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,

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

$$= P\left(\bigcup_{x_2 \in \mathcal{X}_2} (X_1 = x_1) \cap (X_2 = x_2) \right)$$

Now,

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

$$= (X_1 = x_1) \bigcap [x_2 \in \mathcal{X}_2]$$

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

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

$$= P\left((X_1 = x_1) \bigcap \Omega \right)$$

$$= P(X_1 = x_1)$$

2 Entropy

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,

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

$$+ \sum_{x_2 \in \mathcal{X}_2} P(X_2 = x_2) \log \frac{1}{P(X_2 = x_2)}$$

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

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

2.2.1 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 $P(x)$.

2.3 Important Results

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

$$H(X/Y) = \sum_{y \in supp(P_Y)} P_Y(y) H(X/Y = y)$$

$$H(X/Y = y) = - \sum_{x \in supp(P_{X/Y})} P_{X/Y}(x/y) log(P_{X/Y}(x/y))$$

2.4 Chain Rule

$$H(X, Y) = H(X) + H(Y/X) = H(Y) + H(X/Y)$$

Proof:

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

$$+ \sum_{x \in supp(P_X)} P(x) \left(\sum_{y \in supp(P_Y)} P(y/x) log \left(\frac{1}{P(y/x)} \right) \right)$$

$$= - \sum_{x \in supp(P_X)} \left(\sum_{y \in supp(P_Y)} P(x, y) \right) log(P(x))$$

$$+ \sum_{x \in supp(P_X)} \left(\sum_{y \in supp(P_Y)} P(x, y) \right) log \left(\frac{1}{P(y/x)} \right)$$

$$= \sum_{x \in supp(P_X)} \sum_{y \in supp(P_Y)} P(x, y) \left(log \frac{1}{P(x)P(y/x)} \right)$$

$$= - \sum_x \sum_y P(x, y) log P(x, y) = H(x, y)$$

3

2.5 Properties of entropy

Upper and lower bound of entropy:
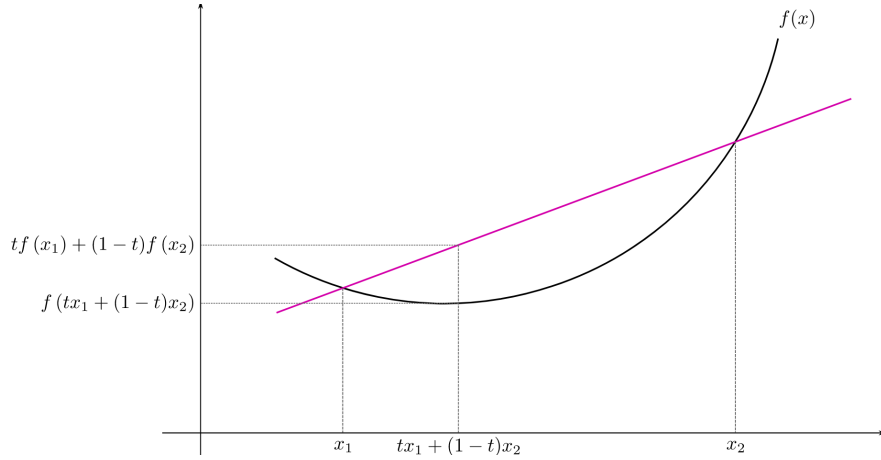
$$0 \leq H(X) \leq |log\mathcal{X}|$$

Proof:

$H(X) = \sum_x P(x) log \frac{1}{P(x)} \rightarrow$ always positive as $P(x) > 0 \ \forall x \in supp(P_X)$.

$H(X) = 0$ when $|supp(P_X)| = 1$.

2.5.1 Jensen's inequality (Concave/convex functions)

$x_1 < x_3 < x_2$ such that $x_3 = tx_1 + (1-t)x_2$

If $f(x_3) \geq tf(x_1) + (1-t)f(x_2) \rightarrow$ Concave function.

If $f(tx_1 + (1-t)x_2) \leq tf(x_1) + (1-t)f(x_2) \rightarrow$ Convex function.

Hence, the above graph is a convex function.

We shall use the concavity property of the log function to obtain the proof for $H(X) \leq |log\mathcal{X}|$.

Proof:

$$H(X) = \sum_{x \in supp(P)} P(x) log \frac{1}{P(x)}$$

Let us assume λ_x is another notation for $P(x)$.

$$H(X) = \sum_{x \in supp(P)} \lambda_x log \frac{1}{P(x)}$$

This is a convex combination of $log \frac{1}{P(x)} : x \in supp(P_X)$ as $\sum_x P(x) = 1$.

$$H(X) \leq log \left(\sum_{x \in supp(P_X)} \lambda_x \frac{1}{P(x)} \right) \leq log |supp(P_x)|$$

$$\Rightarrow H(X) \leq |\mathcal{X}|$$

$H(X)$ is $log|\mathcal{X}|$ only when.

4