16. 实例依赖性下界

在上一章中，我们证明了在[0, 1]中具有次优性间隙的亚高斯老虎机的极大极小遗憾的下界。这样的界限可以作为衡量决策鲁棒性的有用指标，但通常过于保守。本章致力于理解实例依赖性下界，它试图捕捉决策在特定老虎机实例上的最佳性能。

由于遗憾是一个多目标准则，算法设计者可能会尝试设计在某种实例上表现良好的算法。一个极端的例子是为所有 t 选择 $A_t = 1$ 的决策，当第一个臂为最优时，该决策将遭受零遗憾，否则将遭受线性遗憾。这是一个苛刻的权衡，仅在少数情况下将遗憾从对数减少到零的代价是其他情况下的线性遗憾。令人惊讶的是，这就是老虎机游戏的本质。可以为每个实例指定一个难度度量，这样在某些实例上相对于此度量执行得太好的决策会为其他实例付出高昂的代价。情况如图 16.1 所示

在有限的时间内，情况有点混乱，但如果将这些想法推向极限，那么对于许多类别的老虎机来说，可以定义依赖实例的最优性的精确概念。

个人注 1：次优差距(sub-optimality gap)是指该策略与最优策略之间的性能差距。

个人注 2：本章主要论述了适用于任何非结构化类的随机老虎机的通用下界，论证了有限时间实例依赖性下界，同时提供了 $\mathcal{M}$ 对应的 $d_{\inf}\left(P,\mu^*,\mathcal{M}\right)$ 的明确公式。

16.1 渐近界

我们需要准确定义合理决策的含义。如果只关心渐近性，那么一个相当保守的定义就足够了。

定义 16.1。如果对于所有 $\nu \in \varepsilon$ 和 $p > 0$，在一类老虎机 ε 上称决策 π 是一致的，则有

$$\lim_{n\to\infty}\frac{R_n(\pi,\nu)}{n^p}=0. \tag{1.1}$$

ε 上的一致决策类用 $\prod_{cons}(\varepsilon)$ 表示。

定理 7.1 表明 UCB 在 $\varepsilon_{SG}^k(1)$ 上是一致的。总是选择第一个动作的决策在任何 ε 上都是不一致的，除非第一个臂对每个 $\nu \in \varepsilon$ 都是最优的。

一致性是一个渐进的概念。一个决策可以是一致的，但对所有 $t \le 10^{100}$ 都是 $A_t = 1$。因此，一致性假设不足以推导出非渐近下界。在第 16.2 节中，我们介绍了一个有限时间版本的一致性，它允许我们证明有限时间实例依赖性下界。

回想一下，如果 $\varepsilon = \mathcal{M}_1 \times \cdots \times \mathcal{M}_k$ 具有 $\mathcal{M}_1,\cdots,\mathcal{M}_k$ 分布集，则 ε 类随机老虎机是非结构化的。本章的主要定理是一个通用的下界，适用于任何非结构化类的随