清 华 大 学

综 合 论 文 训 练

题目：关于先知不等式以及最优停时问题中的截断策略的研究

系　　别：交叉信息研究院

专　　业：计算机科学与技术

姓　　名：谭子涵

指导教师：王振波副教授

2015年 6月

关于学位论文使用授权的说明

本人完全了解清华大学有关保留、使用学位论文的规定，即：学校有权保留学位论文的复印件，允许该论文被查阅和借阅；学校可以公布该论文的全部或部分内容，可以采用影印、缩印或其他复制手段保存该论文。

(涉密的学位论文在解密后应遵守此规定)

签　名：_____ 导　师　签　名：_____ 日　　期：_____

中文摘要

最优停时问题是随机优化研究的广泛研究中一个很重要的问题：在一场多轮博弈中，一边获得对于未知或者已知分布的知识，一边又同时需要做出决断选择一个近似最优的策略以优化自己的收益。这类型的问题与在线学习和在线优化等计算机领域科学，运筹学领域研究也有紧密的联系。

本文主要对于随机优化中的最优停时问题做出了理论方面研究与讨论，以先知不等式问题入手，对于一般性分布，讨论了特殊情况下先知的优势比例极限。并且在相应问题中详尽分析了最优停时策略为阶段策略，证明了截断策略的一般性以及成立条件，并且对于不同轮之间的截断值的性质做出了分析。最后，对于随机优化中的特殊分布进行讨论，并分析了在特殊分布之下计算期望收益的算法复杂性与近似性。

关键词：最优停时；截断策略；先知不等式；随机优化

ABSTRACT

The optimal stopping time has been an important problem in stochastic optimization for long. In a multi-stage game, the player is required to learn information from the (known or unknown) latent distribution as well as optimize his utility. Such problem has concrete connection with online learning, a modern topic lying in the intersection between theoretical computer science and operation search.

This thesis mainly focused on the theoretical part of optimal stopping time. Beginning with the setting of prophet inequality problem, we explored the advantage ratio in special cases. Then we studied the pattern of optimal stopping strategy in such type of game and found that threshold strategy can be proved optimal. We then proved some property of threshold values in a threshold strategy. Finally we analyze the computational complexity and approximability of the optimal expected reward with special types of distribution.

Key words: Stopping time; Threshold strategy; Prophet inequality; Stochastic optimization

目 录

第 1 章　引言

随机优化是近年来兴起的一个领域，其本质是在于多轮重复博弈中，在对于未知的概率分布获取信息与优化博弈的自身收益这两个优化目标之中达到一个平衡。随机优化由于大数据时代的到来而显得更加吸引人，有趣而有深度的问题不断地涌现，比如[1] 文章中首次提出的多选择赌博机问题，由于一开始游戏参与者并不知道每一个赌博选择的收益分布，参与者只能够通过不断地选择从而完成对于位置分布的取样，然后确定出最佳的选择。但是在不断取样的过程中游戏收益又会受到损失，如何在两者之间达到平衡就成为一个值得研究的问题。另一个有名的问题是秘书问题，其目标就是在对于未来的应聘秘书的质量没有全部知识的情况下选定一个聘选策略使得最终雇佣的秘书能力值在某种期望成都之下最佳，具体可见Freeman写的详尽的survey[3].

研究者们在研究这个问题的时候一般会做出一些假设，例如：秘书们是根据一个完全随机的顺序来应聘的，或者说，每一位秘书带给公司的效用值服从某一个特定的分布，等等。假设的种类多种多样，求解问题的难度也参差不齐，技巧也变化万千。

我们在本文中研究的便是随机优化中的一种特殊情形，整个问题来源于先知不等式的问题设定，来自[2] ,事实上我们的问题来自上述综述的最后提出的open problems。具体来说，我们的问题可以陈述如下。

给定一族随机变量列$C = \{\mathbf{X} \mid \mathbf{X} = (X_1, X_2, \cdots, X_n)\}$，找寻一个对于$C$中的序列普适的比较“随机变量的最大值期望”和“最优停时策略可以获得的期望”的不等式，定义具体描述如下：

