Problem 1: Suppose P is a probability measure defined on $S = \{1, 2, 3, 4, 5\}$ with mass function

s	$p(s)$
1	0.1
2	0.3
3	0.2
4	0.3
5	0.1

Compute the information content $I(s)$ of each sample point and the entropy $H(P)$.

We compute

s	$p(s)$	$I(s)$
1	0.1	3.322
2	0.3	1.737
3	0.2	2.322
4	0.3	1.737
5	0.1	3.322

where the surprisals are rounded to three places after the decimal point. The entropy is $H(P) \approx 2.171$.

Problem 2: Let $X \sim \mathcal{B}er(\theta)$ for $\theta \in [0, 1]$. Compute a formula for $H(X)$ in terms of θ.

We compute:

$$H(X) = -\sum_{x=0}^{1} p(x)\log_2(p(x)) = -(1-\theta)\log_2(1-\theta) - \theta\log_2(\theta).$$

Problem 3: Suppose P and Q are probability measures defined on $S = \{1, 2, 3, 4, 5\}$ with mass functions

s	$p(s)$	$q(s)$
1	0.1	0.05
2	0.3	0.15
3	0.2	0.7
4	0.3	0.03
5	0.1	0.07

Compute the cross entropies $H_P(Q)$ and $H_Q(P)$.

$$H_P(Q) = -\sum_{s=1}^{5} p(s)\log_2(q(s)) \approx 3.258, \quad H_Q(P) = -\sum_{s=1}^{5} q(s)\log_2(p(s)) \approx 2.337.$$

Problem 4: Suppose P and Q are probability measures defined on $S = \{1, 2, 3, 4, 5\}$ with mass functions

s	$p(s)$	$q(s)$
1	0.1	0.05
2	0.3	0.15
3	0.2	0.7
4	0.3	0.03
5	0.1	0.07

Compute $D(P \parallel Q)$ and $D(Q \parallel P)$.

$$D(P \parallel Q) = \sum_{s=1}^{5} p(s) \log_2 \left(\frac{p(s)}{q(s)} \right) \approx 1.087 \quad \text{and} \quad D(Q \parallel P) = \sum_{s=1}^{5} q(s) \log_2 \left(\frac{q(s)}{p(s)} \right) \approx 0.929.$$

Problem 5: While it is **not** true that $D(P \parallel Q) = D(Q \parallel P)$ for **all** P and Q, there are examples of special distributions P and Q for which equality does hold. Find examples.

For $\theta \in [0, 1]$, define P_θ and Q_θ on the sample space $S = \{0, 1\}$ by

s	$p_\theta(s)$	$q_\theta(s)$
0	θ	$1 - \theta$
1	$1 - \theta$	θ

One then easily proves $D(P_\theta \parallel Q_\theta) = D(Q_\theta \parallel P_\theta)$.