香港中文大學
The Chinese University of Hong Kong

2.5 Informational Divergence

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:

  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:

    1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:

  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:

  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$
  2. $c \log \frac{c}{0} = \infty$ for $c > 0$

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:
  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$
  2. $c \log \frac{c}{0} = \infty$ for $c > 0$
  3. If $D(p\|q) < \infty$, then $p(x) > 0 \Rightarrow q(x) > 0$, or $\mathcal{S}_p \subset \mathcal{S}_q$.

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:
  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$
  2. $c \log \frac{c}{0} = \infty$ for $c > 0$
  3. If $D(p\|q) < \infty$, then $p(x) > 0 \Rightarrow q(x) > 0$, or $\mathcal{S}_p \subset \mathcal{S}_q$.
- $D(p\|q)$ measures the "distance" between p and q.

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:
  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$
  2. $c \log \frac{c}{0} = \infty$ for $c > 0$
  3. If $D(p\|q) < \infty$, then $p(x) > 0 \Rightarrow q(x) > 0$, or $\mathcal{S}_p \subset \mathcal{S}_q$.
- $D(p\|q)$ measures the "distance" between p and q.
- $D(p\|q)$ is not symmetrical in p and q, so $D(\cdot\|\cdot)$ is not a true metric.

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

- Convention:
  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$
  2. $c \log \frac{c}{0} = \infty$ for $c > 0$
  3. If $D(p\|q) < \infty$, then $p(x) > 0 \Rightarrow q(x) > 0$, or $\mathcal{S}_p \subset \mathcal{S}_q$.
- $D(p\|q)$ measures the "distance" between p and q.
- $D(p\|q)$ is not symmetrical in p and q, so $D(\cdot\|\cdot)$ is not a true metric.
- $D(\cdot\|\cdot)$ does not satisfy the triangular inequality.

Definition 2.28 The informational divergence between two probability distributions p and q on a common alphabet $\mathcal{X}$ is defined as

$$D(p\|q) = \sum_x p(x) \log \frac{p(x)}{q(x)} = E_p \log \frac{p(X)}{q(X)},$$

where E_p denotes expectation with respect to p.

