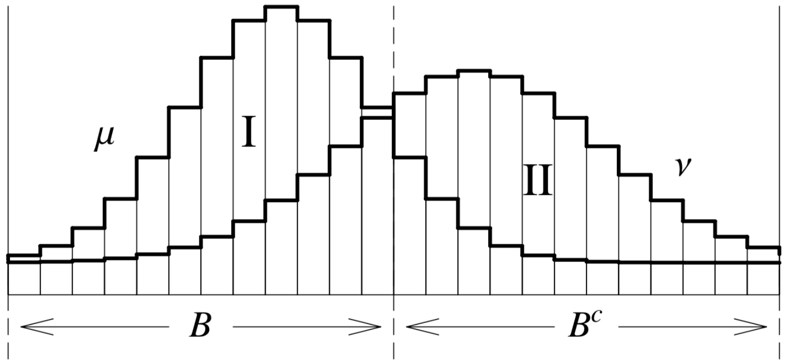
0.1 Fact About Total Variation Distance



定义 0.1. total variation distance

π 和 π' 是两个不同的概率分布，他们都定义在一个可数集合 Ω 上. 我们称 pi 和 π' 之间的 total variation distance 为:

$$||\mu - \nu||_{TV} := \max_{A \subseteq \Omega} |\mu(A) - \nu(A)|$$

♣

命题 0.1

Proposition 4.2. *Let μ and ν be two probability distributions on $\mathcal{X}$. Then*

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)| . \tag{4.2}$$

Proof. Let $B = \{x : \mu(x) \geq \nu(x)\}$ and let $A \subset \mathcal{X}$ be any event. Then

$$\mu(A) - \nu(A) \leq \mu(A \cap B) - \nu(A \cap B) \leq \mu(B) - \nu(B). \tag{4.3}$$

The first inequality is true because any $x \in A \cap B^c$ satisfies $\mu(x) - \nu(x) < 0$, so the difference in probability cannot decrease when such elements are eliminated. For the second inequality, note that including more elements of B cannot decrease the difference in probability.

By exactly parallel reasoning,

$$\nu(A) - \mu(A) \leq \nu(B^c) - \mu(B^c). \tag{4.4}$$

Fortunately, the upper bounds on the right-hand sides of (4.3) and (4.4) are actually the same (as can be seen by subtracting them; see Figure 4.1). Furthermore, when we take $A = B$ (or B^c), then $|\mu(A) - \nu(A)|$ is equal to the upper bound. Thus

$$\|\mu - \nu\|_{\mathrm{TV}} = \frac{1}{2} [\mu(B) - \nu(B) + \nu(B^c) - \mu(B^c)] = \frac{1}{2} \sum_{x \in \mathcal{X}} |\mu(x) - \nu(x)|.$$

■

♠