我们令M表示随机变量列中最大值的期望，

$$M(\mathbf{X}) = \mathbb{E}[\sup_i X_i] \tag{1-1}$$

同时令V表示最佳停时可以得到的随机变量值的期望（$\Gamma = \Gamma(X)$ 表示停止时间的集合），即

$$V(\mathbf{X}) = \sup_{t\in\Gamma} \mathbb{E}[X_t] \tag{1-2}$$

这个问题在现实中可以表述为一种特殊的秘书问题：我们需要聘用一个秘

书，有100（$n = 100$）个同一个学校培养出来的学生来应聘这个秘书岗位,我们看得到每一个人的简历所以可以对于他们的价值有一个提前认知（知道每一个X_i的分布），但是我们不通过面试的话就不能够知道这个值。于是我们按照他们来的顺序对他们一一进行面试，在面试完之后我们要立刻决定要不要聘用这个人，如果要的话我们就没有机会再面试后面的人，如果不要的话我们再也没有机会在未来重新聘用这个人（根据停时的定义）。于是我们在和先知较量（先知是一个了解所有人的真实值的个体，因而它永远可以做出最优的选择但我们不能），这就是M和V的比较。

我们寻求一个比较M和V的不等式，一个著名的例子是[2]：当C包含全部只取正值且相互独立的随机变量列的时候（即，$\forall \mathbf{X} \in C$, 如果$i \neq j$ 那么X_i和X_j是相互独立的，并且C包含所有这样的随机变量列），我们有：

$$V \leq M \leq 2V$$

并且2被证明是不可改进的。

在本文中，我们研究上述问题的两个变形，即改变目标函数，我们的目标不再是聘用某一个应聘者，而是某种累积函数，具体如下：

问题 1.1：（时间平均）

如果$Y_1, \cdots Y_n$是在$[0, 1]$中取值的独立同分布随机变量，令$X_j = \frac{1}{j}\sum_{i=1}^{j} Y_i$, 那么$M$与$V$之间的比较关系如何？换言之，找到最小的$k$使得下式成立:

$$V \leq M \leq kV$$

问题 1.2：（递减累积）

如果$Y_1, \cdots Y_n$是在$[0, 1]$中取值的独立同分布随机变量，令$0 < \alpha < 1$为递减因子并令$X_j = \sum_{i=1}^{j} \alpha^{j-i} Y_i$, 那么$M$与$V$之间的比较关系如何？换言之，找到最小的$k$使得下式成立:

$$V \leq M \leq kV$$

对于这个问题，我们在本文中仅对于$n = 2$给出了完全解答，我们在下文也会看到问题的困难程度在$n = 3$的时候急剧增加。对于一般的n，虽然我们没有能够算出这个最优k值，但是我们对于最佳策略进行了深入地分析，并证明了：在这样两种问题中最佳策略一定是具有“截断策略”（我们后文会给出精确

定义）的形式。所以求最佳策略的问题就转化为求截断值的问题。对于这两个问题的截断值我们也有详尽的讨论。除此之外，我们证明了更一般性的一个定理，说明了更广泛一类问题中最佳策略都是“截断策略”。最后，对于随机变量取特殊的分布的情况，我们给出了一些分析，并且讨论了在这些特定情况下求出最佳策略和期望收益值的计算复杂性。这些结论将在后续的三个章节中分别讨论。

第 2 章　$n = 2$时的先知不等式

在本节中我们分析随机变量序列只有两个元素时的情形，在下文我们使用$p(\cdot)$表示独立同分布随机变量Y_i分布的概率密度函数，在本节中我们假设$p(\cdot)$是连续函数。

2.1　时间平均情况

我们先证明在$n = 2$的情况下，决策者的最优策略（即：最佳停机时间）是下述的“截断策略”。

引理 2.1：　$n = 2$时，最佳策略应该具有如下的形式：先看Y_1,如果Y_1的值大于某一个阈值T(T由$p(\cdot)$确定，事实上$T = E[Y]$),那么我们停止并且接受Y_1，否则继续看Y_2并且接受Y_1, Y_2。

证明 事实上，在$n = 2$的情况下，任意的策略可以被写成一个(可以是随机的)函数$f : [0, 1] \to \{0, 1\}$,这里自变量就是Y_1的值，而因变量就是是否继续查看下一个Y_2的值。我们这里说函数可以是随机的，因为一个策略可以是随机的。在这里我们不妨设$f(x) = 1$表示策略会继续查看并接受Y_2，反之$f(x) = 0$。

事实上，一个随机的函数可以被写成非随机的函数的概率组合，而随机策略的收益也就自然是这些非随机策略的收益的线性组合，由于我们希望得到一个收益最大化的策略，我们不妨只考虑确定性的策略。

