

## 第一章 文献翻译&个人理解

### 15 极大极小下界

在短暂的讨论信息论偏移之后，让我们回到 **k-armed** 随机赌博机上。在下面的内容中，我们规定范围  $n > 0$  和动作次数  $k > 1$ 。本章有两个组成部分。首先是针对固定策略和不同的赌博机，精确计算典型赌博机模型中措施之间的相对熵。在第二部分中，我们证明了一个极大极小下界，将第 13 章给出的直觉论点形式化。

**个人注：**本章主要讨论决策理论问题中极大极小风险的下界。这种界限对于评估决策规则的质量很有用。

#### 15.1 赌博机之间的相对熵

下面的结果将被反复使用，练习题中提供了一些概括。

**引理 15.1：散度分解**

令  $\nu = (P_1, \dots, P_k)$  为一个 **k-armed** 赌博机相关的奖励分布，令  $\nu' = (P'_1, \dots, P'_k)$  为另一个 **k-armed** 赌博机相关的奖励分布。固定一些政策  $\pi$ ，同时令  $\mathbb{P}_\nu = \mathbb{P}_{\nu\pi}$  和  $\mathbb{P}_{\nu'} = \mathbb{P}_{\nu'\pi}$  为由  $\pi$  和  $\nu$ （ $\pi$  和  $\nu'$ ）的  $n$  轮相互联络所得到的正则赌博机模型（4.6 节）的概率测度。因此有：

$$D(\mathbb{P}_\nu, \mathbb{P}_{\nu'}) = \sum_{i=1}^k \mathbb{E}_\nu[T_i(n)] D(P_i, P'_i)$$

证明：

假设对于所有  $i \in [k]$  都有  $D(P_i, P'_i) < \infty$ 。由此可见  $P_i \ll P'_i$ 。定义  $\lambda = \sum_{i=1}^k P_i + P'_i$ ，

$\lambda(A) = \sum_{i=1}^k (P_i(A) + P'_i(A))$  为任意可测集  $A$  定义的度量，定理 14.1 表明，只要

$(\frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu'}}) < +\infty$ ，则有：

$$D(\mathbb{P}_\nu, \mathbb{P}_{\nu'}) = \sum_{i=1}^k \mathbb{E}_\nu[\log(\frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu'}})]$$

回顾  $\rho$  是  $[k]$  上的计数测度，我们发现  $\mathbb{P}_\nu$  关于乘积测度  $(\rho \times \lambda)^n$  的 Radon-Nikodym 导数在式（4.7）中给出：

$$p_{\nu\pi}(a_1, x_1, \dots, a_n, x_n) = \prod_{t=1}^n \pi_t(a_t | a_1, x_1, \dots, a_{t-1}, x_{t-1}) p_{a_t}(x_t)$$

除了  $p_{a_t}$  被  $p'_{a_t}$  取代外， $\mathbb{P}_\nu$  的密度相同，则有：

$$\log \frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu'}}(a_1, x_1, \dots, a_n, x_n) = \sum_{t=1}^n \log \frac{p_{a_t}(x_t)}{p'_{a_t}(x_t)}$$

其中，我们使用了 Radon-Nikodym 衍生品的链规则和涉及策略的条款取消的事实。取双方的期望：

$$\mathbb{E}_\nu[\log(\frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu'}}(A_1, X_1, \dots, A_n, X_n))] = \sum_{t=1}^n \mathbb{E}_\nu[\log \frac{p_{A_t}(X_t)}{p'_{A_t}(X_t)}]$$

以及

$$\mathbb{E}_\nu[\log \frac{p_{A_t}(X_t)}{p'_{A_t}(X_t)}] = \mathbb{E}_\nu[\mathbb{E}_\nu[\log \frac{p_{A_t}(X_t)}{p'_{A_t}(X_t)} | A_t]] = \mathbb{E}_\nu[D(P_{A_t}, P'_{A_t})]$$