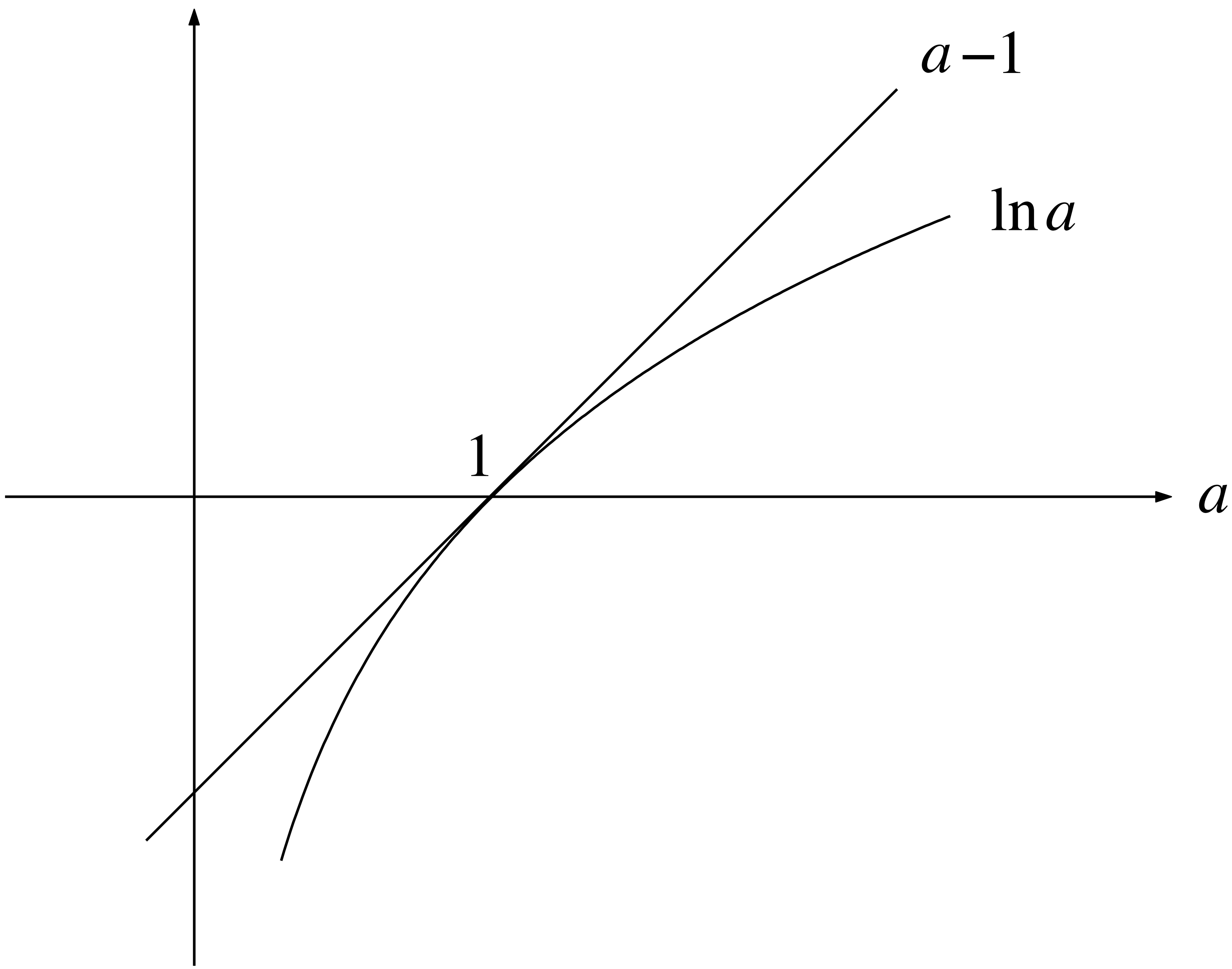
- Convention:
  1. Summation is over $\mathcal{S}_p$, i.e., $\sum_{x \in \mathcal{S}_p}$
  2. $c \log \frac{c}{0} = \infty$ for $c > 0$
  3. If $D(p\|q) < \infty$, then $p(x) > 0 \Rightarrow q(x) > 0$, or $\mathcal{S}_p \subset \mathcal{S}_q$.
- $D(p\|q)$ measures the "distance" between p and q.
- $D(p\|q)$ is not symmetrical in p and q, so $D(\cdot\|\cdot)$ is not a true metric.
- $D(\cdot\|\cdot)$ does not satisfy the triangular inequality.
- Also called *relative entropy* or the *Kullback-Leibler distance*.

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

$$\ln a \leq a - 1$$



Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \le a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \ge 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \le a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \ge 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \le a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \ge 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \le a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \ge 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\ln \frac{1}{b} \quad \le \quad \frac{1}{b} - 1$$

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\begin{aligned} \ln \frac{1}{b} &\leq \frac{1}{b} - 1 \\ -\ln b &\leq \frac{1}{b} - 1 \end{aligned}$$

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\begin{aligned}
\ln \frac{1}{b} &\leq \frac{1}{b} - 1 \\
-\ln b &\leq \frac{1}{b} - 1 \\
\ln b &\geq 1 - \frac{1}{b}
\end{aligned}$$

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\begin{aligned} \ln \frac{1}{b} &\leq \frac{1}{b} - 1 \\ -\ln b &\leq \frac{1}{b} - 1 \\ \ln b &\geq 1 - \frac{1}{b} \end{aligned}$$

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\begin{aligned} \ln \frac{1}{b} &\leq \frac{1}{b} - 1 \\ -\ln b &\leq \frac{1}{b} - 1 \\ \ln b &\geq 1 - \frac{1}{b} \end{aligned}$$

Equality holds if and only if $\frac{1}{b} = a = 1$, or $b = 1$.

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\begin{aligned} \ln \frac{1}{b} &\leq \frac{1}{b} - 1 \\ -\ln b &\leq \frac{1}{b} - 1 \\ \ln b &\geq 1 - \frac{1}{b} \end{aligned}$$

Equality holds if and only if $\frac{1}{b} = a = 1$, or $b = 1$.

