Statistical measures of information

Ex. 1 (15 points)

A bag contains 4 yellow balls, 3 black balls, and 2 white balls. Keep results with 3 significant figures.

1) We randomly draw 3 balls **without replacement**. Calculate the amount of information (in bits) when all 3 balls are yellow. (5 points)

$$p(E_1) = \frac{C_4^3}{C_9^3}, \quad I(E_1) = -\,log_2\big(p(E_1)\big) = 4.39\; bits$$

2) We randomly draw 3 balls **with replacement**. Calculate the amount of information (in bits) for drawing three balls of different colors. (5 points)

$$p(E_2) = 3! \times \frac{4 \times 3 \times 2}{9^3}, \quad I(E_2) = -\,log_2\big(p(E_2)\big) = 2.34\; bits$$

3) We randomly draw 3 balls **without replacement**. Calculate the amount of information (in bits) when at least one of the balls is white. (5 points)

$$p(E_3) = 1 - \frac{C_7^3}{C_9^3}, \quad I(E_3) = -\,log_2\big(p(E_3)\big) = 0.778\; bits$$

Ex. 2 (20 points)

A fitness center surveyed its members to understand their exercise preferences. It was found that:

- 70% of members prefer cardio exercises**(C),** while 30% prefer strength training**(S).**
- Among cardio enthusiasts, 50% exercise in the morning**(M).**
- Among strength training enthusiasts, 20% exercise in the morning.

1) What is the probability that a person who prefers cardio exercises knowing that he likes to exercise in the morning? (5 points):

First, calculate $P(M)$(total probability of exercising in the morning):

$$P(M) = P(M|C) \cdot P(C) + P(M|S) \cdot P(S) = 0.5 \cdot 0.7 + 0.2 \cdot 0.3 = 0.41$$

Now, use Bayes' theorem to find $P(C \mid M)$:

$$P(C|M) = \frac{P(M|C) \cdot P(C)}{P(M)} = \frac{0.5 \cdot 0.7}{0.41} \approx 0.854$$

2) If you know that someone prefers cardio exercises, how much information (in bits) do you gain by learning that they exercise in the morning? (5 points)

$$I(M|C) = -log_2 P(M|C) = -log_2(0.5) = 1\; bit$$

3) If you know that someone prefers strength training, how much information (in bits) do you gain by learning that they exercise in the morning? (5 points)

$$I(M|S) = -log_2 P(M|S) = -log_2(0.2) \approx 2.32\; bits$$

4) Finally, if you know that someone exercises in the morning, how much information (in bits) do you gain by learning that they prefer strength training? (5 points)

$$P(S|M) = \frac{P(M|S) \cdot P(S)}{P(M)} = \frac{0.2 \cdot 0.3}{0.41} \approx 0.146$$
$$I(S|M) = -log_2 P(S|M) = -log_2(0.146) \approx 2.77\ bits$$

Ex. 3 (20 points)

We have 10 coins:

- **8 ordinary coins(O)**: Equal probability of heads(H) or tails(T).
- **2 double-headed coins(D)**: Always heads.

1) If we toss a coin, how much information can we gain when it comes up heads? And, how much information can we gain when it comes up tails? (10 points)
$$I(H) = -\,log_2\, p\,(H) = -\,log_2(0.6) = 0.737\ bits,$$
$$I(T) = -\,log_2\, p\,(T) = -\,log_2(0.4) = 1.322\ bits.$$
2) If we toss the coin and it comes up heads, how much information can we gain after we know it is one of the eight ordinary coins? (5 points)
$$P(O|H) = \frac{P(H|O)p(O)}{P(H)} = \frac{0.5 \times 0.8}{0.6},$$
$$I(O|H) = -\,log_2\big(p(O|H)\big) = 0.585\ bits.$$
3) If we toss the coin one further time and it comes up head again, how much information can we gain after we know it is one of the eight ordinary coins? (5 points)
$$P(2H) = P(2H|O) \times P(O) + P(2H|D) \times P(D) = 0.25 \times 0.8 + 1 \times 0.2 = 0.4,$$
$$P(O|2H) = \frac{P(2H|O)P(O)}{P(2H)} = 0.5,$$
$$I(O|2H) = -\,log_2\big(p(O|2H)\big) = 1\ bit.$$

Information source entropy

Ex. 4 (30 points)

A source sends messages to a noisy communication channel. The output of the source is represented by a random variable X whose sample space is $\{a, b, c, d\}$ with equal probabilities. The output from this channel is represented by a random variable Y who has the identical sample space. The joint probability of these two random variables is given as follows:

$$\begin{array}{ccccc}
 & x=a & x=b & x=c & x=d \\
y=a & \frac{1}{8} & \frac{1}{16} & \frac{1}{16} & 0 \\
y=b & \frac{1}{16} & \frac{1}{8} & \frac{1}{16} & \frac{1}{4} \\
y=c & \frac{1}{32} & 0 & \frac{1}{16} & 0 \\
y=d & \frac{1}{32} & \frac{1}{16} & \frac{1}{16} & 0
\end{array}$$