令$E = \mathbb{E}[Y]$, 我们有:

$$E[\text{f的收益}] = \int_{y:f(y)=0} yp(y)dy + \int_{y:f(y)=1} \frac{1}{2}(y + E)p(y)dy$$

从上式我们不难看出，如果$y \geq \frac{1}{2}(y + E)$, 我们就应该令$f(y) = 0$，否则令$f(y) = 1$，于是命题得证。

□

从上面的引理我们可以证明下面的定理：

定理 2.1： 如果 Y_1, Y_2 是在 $[0,1]$ 上取值的独立同分布随机变量, 对于 $j = 1, 2$, 令 $X_j = \frac{1}{j}\sum_{i=1}^{j} Y_i$, 那么按照1,2式定义的$M, V$,对于$C$为只包含具有上述特征的$X_1, X_2$这样长度为2的随机变量列类，我们有

$$V \le M \le (6 - 2\sqrt{6})V$$

证明 从引理1的结论，我们可以如下计算出V的表达式

$$V = \int_E^1 yp(y)dy + \int_0^E \frac{1}{2}(y+E)p(y)dy = \frac{1}{2}E + \frac{1}{2}\int_E^1 yp(y)dy + \frac{1}{2}E\int_0^E p(y)dy$$

对于先知者，容易发现如果$y_1 \ge y_2$那么它应该停在Y_1，反之则继续查看Y_2，由于Y_1,Y_2是独立同分布的，不难证明$\Pr[y_1 \ge y_2] = \frac{1}{2}$。于是对于$M$,我们有下面的表达式：

$$M = \frac{1}{2}E + \frac{1}{2}\mathbb{E}[\max\{y_1, y_2\}]$$

我们使用下述的来自[2] 的引理。

引理 2.2： (极化引理)

设 X 是任意可积的随机变量，$-\infty < a < b < +\infty$, 令$(X)_a^b$ 是一个满足如果$X \notin [a,b]$，则$(X)_a^b = X$, 以概率$(b-a)^{-1}\int_{X\in[a,b]}(b-X)dP(x)$取 $(X)_a^b = a$，同时以概率 $(b-a)^{-1}\int_{X\in[a,b]}(X-a)dP(x)$取$(X)_a^b = b$ 的随机变量, 这样定义的 $(X)_a^b$ 称作 X 在区间 $[a,b]$上的极化,它满足下面两条性质：

(1)$\mathbb{E}[X] = \mathbb{E}[(X)_a^b]$.

(2)对于任意独立于X和$(X)_a^b$的随机变量Z, 有$\mathbb{E}[\max\{X,Z\}] \le \mathbb{E}[\max\{(X)_a^b, Z\}]$.

由于我们想要找到$\sup_{p(\cdot)}\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$。我们观察到$V$依赖于三个量：$E$, $\int_E^1 yp(y)dy$ 和 $\int_0^E p(y)dy$，所以我们只需要考虑极大化$\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$的$p(\cdot)$(当它和其他的概率密度函数$p'(\cdot)$相比具有同样的$E$, $\int_E^1 yp(y)dy$ 和 $\int_0^E p(y)dy$时)。

根据极化引理，这样的$p(\cdot)$ 应该具有“离散”的形式（注意到这样的“离散”形式并不满足$p(\cdot)$的连续性，但是可以由连续的$p(\cdot)$序列来逼近，所以当我们关注最大化比例$\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$的时候，我们可以考虑这样的离散型概率分布）

$$p(0) = a; p(E-\epsilon) = b; p(E+\epsilon) = c; p(1) = d; a+b+c+d = 1; (b+c)E + d = E;$$

具体来说，我们可以对Y做[0,E]区间上的极化，然后再对得到的随机变量做$[E,1]$区间上的极化，得到上面的形式，根据极化定理我们知道上述操作只会增加比例$\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$。

所以我们只需要找到最佳的a,b,c,d使得比例被最大化。

鉴于 $a+b+c+d=1$ 以及 $(b+c)E+d=E$, 我们有 $E=\frac{d}{a+d}$，我们继续做下列计算

$$\mathbb{E}[\max Y_1,Y_2]=[1-(1-d)(1-d)]\cdot 1+[2(b+c)(1-d)-(b+c)^2]\cdot E=(2-d)d+(1-a-d)(1-d+a)E$$

$$V=\frac{1}{2}E+\frac{1}{2}(cE+d)+\frac{1}{2}E(a+b)=\frac{1}{2}E(2-d)+\frac{1}{2}d$$

$$\text{ratio}=\frac{X}{V}=\frac{\frac{1}{2}E+\frac{1}{2}((2-d)d+(1-a-d)(1-d+a)E)}{\frac{1}{2}E(2-d)+\frac{1}{2}d}=1+\frac{a(1-a-d)}{2+a}$$

