Module #16:
Probability Theory

Rosen 5th ed., ch. 5

23 slides, ~1 lecture

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*.
- Useful in weighing evidence, diagnosing problems, and analyzing situations whose exact details are unknown.

Random Variables

- A *random variable V* is a variable whose value is unknown, or that depends on the 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]=$\{v_1,\ldots,v_n\}$
- The proposition $V=v_i$ may be uncertain, and be assigned a *probability*.

Amount of Information

- The *amount of information* $\mathbf{I}[V]$ in a random variable V is the logarithm of the size of the domain of V, $\mathbf{I}[V]=\log|\mathbf{dom}[V]|$.
  - The base of the logarithm determines the information unit; base 2 gives a unit of 1 bit.
- Example: An 8-bit register has $2^8 = 256$ possible values. Log 256 = 8 bits.

Experiments

- A (stochastic) *experiment* is a process by which a given random variable gets assigned a specific value.
- The *sample space* S of the experiment is the domain of the random variable.
- The *outcome* of the experiment is the specific value of the random variable that is selected.

Events

- An *event* E is a set of possible outcomes
  - That is, $E \subseteq S = \mathbf{dom}[V]$.
- We say that event E *occurs* when $V \in E$.
- Note that $V \in E$ is the (uncertain) proposition that the actual outcome 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 $V \in E$ is true.
  - If $\Pr[E] = 0$, then E is absolutely certain *not* to occur,
    - thus $V \in E$ is false.
  - If $\Pr[E] = ½$, then we are *completely uncertain* about whether E will occur; that is,
    - $V \in E$ and $V \notin E$ are considered *equally likely*.
  - What about other cases?

Four Definitions of Probability

- Several alternative definitions of probability are commonly encountered:
  - Frequentist, Bayesian, Laplacian, Axiomatic
- They have different strengths & weaknesses.
- Fortunately, they coincide and work well with each other in most cases.

Probability: Frequentist Definition

- The probability of an event E is the limit, as $n\rightarrow\infty$, of the fraction of times that $V\in E$ in n repetitions of the same experiment.
- Problems:
  - Only well-defined for experiments that are infinitely repeatable (at least in principle).
  - Can never be measured exactly in finite time!
- Advantage: Objective, mathematical def'n.

Probability: Bayesian Definition

- Suppose a rational entity *R* is offered a choice between two rewards:
  - Winning $1 if event *E* occurs.
  - Receiving *p* dollars (where $p \in [0,1]$) unconditionally.
- If *R* is indifferent between these two rewards, then we say *R*'s probability for *E* is *p*.
- Problem: Subjective definition, depends on the reasoner *R*, and his knowledge & rationality.

Probability: Laplacian Definition

- First, assume that all outcomes in the sample space are *equally likely*
  - This term still needs to be defined.
- Then, the probability of event E, $\Pr[E] = |E|/|S|$. Very simple!
- Problems: Still needs a definition for *equally likely*, and depends on existence of a finite sample space with all equally likely outcomes.

Probability: Axiomatic Definition

- Let p be any function $p{:}S\rightarrow[0,1]$, such that:
  - $0 \leq p(s) \leq 1$ for all outcomes $s\in S$.
  - $\sum p(s) = 1$.
- Such a p is called a *probability distribution*.
- Then, the probability of any event $E\subseteq S$ is just:

$$\Pr[E] = \sum_{s\in E} p(s)$$

- Advantage: Totally mathematically well-defined.
- Problem: Leaves operational def'n unspecified.

Probability of Complementary Events

- Let E be an event in a sample space S.
- Then, $\overline{E}$ represents the *complementary* event that $V \notin E$.
- $\Pr[\overline{E}] = 1 - \Pr[E]$

Probability of Unions of Events

- Let $E_1, E_2 \subseteq S = \mathbf{dom}[V]$.
- Then:
  $\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2] - \Pr[E_1 \cap E_2]$
  - By the inclusion-exclusion principle.

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,

$$\Pr[E_1 \cup E_2] = \Pr[E_1] + \Pr[E_2].$$

Exhaustive Sets of Events

- A set $\boldsymbol{E} = \{E_1, E_2, \ldots\}$ of events in the sample space S is *exhaustive* if $\bigcup E_i = S$.
- An exhaustive set of events that are all mutually exclusive with each other has the property that

$$\sum \Pr[E_i] = 1$$

Independent Events

- Two events E,F are *independent* if $\Pr[E\cap F] = \Pr[E]\cdot\Pr[F]$.
- Relates to product rule for number of ways of doing two independent tasks
- Example: Flip a coin, and roll a die.

  Pr[quarter is heads $\cap$ die is 1] =

  Pr[quarter is heads] $\times$ Pr[die is 1]

Conditional Probability

- Let E,F be events such that $\Pr[F]>0$.
- Then, the *conditional probability of E given F*, written $\Pr[E|F]$, is defined as $\Pr[E\cap F]/\Pr[F]$.
- This is the probability that E would turn out to be true, given just the information that F is true.
- If E and F are independent, $\Pr[E|F] = \Pr[E]$.

Bayes's Theorem

- Allows one to compute the probability that a hypothesis H is correct, given data D:

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

- Easy to prove from def'n of conditional prob.
- Extremely useful in artificial intelligence apps:
  - Data mining, automated diagnosis, pattern recognition, statistical modeling, evaluating scientific hypotheses.

Expectation Values

- For a random variable V having a numeric domain, its *expectation value* or *expected value* or *weighted average value* or *arithmetic mean value* $\mathbf{Ex}[V]$ is defined as
$$\sum_{v\in\mathbf{dom}[V]} v\cdot p(v).$$
- The term "expected value" is widely used, but misleading since the expected value might be totally unexpected or impossible!

Derived Random Variables

- Let S be a sample space over values of a random variable V (representing possible outcomes).
- Then, any function f over S can also be considered to be a random variable (whose value is derived from the value of V).
- If the range R = **range**[f] of f is numeric, then **Ex**[f] can still be defined, as $\sum_{s\in S} p(s)\cdot f(s)$

Linearity of Expectation

- Let X_1, X_2 be any two random variables derived from the same sample space. Then:
- $\mathbf{Ex}[X_1+X_2] = \mathbf{Ex}[X_1] + \mathbf{Ex}[X_2]$
- $\mathbf{Ex}[aX_1 + b] = a\mathbf{Ex}[X_1] + b$

Variance

- The *variance* $\mathbf{Var}[X] = \sigma^2(X)$ of a random variable X is the expected value of the square of the difference between the value of X and its expectation value $\mathbf{Ex}[X]$:

$$\mathbf{Var}[X] :\equiv \sum_{s \in S} (X(s) - \mathbf{Ex}[X])^2 \, p(s)$$

- The *standard deviation* or *root-mean-square* (RMS) *difference* of X, $\sigma(X) :\equiv \mathbf{Var}[X]^{1/2}$.

Entropy

- The *entropy* H of a probability distribution p over a sample space S over outcomes is a measure of our degree of uncertainty about the outcome.
  - It measures the expected amount of increase in known information from learning the actual outcome.

$$H = \sum_{s \in S} p(s) \log(1/p(s))$$

- The base of the logarithm gives the unit of entropy; base 2 → 1 bit, base e → 1 nat
  - 1 nat is also known as "Boltzmann's constant" k_B & as the "ideal gas constant" R, first discovered physically