

# 1 情報理論

## 1.1 推定

集団から一部を取り出し母集団の性質を推測すること。

### 1.1.1 標本平均

標本データの平均をとったもの。一致性と不偏性を持つため点推定によく用いられる。

### 1.1.2 標本分散

標本データの分散をとったもの。一致性は持つが不偏性は持たない。 $\frac{n}{n-1}$ をかけたものを不偏分散といい、これは不偏性を持つ。

## 1.2 情報量

あるできごと（事象）が起きた際、それがどれほど起こりにくいかを表す尺度である。

### 1.2.1 自己情報量

事象  $x$  が確率  $P(x)$  で起こるとき、自己情報量  $I(x)$  の定義を以下に示す。

$$I(x) = -\log(P(x))$$

また、自己情報量の期待値をとったものをシャノンエントロピー  $H(x)$  といい、以下のように表される。

$$H(x) = E(I(x)) = -\sum (P(x) \log P(x))$$

### 1.2.2 KL ダイバージェンス

2つの確率分布の違いを表すのに用いられる。確率分布  $P(x)$ 、 $Q(x)$  の KL ダイバージェンス  $D_{KL}(P||Q)$  の定義を以下に示す。

$$D_{KL}(P||Q) = E_{x \sim P}(-\log \frac{Q(x)}{P(x)})$$

$$D_{KL}(P||Q) = \sum_x P(x)((-\log Q(x)) - (-\log P(x)))$$

### 1.2.3 交差エントロピー

KL ダイバージェンスの一部。確率分布  $P(x)$ 、 $Q(x)$  の交差エントロピー  $H(P, Q)$  は以下のように表される。

$$H(P, Q) = E_{x \sim P}(-\log Q(x)) = -\sum (P(x) \log Q(x))$$