我们只需在$a,d\geq 0$ 与 $a+d\leq 1$的条件下接着极大化 $\frac{a(1-a-d)}{2+a}$. 显然我们应该设定 $d=0$, 做变量转化$t=a+2$,我们得到

$$\frac{a(1-a-d)}{2+a}=\frac{-t^2+5t-6}{t}\leq 5-(t+\frac{6}{t})\leq 5-2\sqrt{6}$$

另一方面，由于上述不等式放缩每一步都可以同时取等或者在极限过程中取等，我们可以合适的设定 a,b,c,d使得比例$\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$ 无限的接近 $6-2\sqrt{6}$，这让我们完成了定理的证明。

□

我们注意到，上述方法对于$n\geq 3$的情况并不能够奏效，根本原因在于M不能够写成一个简单的表达式（表达式中或出现累积概率分布反函数等项），因此我们很难直接应用计划定理来把问题进行有效地简化。

对于一般的n，我们没有能够给出一般的结论，但是我们有下面的分析，从而引领我们做出本小节末的猜想。根据弱大数定律，我们知道当n趋向于无穷大的时候，X_n得概率分布高度集中于E的附近。所以，如果一个策略以一定的概率允许查看很多的Y_i的话，在这种情况下它的收益一定基本上等于E。比如，下面的合理的策略的期望收益基本上就是E：

如果$X_i\leq E$,那么继续查看Y_{i+1},一旦看到当前的$X_i>E$，停止。

另一方面，我们可以合理的相信，当我们允许的随机变量序列更长的时候，先知对于我们具有的优势就更少，因为当我们已经接受了当前的$Y_1,\cdots Y_t$而

准备继续接受Y_{t+1}的时候，我们的收益中$Y_1, \cdots Y_t$已经占有了很大的权重，而不像我们刚刚接受Y_1准备去查看Y_2时候的收益变化那么大。所以通常来说，先知的优势主要在于在短短前几个随机变量内做出我们不能做出的最优决策。所以我们有如下猜测：

猜想 2.1： 如果$Y_1, \cdots Y_n$是在$[0,1]$中取值的独立同分布随机变量，令$X_j = \frac{1}{j}\sum_{i=1}^{j} Y_i$, 那么我们有

$$V \leq M \leq (6 - 2\sqrt{6})V$$

2.2 递减累积情况

类似于时间平均情况，在本节中我们仍然先证明最优策略应该是一种截断策略，然后我们通过最优化参数来计算出k值。

引理 2.3： $n = 2$时，最佳策略应该具有如下的形式：先看Y_1,如果Y_1的值大于某一个阈值T(T由$p(\cdot)$确定，事实上$T = \frac{E[Y]}{1-\alpha}$), 那么我们停止并且接受$Y_1$，否则继续看$Y_2$并且接受$Y_2$。

证明 令$E = \mathbb{E}[Y]$, 类似于前面的引理的证明，我们有:

$$E[\text{f的收益}] = \int_{y:f(y)=0} yp(y)dy + \int_{y:f(y)=1} (\alpha y + E)p(y)dy$$

从上式我们不难看出，如果$y \geq (\alpha y + E)$, 我们就应该令$f(y) = 0$，否则令$f(y) = 1$，于是命题得证。

□

我们可以应用上述引理和之前提到的计划引理来证明下面的$n = 2$结论：

定理 2.2： 如果 Y_1, Y_2 是在 $[0,1]$ 上取值的独立同分布随机变量, 对于 $j = 1, 2$, 令 $X_j = \sum_{i=1}^{j} \alpha^{j-i} Y_i$, 那么按照1,2式定义的$M, V$,对于$C$为只包含具有上述特征的$X_1, X_2$这样长度为2的随机变量列类，我们有

$$V \leq M \leq \frac{2}{1+\alpha} V$$

证明

从引理1的结论，我们可以如下计算出V的表达式

$$V = \int_{\frac{E}{1-\alpha}}^{1} yp(y)dy + \int_{0}^{\frac{E}{1-\alpha}} (\alpha y + E)p(y)dy = \alpha \int_{0}^{\frac{E}{1-\alpha}} yp(y)dy + E \int_{0}^{\frac{E}{1-\alpha}} p(y)dy + \int_{\frac{E}{1-\alpha}}^{1} yp(y)dy$$