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\begin{aligned} \ln \frac{1}{b} &\leq \frac{1}{b} - 1 \\ -\ln b &\leq \frac{1}{b} - 1 \\ \ln b &\geq 1 - \frac{1}{b} \end{aligned}$$

Equality holds if and only if $\frac{1}{b} = a = 1$, or $b = 1$.

Lemma 2.29 (Fundamental Inequality) For any $a > 0$,

$$\ln a \leq a - 1$$

with equality if and only if $a = 1$.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Proof Let $a = \frac{1}{b}$ in the fundamental inequality, where $b > 0$. Then

$$\begin{aligned} \ln \frac{1}{b} &\leq \frac{1}{b} - 1 \\ -\ln b &\leq \frac{1}{b} - 1 \\ \ln b &\geq 1 - \frac{1}{b} \end{aligned}$$

Equality holds if and only if $\frac{1}{b} = a = 1$, or $b = 1$.

Theorem 2.31 (Divergence Inequality) For any two probability distributions p and q on a common alphabet $\mathcal{X}$,

$$D(p\|q) \geq 0$$

with equality if and only if $p = q$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$D(p\|q) \quad = \quad \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)}$$

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$D(p\|q) = \sum_{x\in\mathcal{S}_p} p(x) \underline{\log} \frac{p(x)}{q(x)}$$

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \underline{\log} \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \underline{\ln} \frac{p(x)}{q(x)} \end{aligned}$$

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \underline{\log} \frac{p(x)}{q(x)} \\ &= \underline{(\log e)} \sum_{x \in \mathcal{S}_p} p(x) \underline{\ln} \frac{p(x)}{q(x)} \end{aligned}$$

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \qquad (1)$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \end{aligned}$$

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \end{aligned}$$

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\ &\geq (\log e) \sum_{x \in \mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x\in\mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\ &\geq (\log e) \sum_{x\in\mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\ &\geq (\log e) \sum_{x \in \mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\ &\geq (\log e) \sum_{x \in \mathcal{S}_p} \cancel{p(x)} \left(1 - \frac{q(x)}{\cancel{p(x)}}\right) \end{aligned} \tag{2}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x\in\mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\ &\geq (\log e) \sum_{x\in\mathcal{S}_p} \cancel{p(x)} \left(1 - \frac{q(x)}{\cancel{p(x)}}\right) \qquad (2) \\ &= (\log e) \left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_p} q(x)\right] \end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned} D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\ &= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\ &\geq (\log e) \sum_{x \in \mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \\ &= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_p} q(x) \right] \end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\
&= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\
&\geq (\log e) \sum_{x \in \mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_p} q(x) \right]
\end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x)\log\frac{p(x)}{q(x)} \\
&= (\log e)\sum_{x\in\mathcal{S}_p} p(x)\ln\frac{p(x)}{q(x)} \\
&\geq (\log e)\sum_{x\in\mathcal{S}_p} p(x)\left(1-\frac{q(x)}{p(x)}\right) \qquad (2)\\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_p} q(x)\right] \\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_q} q(x)\right]
\end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\
&= (\log e) \sum_{x\in\mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\
&\geq (\log e) \sum_{x\in\mathcal{S}_p} \cancel{p(x)} \left(1 - \frac{q(x)}{\cancel{p(x)}}\right) \qquad (2) \\
&= (\log e) \left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_p} q(x)\right] \\
&= (\log e) \left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_q} q(x)\right]
\end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\
&= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\
&\geq (\log e) \sum_{x \in \mathcal{S}_p} \cancel{p(x)} \left(1 - \frac{q(x)}{\cancel{p(x)}}\right) \qquad (2) \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_p} q(x)\right] \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_q} q(x)\right] \\
&= (\log e)[1 - 1]
\end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\
&= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\
&\geq (\log e) \sum_{x \in \mathcal{S}_p} \cancel{p(x)} \left(1 - \frac{q(x)}{\cancel{p(x)}}\right) \qquad (2) \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_p} q(x)\right] \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_q} q(x)\right] \\
&= (\log e)[1 - 1] \\
&= 0.
\end{aligned}$$

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x)\log\frac{p(x)}{q(x)} \\
&= (\log e)\sum_{x\in\mathcal{S}_p} p(x)\ln\frac{p(x)}{q(x)} \\
&\geq (\log e)\sum_{x\in\mathcal{S}_p} \cancel{p(x)}\left(1-\frac{q(x)}{\cancel{p(x)}}\right) \qquad (2)\\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_p} q(x)\right] \\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_q} q(x)\right] \\
&= (\log e)[1-1] \\
&= 0.
\end{aligned}$$

This proves (1).