其中，第二个等式在  $\mathbb{P}_\nu(\cdot | A_t)$  条件下， $X_t$  的分布是  $dP_{A_t} = p_{A_t} d\lambda$ ，回插到前式中得：

$$\begin{aligned} \mathbb{E}_\nu[\log(\frac{d\mathbb{P}_\nu}{d\mathbb{P}_{\nu'}}(A_1, X_1, \dots, A_n, X_n))] &= \sum_{t=1}^n \mathbb{E}_\nu[\log \frac{p_{A_t}(X_t)}{p'_{A_t}(X_t)}] \\ &= \sum_{t=1}^n \mathbb{E}_\nu[D(P_{A_t}, P'_{A_t})] = \sum_{i=1}^k \mathbb{E}_\nu[\sum_{t=1}^n \mathbb{I}\{A_t = i\} D(P_{A_t}, P'_{A_t})] \\ &= \sum_{i=1}^k \mathbb{E}_\nu[T_i(n)] D(P_{A_i}, P'_{A_i}) \end{aligned}$$

当（15.1）式的右边是无穷大时，由我们前面的计算不难看出，左边也是无穷大的。

我们注意到，无论动作集是否离散，发散分解都成立。在更一般的形式下，有关行动的总和必须用适当的非负措施的积分来代替，该积分概括了预期的臂拔出数量。详情见练习 15.8。

**个人注：**相对熵（relative entropy）又称为 KL 散度（Kullback–Leibler divergence，简称 KLD），信息散度（information divergence），信息增益（information gain）。KL 散度是两个概率分布  $P$  和  $Q$  差别的非对称性的度量。KL 散度是用来度量使用基于  $Q$  的编码来编码来自  $P$  的样本平均所需的额外的位元数。典型情况下， $P$  表示数据的真实分布， $Q$  表示数据的理论分布，模型分布，或  $P$  的近似分布。简单来说，相对熵用来衡量两个取值为正的函数或概率分布之间的差异

#### 15.2 极大极小下界

记  $\varepsilon_N^k(\mathbf{1})$  是具有单位方差的高斯赌博机类，可以通过它们的平均向量进行参数化

$\mu \in \mathbb{R}^k$ ，令  $\mu \in \mathbb{R}^k$ ， $v_\mu$  为第  $i$  个有奖励分布为  $N(\mu_i, 1)$  的高斯赌博机。

**引理 15.2：**

令  $k > 1$  且  $n \geq k - 1$ ，因此对于任何政策  $\pi$ ，存在一个平均向量  $\mu \in [0, 1]^k$ ，使得：

$$R_n(\pi, v_\mu) \geq \frac{1}{27} \sqrt{(k-1)n}$$

从而  $v_\mu \in \varepsilon_N^k(\mathbf{1})$ ，因此，当  $n \geq k - 1$  时，极大极小边界  $\varepsilon_N^k(\mathbf{1})$  的极小边界当由上述等