对于先知者，容易发现如果$y_1 \geq y_2$那么它应该停在Y_1，反之则继续查看Y_2，于是对于M,我们有下面的表达式：

$$M = E[\max\{\alpha Y_1 + Y_2, Y_1\}] = E[\alpha Y_1] + E[\max\{Y_2, (1-\alpha)Y_1\}]$$

由于我们想要找到$\sup_{p(\cdot)} \frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$。我们观察到$V$依赖于三个量：$E$, $\int_{\frac{E}{1-\alpha}}^{1} yp(y)dy$ 和 $\int_{0}^{\frac{E}{1-\alpha}} p(y)dy$，所以我们只需要考虑极大化$\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$的$p(\cdot)$(当它和其他的概率密度函数$p'(\cdot)$相比具有同样的$E$, $\int_{\frac{E}{1-\alpha}}^{1} yp(y)dy$ 和 $\int_{0}^{\frac{E}{1-\alpha}} p(y)dy$时)。

根据极化引理，这样的$p(\cdot)$ 应该具有“离散”的形式（注意到这样的“离散”形式并不满足$p(\cdot)$的连续性，但是可以由连续的$p()$序列来逼近，所以当我们关注最大化比例$\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$的时候，我们可以考虑这样的离散型概率分布）

如果 $\frac{E}{1-\alpha} < 1$, 那么

$$p(0) = a; p(\frac{E}{1-\alpha} - \epsilon) = b; p(\frac{E}{1-\alpha} + \epsilon) = c; p(1) = d;$$

$$a + b + c + d = 1; (b + c)\frac{E}{1-\alpha} + d = E;$$

如果 $\frac{E}{1-\alpha} \geq 1$, 那么

$$p(0) = a, p(1) = 1 - a$$

对于第一种情况，具体来说，我们可以对Y做[0,E]区间上的极化，然后再对得到的随机变量做$[E, 1]$区间上的极化，得到上面的形式，根据极化定理我们知道上述操作只会增加比例$\frac{M(Y_1,Y_2)}{V(Y_1,Y_2)}$。

所以我们只需要找到最佳的a, b, c, d使得比例被最大化。

鉴于 $a + b + c + d = 1$ 以及 $(b + c)E + d = E$, 我们有 $E = \frac{d}{a+d}$，我们继续分三种情况做下列计算：

情况1 $\frac{E}{1-\alpha} \geq 1$

最佳策略永远是连续查看 Y_1 和 Y_2, 此时 $E = (1 - a)$.

$$V = \int_0^1 (\alpha y + E)p(y)dy = (1+\alpha)E$$

$$M = a^2 \cdot 0 + 2(1-a)a \cdot 1 + (1-a)^2 \cdot (1+\alpha) = (1-a)(1+\alpha)E + 2aE$$

$$\frac{M}{V} = \frac{(1-\alpha)a + (1+\alpha)}{1-\alpha}$$

我们可以令 $a \to 1$ ，从而 $\frac{M}{V} \to \frac{2}{1+\alpha}$.

情况2 $(1-\alpha) < \frac{E}{1-\alpha} < 1$

$$p(0) = a; p(\frac{E}{1-\alpha} - \epsilon) = b; p(\frac{E}{1-\alpha} + \epsilon) = c; p(1) = d; a+b+c+d = 1; (b+c)\frac{E}{1-\alpha} + d = E;$$

由于 $a+b+c+d = 1$ 并且 $(b+c)\frac{E}{1-\alpha} + d = E$, 我们得到 $E = \frac{d(1-\alpha)}{a+d-\alpha}$, 以及 $a > \alpha$.

$$V = \alpha E + (a+b)E + (1-\alpha)(c \cdot \frac{E}{1-\alpha} + d) = (1-\alpha)d\frac{1+a}{a+d-\alpha}$$

$$M = \alpha E + E[\max\{Y_1, (1-\alpha)Y_1\}] = \alpha E + d + (1-a-d)\frac{E}{1-\alpha} + a(1-\alpha)E = \frac{(1-\alpha)d}{a+d-\alpha}(1+\alpha+(1-\alpha)a)$$

$$\frac{M}{V} = 1 - \alpha + \frac{2\alpha}{1+a}$$

令 $a \to \alpha$ 则比例趋近于 $1 - \alpha + \frac{2\alpha}{1+\alpha}$.