> **Corollary 2.30** For any $a > 0$,
>
> $$\ln a \geq 1 - \frac{1}{a}$$
>
> with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\
&= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\
&\geq (\log e) \sum_{x \in \mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_p} q(x)\right] \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_q} q(x)\right] \\
&= (\log e)[1 - 1] \\
&= 0.
\end{aligned}$$

This proves (1).

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\
&= (\log e) \sum_{x\in\mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\
&\geq (\log e) \sum_{x\in\mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \\
&= (\log e) \left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_p} q(x)\right] \\
&= (\log e) \left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_q} q(x)\right] \\
&= (\log e)[1 - 1] \\
&= 0.
\end{aligned}$$

This proves (1).

3. For equality to hold in (2), we see from Corollary 2.30 that this is the case if and only if

$$\frac{p(x)}{q(x)} = 1 \text{ or } p(x) = q(x) \quad \text{for all } x \in \mathcal{S}_p.$$

This proves the theorem.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x \in \mathcal{S}_p} p(x) \log \frac{p(x)}{q(x)} \\
&= (\log e) \sum_{x \in \mathcal{S}_p} p(x) \ln \frac{p(x)}{q(x)} \\
&\geq (\log e) \sum_{x \in \mathcal{S}_p} p(x) \left(1 - \frac{q(x)}{p(x)}\right) \qquad (2) \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_p} q(x) \right] \\
&= (\log e) \left[\sum_{x \in \mathcal{S}_p} p(x) - \sum_{x \in \mathcal{S}_q} q(x) \right] \\
&= (\log e)[1 - 1] \\
&= 0.
\end{aligned}$$

This proves (1).

3. For equality to hold in (2), we see from Corollary 2.30 that this is the case if and only if

$$\frac{p(x)}{q(x)} = 1 \text{ or } p(x) = q(x) \quad \text{for all } x \in \mathcal{S}_p.$$

This proves the theorem.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \qquad (1)$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x)\log\frac{p(x)}{q(x)} \\
&= (\log e)\sum_{x\in\mathcal{S}_p} p(x)\ln\frac{p(x)}{q(x)} \\
&\geq (\log e)\sum_{x\in\mathcal{S}_p} p(x)\left(1-\frac{q(x)}{p(x)}\right) \qquad (2)\\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_p} q(x)\right] \\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_q} q(x)\right] \\
&= (\log e)[1-1] \\
&= 0.
\end{aligned}$$

This proves (1).

3. For equality to hold in (2), we see from Corollary 2.30 that this is the case if and only if

$$\frac{p(x)}{q(x)} = 1 \text{ or } p(x) = q(x) \quad \text{for all } x \in \mathcal{S}_p.$$

This proves the theorem.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.31 (Divergence Inequality)

$$D(p\|q) \geq 0 \tag{1}$$

with equality if and only if $p = q$.

Proof

1. For simplicity, assume $\mathcal{S}_p = \mathcal{S}_q$. For a proof without this assumption, see the textbook.

2. Consider

$$\begin{aligned}
D(p\|q) &= \sum_{x\in\mathcal{S}_p} p(x)\log\frac{p(x)}{q(x)} \\
&= (\log e)\sum_{x\in\mathcal{S}_p} p(x)\ln\frac{p(x)}{q(x)} \\
&\geq (\log e)\sum_{x\in\mathcal{S}_p} p(x)\left(1-\frac{q(x)}{p(x)}\right) \qquad (2) \\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_p} q(x)\right] \\
&= (\log e)\left[\sum_{x\in\mathcal{S}_p} p(x) - \sum_{x\in\mathcal{S}_q} q(x)\right] \\
&= (\log e)[1-1] \\
&= 0.
\end{aligned}$$

This proves (1).

