Week 2

1 Independent random variables

Definition: Two random variables X_1, X_2 are said to be independent if,

$$P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)P(X_2 = x_2) \quad \forall\, x_1 \in \mathcal{X}_1 \,\&\, x_2 \in \mathcal{X}_2$$

$\boxed{P(X_1 = x_1, X_2 = x_2) \text{ means that } X_1 = x_1 \textbf{ AND } X_2 = x_2 \text{ occur.}}$

1.1 Claim

$$\sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) = P(X_1 = x_1)$$

Proof: Suppose A & B are disjoint/mutually exclusive. Then,

$$P(A \cap B) = 0 \quad \& \quad P(A \cup B) = P(A) + P(B)$$

Now the events $(X_1 = x_1) \cap (X_2 = x_2)$ are disjoint for different values of $x_2 \in \mathcal{X}_2$.(if $x_2 \neq x_2' \;\forall\, x_2 \in \mathcal{X}_2$)
Thus,

$$\begin{aligned}
\sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) &= \sum_{x_2 \in \mathcal{X}_2} P\big((X_1 = x_1) \cap (X_2 = x_2)\big) \\
&= P\left(\bigcup_{x_2 \in \mathcal{X}_2} (X_1 = x_1) \cap (X_2 = x_2)\right)
\end{aligned}$$

Now,

$$\begin{aligned}
\bigcup_{x_2 \in \mathcal{X}_2} [(X_1 = x_1) \cap (X_2 = x_2)] &= (X_1 = x_1) \bigcap \left[\bigcup_{x_2 \in \mathcal{X}_2} (X_2 = x_2)\right] \\
&= (X_1 = x_1) \bigcap [x_2 \in \mathcal{X}_2]
\end{aligned}$$

As $[x_2 \in \mathcal{X}_2]$ forms the entire sample space,

$$\begin{aligned}
P\left(\bigcup_{x_2 \in \mathcal{X}_2} (X_1 = x_1) \cap (X_2 = x_2)\right) &= P\Big((X_1 = x_1) \bigcap [x_2 \in \mathcal{X}_2]\Big) \\
&= P\Big((X_1 = x_1) \bigcap \Omega\Big) \\
&= P(X_1 = x_1)
\end{aligned}$$

2 Lemma

Suppose $X_1 \in \mathcal{X}_1$ & $X_2 \in \mathcal{X}_2$ are **independant** random variables. Then,

$$H(X_1,\ X_2) = H(X_1) + H(X_2)$$

Proof:

$$\sum_{\substack{x_1 \in \mathcal{X}_1 \\ x_2 \in \mathcal{X}_2}} P(X_1 = x_1, X_2 = x_2) \log\left(\frac{1}{P(X_1 = x_1, X_2 = x_2)}\right)$$

$$= \sum_{x_1 \in \mathcal{X}_1} \sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2) \left[\log \frac{1}{P(X_1 = x_1)} + \log \frac{1}{P(X_2 = x_2)}\right]$$

$$= \sum_{x_1 \in \mathcal{X}_1} \log \frac{1}{P(X_1 = x_1)} \left(\sum_{x_2 \in \mathcal{X}_2} P(X_1 = x_1, X_2 = x_2)\right)$$

$$+ \sum_{x_2 \in \mathcal{X}_2} \log \frac{1}{P(X_2 = x_2)} \left(\sum_{x_1 \in \mathcal{X}_1} P(X_1 = x_1, X_2 = x_2)\right)$$

From the previous claim,

$$\begin{aligned} H(X_1, X_2) = &\sum_{x_1 \in \mathcal{X}_1} P(X_1 = x_1) \log \frac{1}{P(X_1 = x_1)} \\ &+ \sum_{x_2 \in \mathcal{X}_2} P(X_2 = x_2) \log \frac{1}{P(X_2 = x_2)} \\ &= H(X_1) + H(X_2) \end{aligned}$$

2.1 Conditional probability distribution

What if X_1 & X_2 are not independent? Then we would use conditional probability distribution. i.e.

$$P(X_2 = x_2 / X_1 = x_1) := \frac{P(X_2 = x_2, X_1 = x_1)}{P(X_1 = x_1)}, \quad P(X_1 = x_1) \neq 0$$

This definition for conditional probability satisfies the probability axioms and hence it is a valid probability measure.

2.2 Conditional entropy

Definition:

$$H(X_2/X_1) := \sum_{x_1 \in \mathcal{X}_1} P(X_2 = x_2) H(X_2/X_1 = x_1)$$

where,

$$H(X_2/X_1 = x_1) := \sum_{x_2 \in \mathcal{X}_2} P(X_2 = x_2/X_1 = x_1) \log \frac{1}{P(X_2 = x_2/X_1 = x_1)}$$

2.2.1 Support of a function

When $P(X = x) = 0$, it is not considered in calculating the entropy.

$$H(X) = - \sum_{\{x \in \mathcal{X} : P(X=x) \neq 0\}} P(X = x) log(P(X = x))$$

Suppose that f : X → R is a real-valued function whose domain is an arbitrary set X. The set-theoretic support of f, written supp(f), is the set of points in X where f is non-zero:

$$supp(X) = \{x \in X : f(x) \neq 0\}$$

Note: $P(X = x)$ can be denoted as $P_X(x)$ or $P(x)$. Hence:

$$H(X) = - \sum_{x \in supp(P)} P(x) log(P(x))$$

2.3 Chain Rule

$$\begin{aligned} H(X_1, X_2) &= H(X_1) + H(X_2/X_1) \\ &= H(X_2) + H(H_1/X_2) \end{aligned}$$

Proof: