!
- **Problems:** Still needs a definition for *equally likely*, and depends on the existence of *some* finite sample space S in which all outcomes in S are, in fact, equally likely.

Probability: Axiomatic Definition

- Let p be any total function $p:S\to[0,1]$ such that $\sum_s p(s) = 1$.
- Such a p is called a *probability distribution*.
- Then, the *probability under p* of any event $E\subseteq S$ is just:

$$\Pr_p[E] :\equiv \sum_{s\in E} p(s)$$

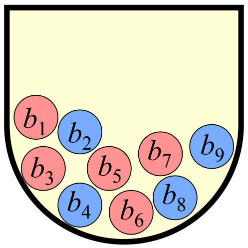
- **Advantage:** Totally mathematically well-defined!
  - This definition can even be extended to apply to infinite sample spaces, by changing $\sum\to\int$, and calling p a *probability density function* or a probability *measure*.
- **Problem:** Leaves operational meaning unspecified.

Probabilities of Mutually Complementary Events

- Let E be an event in a sample space S.
- Then, $\overline{E}$ represents the *complementary* event, saying that the actual value of $V \notin E$.
- **Theorem**: $\Pr[\overline{E}] = 1 - \Pr[E]$
  - This can be proved using the Laplacian definition of probability, since $\Pr[\overline{E}] = |\overline{E}|/|S| = (|S|-|E|)/|S| = 1 - |E|/|S| = 1 - \Pr[E]$.
    - Other definitions can also be used to prove it.

Example 1: Balls-and-Urn

- Suppose an urn contains 4 blue balls and 5 red balls.
- An example **experiment:** Shake up the urn, reach in (without looking) and pull out a ball.
- A **random variable** V: Identity of the chosen ball.
- The **sample space** S: The set of all possible values of V:
  - In this case, $S = \{b_1,\ldots,b_9\}$
- An **event** E: "The ball chosen is blue": $E = \{$ ________________ $\}$
- What are the odds in favor of E?
- What is the probability of E? (Use Laplacian def'n.)



Example 2: Seven on Two Dice

- **Experiment:** Roll a pair of fair (unweighted) 6-sided dice.

- Describe a sample space for this experiment that fits the Laplacian definition.
- Using this sample space, represent an event E expressing that "the upper spots sum to 7."
- What is the probability of E?

Probability of Unions of Events

- Let $E_1,E_2 \subseteq S = \textbf{dom}[V]$.
- Then we have: **Theorem:** $\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2]$
  - By the inclusion-exclusion principle, together with the Laplacian definition of probability.
- You should be able to easily flesh out the proof yourself at home.

Mutually Exclusive Events

- Two events E_1, E_2 are called *mutually exclusive* if they are disjoint: $E_1 \cap E_2 = \varnothing$
  - Note that two mutually exclusive events *cannot both occur* in the same instance of a given experiment.
- For mutually exclusive events,
  $\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2]$.
  - Follows from the sum rule of combinatorics.

Exhaustive Sets of Events

- A set $\boldsymbol{E} = \{E_1, E_2, \ldots\}$ of events in the sample space S is called *exhaustive* iff $\bigcup E_i = S$.
- An exhaustive set $\boldsymbol{E}$ of events that are all mutually exclusive with each other has the property that $\sum \Pr[E_i] = 1.$
- You should be able to easily prove this theorem, using either the Laplacian or Axiomatic definitions of probability from earlier.

Independent Events

- Two events E,F are called *independent* if
  $\Pr[E \cap F] = \Pr[E] \cdot \Pr[F]$.
- Relates to the product rule for the number of ways of doing two independent tasks.
- **Example:** Flip a coin, and roll a die.
  Pr[(coin shows heads) ∩ (die shows 1)] =
  Pr[coin is heads] × Pr[die is 1] = ½×1/6 =1/12.

Conditional Probability

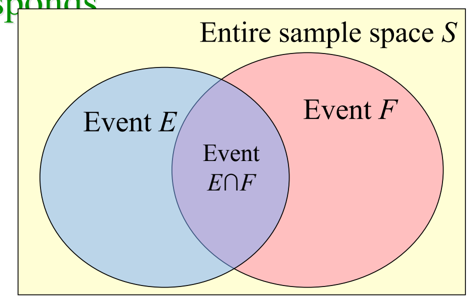
- Let E,F be any events such that $\Pr[F]>0$.
- Then, the *conditional probability of E given F*, written $\Pr[E|F]$, is defined as
  $\Pr[E|F] :\equiv \Pr[E \cap F]/\Pr[F]$.
- This is what our probability that E would turn out to occur should be, if we are given *only* the information that F occurs.
- If E and F are independent then $\Pr[E|F] = \Pr[E]$.
  $\because \Pr[E|F] = \Pr[E \cap F]/\Pr[F] = \Pr[E]\times\Pr[F]/\Pr[F] = \Pr[E]$

Prior and Posterior Probability

- Suppose that, before you are given any information about the outcome of an experiment, your personal probability for an event E to occur is $p(E) = \Pr[E]$.
  - The probability of E in your original probability distribution p is called the *prior* probability of E.
    - This is its probability *prior* to obtaining any information about the outcome.
- Now, suppose someone tells you that some event F (which may overlap with E) actually occurred in the experiment.
  - Then, you should *update* your personal probability for event E to occur, to become $p'(E) = \Pr[E|F] = p(E \cap F)/p(F)$.
    - The conditional probability of E, *given* F.
  - The probability of E in your *new* probability distribution p' is called the *posterior* probability of E.
    - This is its probability *after* learning that event F occurred.
- After seeing F, the posterior distribution p' is defined by letting $p'(v) = p(\{v\} \cap F)/p(F)$ for each individual outcome $v \in S$.

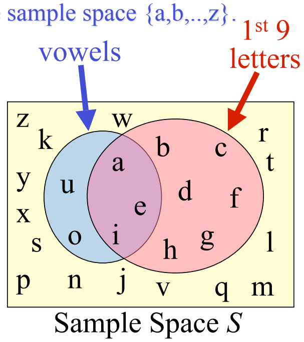
Visualizing Conditional Probability

- If we are given that event F occurs, then
  - Our attention gets restricted to the subspace F.
- Our *posterior* probability for E (after seeing F) corresponds to the *fraction* of F where E occurs also.
- Thus, $p'(E) = p(E \cap F)/p(F)$.



Conditional Probability Example

- Suppose I choose a single letter out of the 26-letter English alphabet, totally at random.
  - Use the Laplacian assumption on the sample space {a,b,..,z}.
  - What is the (prior) probability that the letter is a vowel?
    - Pr[Vowel] = __ / __ .
- Now, suppose I tell you that the letter chosen happened to be in the first 9 letters of the alphabet.
  - Now, what is the conditional (or posterior) probability that the letter is a vowel, given this information?
    - Pr[Vowel | First9] = ___ / ___ .

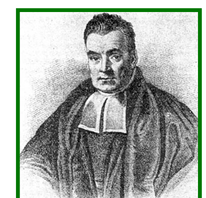


Bayes' Rule

- One way to compute the probability that a hypothesis H is correct, given some data D:

$$\Pr[H \mid D] = \frac{\Pr[D \mid H] \cdot \Pr[H]}{\Pr[D]}$$

- This follows directly from the definition of conditional probability! (**Exercise:** Prove it)
- This rule is the foundation of *Bayesian methods* for probabilistic reasoning, which are very powerful, and widely used in artificial intelligence applications:
  - For data mining, automated diagnosis, pattern recognition, statistical modeling, even evaluating scientific hypotheses!


Rev. Thomas Bayes 1702-1761