3. For equality to hold in (2), we see from Corollary 2.30 that this is the case if and only if

$$\frac{p(x)}{q(x)} = 1 \text{ or } p(x) = q(x) \quad \text{for all } x \in \mathcal{S}_p.$$

This proves the theorem.

Corollary 2.30 For any $a > 0$,

$$\ln a \geq 1 - \frac{1}{a}$$

with equality if and only if $a = 1$.

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i}$$

with the convention that $\log \frac{a_i}{0} = \infty$. Moreover, equality holds if and only if $\frac{a_i}{b_i} = \mathit{constant}$ for all i.

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i}$$

with the convention that $\log \frac{a_i}{0} = \infty$. Moreover, equality holds if and only if $\frac{a_i}{b_i} = \mathit{constant}$ for all i.

Example:

$$a_1 \log \frac{a_1}{b_1} + a_2 \log \frac{a_2}{b_2} \geq (a_1 + a_2) \log \frac{a_1 + a_2}{b_1 + b_2}.$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \mathit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$0 \leq \sum_i a'_i \log \frac{a'_i}{b'_i}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$0 \leq \sum_i a_i' \log \frac{a_i'}{b_i'}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \mathit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$0 \leq \sum_i a'_i \log \frac{a'_i}{b'_i}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a_i' \log \frac{a_i'}{b_i'} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a_i' \log \frac{a_i'}{b_i'} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a_i' \log \frac{a_i'}{b_i'} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \mathit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\ &= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\ &= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right]
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right]
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\ &= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\ &= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right] \end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j}\right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j}\right]
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right]
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right]
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a_i' \log \frac{a_i'}{b_i'} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \underline{\frac{\sum_j a_j}{\sum_j b_j}} \right]
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \underline{\frac{\sum_j a_j}{\sum_j b_j}} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \underline{\left(\sum_i a_i\right)} \log \frac{\sum_j a_j}{\sum_j b_j} \right],
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i\right) \log \frac{\sum_j a_j}{\sum_j b_j} \right],
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i\right) \log \frac{\sum_j a_j}{\sum_j b_j} \right],
\end{aligned}$$

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} =$ *constant* for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i\right) \log \frac{\sum_j a_j}{\sum_j b_j} \right],
\end{aligned}$$

which implies (1).

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \tag{1}$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \textit{constant}$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a_i' \log \frac{a_i'}{b_i'} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i\right) \log \frac{\sum_j a_j}{\sum_j b_j} \right],
\end{aligned}$$

which implies (1).

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \mathit{constant}$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a_i' \log \frac{a_i'}{b_i'} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j}\right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j}\right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i\right) \log \frac{\sum_j a_j}{\sum_j b_j}\right],
\end{aligned}$$

which implies (1).

3. Equality holds if and only if for all i,

$$a_i' = b_i' \quad \text{or} \quad \frac{a_i}{b_i} = \mathit{constant}.$$

The theorem is proved.

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a_i' = a_i / \sum_j a_j$ and $b_i' = b_i / \sum_j b_j$. Then $\{a_i'\}$ and $\{b_i'\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a_i' \log \frac{a_i'}{b_i'} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j}\right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j}\right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i\right) \log \frac{\sum_j a_j}{\sum_j b_j}\right],
\end{aligned}$$

which implies (1).

3. Equality holds if and only if for all i,

$$a_i' = b_i' \quad \text{or} \quad \frac{a_i}{b_i} = constant.$$

The theorem is proved.

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i \right) \log \frac{\sum_j a_j}{\sum_j b_j} \right],
\end{aligned}$$

which implies (1).

3. Equality holds if and only if for all i,

$$a'_i = b'_i \quad \text{or} \quad \frac{a_i}{b_i} = constant.$$

The theorem is proved.

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i\right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = \mathit{constant}$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned}
0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\
&= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j}\right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j}\right] \\
&= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i\right) \log \frac{\sum_j a_j}{\sum_j b_j}\right],
\end{aligned}$$

which implies (1).

3. Equality holds if and only if for all i,

$$a'_i = b'_i \quad \text{or} \quad \frac{a_i}{b_i} = \mathit{constant}.$$

The theorem is proved.