式右边主导的边缘下界：

$$R_n^*(\varepsilon_N^k(1)) \geq \frac{1}{27} \sqrt{(k-1)n}$$

证明的思想如图 15.1 所示。

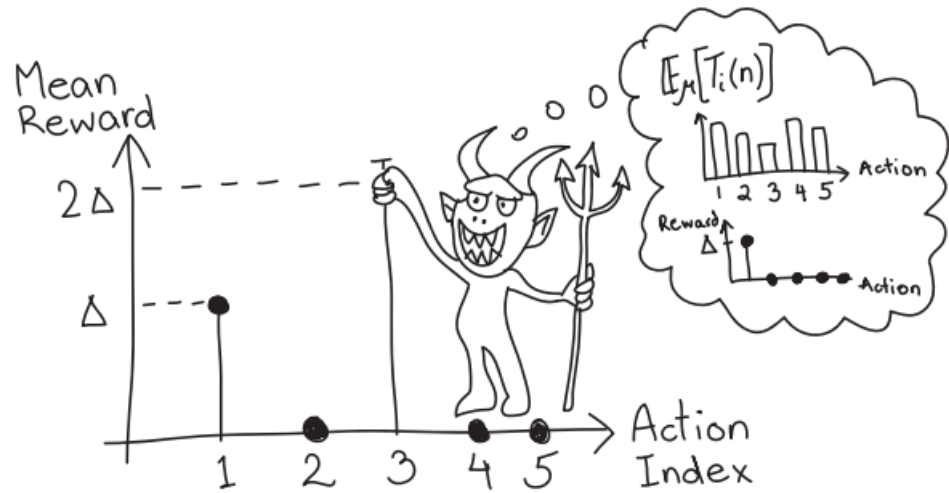


图 15.1 极大极小下界的概念。如果有一个政策和一个环境反对者选择另一个环境，使政策在至少一个环境中遭受巨大的偏差遗憾

证明：

固定策略  $\pi$ ，令  $\Delta \in [0, 1/2]$  是稍后选择的常数。根据第 13 章中的建议，我们从单位方差和平均向量  $\mu = (\Delta, 0, 0, \dots, 0)$  的高斯赌博机开始。这种环境下和  $\pi$  在正则空间上产生了分布  $\mathbb{P}_{v_\mu, \pi}$  和赌博机模型  $(H_n, F_n)$ 。为了简洁起见，我们将使用  $\mathbb{P}_\mu$  来代替  $\mathbb{P}_{v_\mu, \pi}$ 。同时， $\mathbb{P}_\mu$  将用  $\mathbb{E}_\mu$  表示。要选择第二个环境，令

$$i = \arg \min_{j \geq 1} \mathbb{E}_\mu[T_i(n)]$$

从而， $\sum_{j=1}^k \mathbb{E}_\mu[T_j(n)] = n$ ，并且  $\mathbb{E}_\mu[T_i(n)] \leq n/(k-1)$ 。第二个高斯赌博机也是单位方差且均值为

$$\mu' = (\Delta, 0, 0, \dots, 0, 2\Delta, 0, \dots, 0)$$

其中  $\mu'_i = 2\Delta$ 。因此  $\mu_j = \mu'_j$ ，除了指数  $i$  和  $v_\mu$  是第一个臂，而在  $v_\mu$  中，臂  $i$  是最佳的。我们

缩写  $\mathbb{P}_{\mu'} = \mathbb{P}_{v_{\mu'}, \pi}$ 。将引理 4.5 进行一个简单的计算得到

$$R_n(\pi, v_\mu) \geq \mathbb{P}_\mu(T_1(n) \leq n/2) \frac{n\Delta}{2} \quad \text{and} \quad R_n(\pi, v_{\mu'}) \geq \mathbb{P}_{\mu'}(T_1(n) > n/2) \frac{n\Delta}{2}$$

然后，应用前一章中的 Bretagnolle–Huber 不等式（定理 14.2），

$$\begin{aligned} R_n(\pi, v_\mu) + R_n(\pi, v_{\mu'}) &> \frac{n\Delta}{2} (\mathbb{P}_\mu(T_1(n) \leq n/2) + \mathbb{P}_{\mu'}(T_1(n) > n/2)) \\ &\geq \frac{n\Delta}{4} \exp(-D(\mathbb{P}_\mu, \mathbb{P}_{\mu'})) \end{aligned}$$

它保持在上限  $D(\mathbb{P}_\mu, \mathbb{P}_{\mu'})$ 。为此，我们使用引理 15.1 和  $\mu, \mu'$  获得

$$D(\mathbb{P}_\mu, \mathbb{P}_{\mu'}) = \mathbb{E}_\mu[T_i(n)] D(\mathcal{N}(0, 1), \mathcal{N}(2\Delta, 1)) = \mathbb{E}_\mu[T_i(n)] \frac{(2\Delta)^2}{2} \leq \frac{2n\Delta^2}{k-1}$$

将此插入之前的显示，我们发现

$$R_n(\pi, v_\mu) + R_n(\pi, v_{\mu'}) \geq \frac{n\Delta}{4} \exp\left(-\frac{2n\Delta^2}{k-1}\right)$$

