Today’s topics

• Probability
  – Expected value

• Reading: Sections 5.3

• Upcoming
  – Probabilistic inference
Expectation Values

• For any random variable \( V \) having a numeric domain, its expectation value or expected value or weighted average value or (arithmetic) mean value \( \text{Ex}[V] \), under the probability distribution \( \Pr[v] = p(v) \), is defined as

\[
\hat{V} \ni \text{Ex}[V] \equiv \text{Ex}_{p}[V] \equiv \sum_{v \in \text{dom}[V]} v \cdot p(v).
\]

• The term “expected value” is very widely used for this.
  – But this term is somewhat misleading, since the “expected” value might itself be totally unexpected, or even impossible!

  • E.g., if \( p(0)=0.5 \) & \( p(2)=0.5 \), then \( \text{Ex}[V]=1 \), even though \( p(1)=0 \) and so we know that \( V \neq 1 \)!
  • Or, if \( p(0)=0.5 \) & \( p(1)=0.5 \), then \( \text{Ex}[V]=0.5 \) even if \( V \) is an integer variable!
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 actual value $f(V)$ is derived from the actual value of $V$).
- If the range $R = \text{range}[f]$ of $f$ is numeric, then the mean value $\text{Ex}[f]$ of $f$ can still be defined, as
  \[
  \hat{f} = \text{Ex}[f] = \sum_{s \in S} p(s) \cdot f(s)
  \]
Linearity of Expectation Values

• Let $X_1, X_2$ be any two random variables derived from the same sample space $S$, and subject to the same underlying distribution.

• Then we have the following theorems:
  \[
  \text{Ex}[X_1 + X_2] = \text{Ex}[X_1] + \text{Ex}[X_2]
  \]
  \[
  \text{Ex}[aX_1 + b] = a\text{Ex}[X_1] + b
  \]

• You should be able to easily prove these for yourself at home.
Variance & Standard Deviation

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

$$\text{Var}[X] := \sum_{s \in S} (X(s) - \text{Ex}_p[X])^2 p(s)$$

• The standard deviation or root-mean-square (RMS) difference of $X$ is $\sigma(X) := \text{Var}[X]^{1/2}$. 
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 actual outcome.

- It measures the expected amount of increase in our known information that would result from learning the outcome.

$$H(p) \equiv \mathbb{E}_p[\log p^{-1}] = -\sum_{s \in S} p(s) \log p(s)$$

- The base of the logarithm gives the corresponding unit of entropy; base 2 $\rightarrow$ 1 bit, base $e$ $\rightarrow$ 1 nat (as before)

- 1 nat is also known as “Boltzmann’s constant” $k_B$ & as the “ideal gas constant” $R$, and was first discovered physically.
Visualizing Entropy

Sample Nonuniform vs. Uniform Probability Distributions

Improbability (Inverse Probability)

Log Improbability (Information of Discovery)

Boltzmann-Gibbs-Shannon Entropy (Expected Log Improbability)