Theorem 2.32 (Log-Sum Inequality) For positive numbers $a_1, a_2, \cdots$ and nonnegative numbers $b_1, b_2, \cdots$ such that $\sum_i a_i < \infty$ and $0 < \sum_i b_i < \infty$,

$$\sum_i a_i \log \frac{a_i}{b_i} \geq \left(\sum_i a_i \right) \log \frac{\sum_i a_i}{\sum_i b_i} \qquad (1)$$

Moreover, equality holds if and only if $\frac{a_i}{b_i} = constant$ for all i.

Proof

1. Let $a'_i = a_i / \sum_j a_j$ and $b'_i = b_i / \sum_j b_j$. Then $\{a'_i\}$ and $\{b'_i\}$ are probability distributions.

2. Using the divergence inequality, we have

$$\begin{aligned} 0 &\leq \sum_i a'_i \log \frac{a'_i}{b'_i} \\ &= \sum_i \frac{a_i}{\sum_j a_j} \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \\ &= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i / \sum_j a_j}{b_i / \sum_j b_j} \right] \\ &= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \sum_i a_i \log \frac{\sum_j a_j}{\sum_j b_j} \right] \\ &= \frac{1}{\sum_j a_j} \left[\sum_i a_i \log \frac{a_i}{b_i} - \left(\sum_i a_i \right) \log \frac{\sum_j a_j}{\sum_j b_j} \right], \end{aligned}$$

which implies (1).

3. Equality holds if and only if for all i,

$$a'_i = b'_i \quad \text{or} \quad \frac{a_i}{b_i} = constant.$$

The theorem is proved.

Divergence Inequality vs Log-Sum Inequality

- The divergence inequality implies the log-sum inequality.

Divergence Inequality vs Log-Sum Inequality

- The divergence inequality implies the log-sum inequality.
- The log-sum inequality also implies the divergence inequality. (Exercise)

Divergence Inequality vs Log-Sum Inequality

- The divergence inequality implies the log-sum inequality.
- The log-sum inequality also implies the divergence inequality. (Exercise)
- The two inequalities are equivalent.

Theorem 2.33 (Pinsker's Inequality)

$$D(p\|q) \geq \frac{1}{2\ln 2} V^2(p, q).$$

Theorem 2.33 (Pinsker's Inequality)

$$D(p\|q) \geq \frac{1}{2\ln 2} V^2(p,q).$$

- If $D(p\|q)$ or $D(q\|p)$ is small, then so is $V(p,q) = V(q,p)$.

Theorem 2.33 (Pinsker's Inequality)

$$D(p\|q) \geq \frac{1}{2\ln 2} V^2(p,q).$$

- If $D(p\|q)$ or $D(q\|p)$ is small, then so is $V(p,q) = V(q,p)$.
- For a sequence of probability distributions q_k, as $k \to \infty$, if $D(p\|q_k) \to 0$ or $D(q_k\|p) \to 0$, then $V(p,q_k) = V(q_k,p) \to 0$.

Theorem 2.33 (Pinsker's Inequality)

$$D(p\|q) \geq \frac{1}{2\ln 2} V^2(p, q).$$

- If $D(p\|q)$ or $D(q\|p)$ is small, then so is $V(p, q) = V(q, p)$.
- For a sequence of probability distributions q_k, as $k \to \infty$, if $D(p\|q_k) \to 0$ or $D(q_k\|p) \to 0$, then $V(p, q_k) = V(q_k, p) \to 0$.
- That is, "convergence in divergence" is a stronger notion than "convergence in variational distance."

Theorem 2.33 (Pinsker's Inequality)

$$D(p\|q) \geq \frac{1}{2\ln 2} V^2(p,q).$$

- If $D(p\|q)$ or $D(q\|p)$ is small, then so is $V(p,q) = V(q,p)$.
- For a sequence of probability distributions q_k, as $k \to \infty$, if $D(p\|q_k) \to 0$ or $D(q_k\|p) \to 0$, then $V(p,q_k) = V(q_k,p) \to 0$.
- That is, "convergence in divergence" is a stronger notion than "convergence in variational distance."
- See Problems 23 and 24 for details.