通过选择  $\Delta = \sqrt{(k-1)/4n} \leq 1/2$ ，其中不等式遵循定理陈述中的假设。最后一步是下

限  $\exp(-1/2)$  和  $2 \max(a, b) \geq a + b$ 。

我们鼓励读者阅读练习 15.2 中概述的替代证明，其中采取了稍微不同的路径。

### 15.3 注记

1、我们使用高斯噪声模型，因为 KL 发散度在这种情况下很容易计算但我们实际使用的是  $D(P_i, P'_i) = O((\mu_i - \mu'_i)^2)$ ，当平均值之间的差距  $\Delta = \mu_i - \mu'_i$  很小。虽然并非所有情况都是这样，但通常情况下确实如此。为什么呢？令  $\{P_\mu : \mu \in \mathbb{R}\}$  是  $\Omega$  上的一些参数分布族，并假设分布  $p_\mu$  具有平均  $\mu$ 。假设密度是两次可微的，并且所有的都是积分和导数可以交换（几乎总是这样），我们可以用一个关于  $\mu$  的泰勒展开式来表示

$$\begin{aligned}
D(P_\mu, P_{\mu+\Delta}) &\approx \frac{\partial}{\partial \Delta} D(P_\mu, P_{\mu+\Delta}) \Big|_{\Delta=0} \Delta + \frac{1}{2} \frac{\partial^2}{\partial \Delta^2} D(P_\mu, P_{\mu+\Delta}) \Big|_{\Delta=0} \Delta^2 \\
&= \frac{\partial}{\partial \Delta} \int_{\Omega} \log \left( \frac{dP_\mu}{dP_{\mu+\Delta}} \right) dP_\mu \Big|_{\Delta=0} \Delta + \frac{1}{2} I(\mu) \Delta^2 \\
&= - \int_{\Omega} \frac{\partial}{\partial \Delta} \log \left( \frac{dP_{\mu+\Delta}}{dP_\mu} \right) \Big|_{\Delta=0} dP_\mu \Delta + \frac{1}{2} I(\mu) \Delta^2 \\
&= - \int_{\Omega} \frac{\partial}{\partial \Delta} \frac{dP_{\mu+\Delta}}{dP_\mu} \Big|_{\Delta=0} dP_\mu \Delta + \frac{1}{2} I(\mu) \Delta^2 \\
&= - \frac{\partial}{\partial \Delta} \int_{\Omega} \frac{dP_{\mu+\Delta}}{dP_\mu} dP_\mu \Big|_{\Delta=0} \Delta + \frac{1}{2} I(\mu) \Delta^2 \\
&= - \frac{\partial}{\partial \Delta} \int_{\Omega} dP_{\mu+\Delta} \Big|_{\Delta=0} \Delta + \frac{1}{2} I(\mu) \Delta^2 \\
&= \frac{1}{2} I(\mu) \Delta^2
\end{aligned}$$

其中，第二行中引入的  $I_{(\mu)}$  称为族的 Fisher 信息  $P(\mu)_\mu$  在。注意，如果  $\lambda$  是  $\Delta small$  的

( $P_{\mu+\Delta}$ ) 的常用主要度量，则  $dP_{\mu+\Delta} = p_{\mu+\Delta} d\lambda$  我们可以写

$$I(\mu) = - \int \frac{\partial^2}{\partial \Delta^2} \log p_{\mu+\Delta} \Big|_{\Delta=0} p_\mu d\lambda$$

这是小学课文中通常给出的形式。这一切的结果是  $D(P_\mu, P_{\mu+\Delta})$ ，因为  $\Delta small$  实际上是  $\Delta$  的

