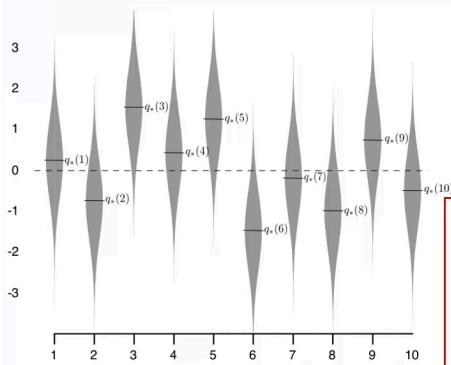


# Greedy和 $\epsilon$ -Greedy策略

区别在于如何处理最优动作的可能性，就是对 $\pi(a|s)$ 不同的处理方式。

## Greedy和 $\epsilon$ -Greedy 温和的贪婪策略



*Reward distribution*

### *Greedy*

### *$\epsilon$ -Greedy*

**Step0:** 奖励列表 $Q(i)$ , 动作概率列表 $P(i)$ ,  $\epsilon$ , 动作选择次数列表 $N(i)$ , 老虎机奖励分布

**Step1:** 初始化 $Q(i) = 0, N(i) = 0, P(i) = 1/10$

**Step2:** 按照  $P(i)$  随机生成动作  $a$ , 产生reward

**Step3:** 根据reward计算更新  $Q(i)$

**Step4:**  $a_{best} = \arg \max_a Q(i)$

**Step5:**  $P(a_{best}) = 1$   
 $P(\text{else}) = 0$

$P(a_{best}) = 1 - \epsilon + \epsilon/10$   
 $P(\text{else}) = \epsilon/10$