1) Construct the mathematical model of X, i.e., $\begin{bmatrix} X \\ P_X \end{bmatrix}$, and compute the entropy $H(X)$ in bit/symbol. (10 points)

$$\begin{bmatrix} X \\ P_X \end{bmatrix} = \begin{bmatrix} a, & b, & c, & d \\ \frac{1}{4}, & \frac{1}{4}, & \frac{1}{4}, & \frac{1}{4} \end{bmatrix}$$

$$H(X) = -4 \cdot \frac{1}{4} \cdot log_2 \frac{1}{4} = 2\ bits/symbol.$$

2) Construct the mathematical model of Y, i.e., $\begin{bmatrix} Y \\ P_Y \end{bmatrix}$, and compute the entropy $H(Y)$ in bit/symbol. (10 points)

$$\begin{bmatrix} Y \\ P_Y \end{bmatrix} = \begin{bmatrix} a, & b, & c, & d \\ \frac{1}{4}, & \frac{1}{2}, & \frac{3}{32}, & \frac{5}{32} \end{bmatrix}$$

$H(Y) = -\frac{1}{4} \cdot \log_2 \frac{1}{4} - \frac{1}{2} \cdot \log_2 \frac{1}{2} - \frac{3}{32} \cdot \log_2 \frac{3}{32} - \frac{5}{32} \cdot \log_2 \frac{5}{32} = 1.74$ bit/symbol.

3) What is the joint entropy $H(XY)$ of the two random variables in bit/symbol? (5 points)

$$H(XY) = -\sum_{x,y} P_{XY}(x,y)\, log_2\, P_{XY}\,(x,y)$$

$$= -2 \cdot \frac{1}{8} \cdot \log_2 \frac{1}{8} - 7 \cdot \frac{1}{16} \cdot \log_2 \frac{1}{16} - 2 \cdot \frac{1}{32} \cdot \log_2 \frac{1}{32} - \frac{1}{4} \cdot \log_2 \frac{1}{4} = 3.3125\ bits/symbol$$

4) What is the conditional entropy $H(Y|X)$ in bit/symbol? (5 points)

$$H(Y|X) = H(XY) - H(X) = 1.3125\ bits/symbol.$$

Ex. 5 (15 points)

We consider the information theory of languages. In order to understand naturally occurring languages, we consider the models for finite languages $\mathcal{L}$ consisting of strings of fixed finite length N together with a probability function P which models the natural language. In what follows, for two strings X and Y we denote their concatenation by XY.

1) Consider the language of 1-character strings over $\{A, B, C, D\}$ with associated probabilities 1/4, 1/8, 1/2, and 1/8. What is its corresponding entropy? (5 points)

$$H(L_1) = \frac{1}{4} \cdot \log_2 4 + \frac{1}{8} \cdot \log_2 8 + \frac{1}{2} \cdot \log_2 2 + \frac{1}{8} \cdot \log_2 8 = 1.75\ bits/symbol$$

2) Consider the language $\mathcal{L}_2$ of all strings of length 2 in $\{A, B, C, D\}$ defined by the probability function of the above question and 2-character independence: $p(XY) = p(X)p(Y)$. What is

the entropy of this language? (5 points)

$$H(L_2) = H(L_1) + H(L_1) = 2 \cdot H(L_1) \;=\; 3.5\, bits/symbol$$

3) A source emits symbols $X \in \{A, B\}$ with probabilities $P(A) = 3/4,\ P(B) = 1/4$. After transmission through a noisy channel, the output $Y \in \{A, B\}$ has the following conditional probabilities:

$$P(\,Y = A \mid X = A\,) = 0.8,$$
$$P(\,Y = B \mid X = A\,) = 0.2,$$
$$P(\,Y = A \mid X = B\,) = 0.4,$$
$$P(\,Y = B \mid X = B\,) = 0.6.$$

Calculate the mutual information $I(X;Y)$. (5 points)

	$X = A$	$X = B$
$Y = A$	0.6	0.1
$Y = B$	0.15	0.15

$$I(X;Y) = 0.6 \times \log_2 \frac{0.6}{0.7 \times 0.75} + 0.1 \times \log_2 \frac{0.1}{0.7 \times 0.25} + 0.15 \times \log_2 \frac{0.15}{0.3 \times 0.75} + 0.15 \times \log_2 \frac{0.15}{0.3 \times 0.25} = 0.097\ bits/symbol$$