二次方，通过提供的缩放  $I_{(\mu)}$ ，因此，最糟糕的遗憾总是  $O\sqrt{nk}$ ，提供所考虑的分配类别足够充分，也不太奇怪。

2、我们现在已经显示了一个下限  $O\sqrt{nk}$ ，虽然许多上限是  $O\sqrt{\log(n)}$ 。这并不矛盾，

因为对数界限取决于次优间隙的倒数，次优间隙可能非常大。

3、我们的下限仅为  $n \geq k-1$ 。在练习 15.3 中，我们要求您理解当  $n < k-1$  时，有一个赌博机

$$R_n \geq \frac{n(2k-n-1)}{2k} > \frac{n}{2}$$

4、用于证明定理 15.2 的方法可以看作是对统计中的 Le Cam 方法。回想一下，等式 (15.2) 对于任意  $\mu$  和  $\mu'$

$$\inf_{\pi} \sup_v R_n(\pi, v) \geq \frac{n\Delta}{8} \exp(-D(\mathbb{P}_\mu, \mathbb{P}_{\mu'}))$$

为了解释 Le Cam 的方法，我们需要一点符号。令  $\chi$  为结果空间， $\mathcal{P}$  为  $\chi$  上的一系列

措施， $\theta: \mathcal{P} \rightarrow \Theta$ ，其中  $(\Theta, d)$  是一个度量空间。估计器是一个函数  $\hat{\theta}: \mathcal{X}^n \rightarrow \Theta$ 。Le-Cam 方法用于证明期望误差的极大极小下界估计量

$$\inf_{P \in \mathcal{P}} \sup_{X_1, \dots, X_n \sim P^n} \mathbb{E} \left[ d(\hat{\theta}(X_1, \dots, X_n), \theta(P)) \right]$$

这个方法是用来选择  $P_0, P_1 \in \mathcal{P}$  去最大化  $d(\theta(P_0), \theta(P_1)) \exp(-nD(P_0, P_1))$ ，在任意

$P_0, P_1 \in \mathcal{P}$  的基础上

$$Eq.(15.3) \geq \frac{\Delta}{8} \exp(-nD(P_0, P_1))$$

式中， $\Delta = d(\theta(P_0), \theta(P_1))$ 。与赌博机下限相比，有两个区别：(1) 我们处理顺序设置；

(2) 选择  $P_0$  后，我们选择  $P_1$  的方式取决于关于算法。这提供了一个非常需要的额外提升，如果没有它，该方法将是无法捕捉  $\mathcal{P}$  的特征如何反映在极大极小风险中（或遗憾，在我们的案例）

## 15.4 文献综述

我们知道的第一个关于下界的工作是 Vogel [1960] 对双臂伯努利赌博机的非常精确的极大极小分析。Bubeck 等人[2013b]首次将 Bretagnolle–Huber 不等式（定理 14.2）用于赌博机。正如注记中所述，利用这个不等式证明下界，在统计学中被称为 Le Cam 方法[Le Cam, 1973]。定理 15.2 的证明采用了与 Gerchinovitz 和 Lattimore [2016] 相同的思路，而练习题 15.2 中的另一种证明本质上是由 Auer 等人[1995]提出的，他们分析了奖励是伯努利的更困难的情况（见练习题 15.4）。Yu [1997] 描述了一些替代 Le Cam 方法的被动、统计设定。这些备选方案可以（而且经常可以）适应序列设置。

## 第二章 关键证明过程公式推导

### 15.1 公式推导

缩写  $\hat{\theta} = \hat{\theta}(X_1, \dots, X_n)$ ，同时令  $R(P) = \mathbb{E}_P[d(\hat{\theta}, P)]$ 。通过三角不等式得

$$d(\hat{\theta}, P_0) + d(\hat{\theta}, P_1) \geq d(P_0, P_1) = \Delta$$

令  $E = \{d(\hat{\theta}, P_0) \leq \Delta/2\}$  在  $E^c$  上认为  $d(\hat{\theta}, P_0) \geq \Delta/2$ ，在  $E$  上认为

$$d(\hat{\theta}, P_1) \geq \Delta - d(\hat{\theta}, P_0) \geq \Delta/2$$

$$R(P_0)+R(P_1)\geq \frac{\Delta}{2}\Big(P_0\big(E^c\big)+P_1(E)\Big)\geq \frac{\Delta}{4}\exp\big(-D(P_0,P_1)\big)$$

结果如下，因为  $\max\{a,b\} \geq (a+b)/2$