情况3 $\frac{E}{1-\alpha} \leq 1 - \alpha$

$$V = \alpha E + (a+b)E + (1-\alpha)(c \cdot \frac{E}{1-\alpha} + d) = (1-\alpha)d\frac{1+a}{a+d-\alpha}$$

$$M = \alpha E + E[\max\{Y_1, (1-\alpha)Y_1\}] = \alpha E + d + (1-a-d)(\frac{E}{1-\alpha}(1-d) + d(1-\alpha)) + a(1-\alpha)E$$

通过类似的计算，我们得到的最优比例未能优于前两种情况。

比较三种情况的最优比例，由于看到$\frac{2}{1+\alpha} > 1 - \alpha + \frac{2\alpha}{1+\alpha}$, 我们可以知道，最优的$k$值为 $\frac{2}{1+\alpha}$.

□

第 3 章　最优停时的截断策略

3.1　截断策略

由于对于一般的n，准确计算出上述两个问题的最佳比例只是困难的，我们转而分析操作者的最优策略的形态。在$n = 2$的时候我们通过写出期望收益的准确表达式而简单的证明了最佳策略是一种“截断策略”，即看到第一个值大于等于某一个数的时候，停止；小于这个数的时候，继续看第二个数。在本节中我们将对于一般的n证明类似的结论.

首先，我们需要给出“阶段策略”的准确数学定义。类似于前一节的陈述，一个策略可以被定义为一个（可以是随机的）函数$f : [0,1]^n \to [n]$，其中$f(y_1, \cdots, y_n) = t$代表当操作者看到$y_1, \cdots, y_n$的时候，会停在y_t而不再查看后面的值。由于操纵者没有先知，在序列$y_1, \cdots, y_n$之下停在y_t意味着不论在t轮之后随机变量到底是什么值，它都会停在y_t，数学的说法便是

$$\text{如果} f(y_1, \cdots, y_n) = t, \text{ 那么 } \forall y'_{t+1}, \cdots, y'_n \in [0,1], f(y_1, \cdots, y_t, y'_{t+1}, \cdots, y'_n) = t$$

因此我们也可以换另外一种方式来定义一个策略。我们定义 $g_t : [0,1]^t \to \{0,1\}$是第$i$步的决策函数, $g_t(y_1 \cdots y_t) = 0$ 意味着当决策者看到（前提是他没有在之前停下来）序列$y_1 \cdots y_t$的时候会选择停在y_t, $g_t(y_1 \cdots y_t) = 1$ 意味着当决策者看到（前提是他没有在之前停下来）序列$y_1 \cdots y_t$的时候会选择停继续查看y_{t+1}. 注意到 g_t 的取值仅取决于整个序列的前t个变量值 $(y_1, \cdots, y_t)$ 并且$g_t(y_1 \cdots y_t) = 1$蕴含$\forall 0 < i < t, g_i(y_1, \cdots, y_i) = 1$ （同时我们约定，如果$\exists 0 < i < t, g_i(y_1, \cdots, y_i) = 0$则$g_t(y_1 \cdots y_t) = 0$）。不难证明，决策函数 f 等价于函数序列 $\{g_i\}_{i=1}^n$.

因为我们感兴趣的是可以将期望回报最大化的决策，鉴于我们看到随机决策的期望回报等于确定性决策的回报的加权求和，我们可以不考虑非确定性策略。

因此，截断策略可以用如下的方式定义。决策函数f被称为一个阶段策略，如果$\forall (y_1, \cdots, y_t)$只要它满足$\forall 0 < i < t, g_i(y_1, \cdots, y_i) = 1$, 那么就存在一个整数$T_t$, 使得$g_t(y_1, \cdots, y_t) = \mathbf{1}_{y_t \geq T_t}$,这里 $\mathbf{1}_{y_t \geq T_t}$ 代表事件$y_t \geq T_t$的指示函数，即如果事

件发生则取值1，否则取值0.

对于原问题的时间平均和递减累计两种不同的目标函数，我们不加证明的陈述下面的定理。

定理 3.1： 在时间平均目标函数和递减累计目标函数两种情形中，最优决策f一定是只依赖于$p(\cdot)$的截断策略。

3.2 一致截断策略

事实上，我们可以证明关于最优策略的更强的结论。首先我们加强定义“截断策略”的限制，得到一个更强的“一致阶段策略”的定义如下。

我们说一个截断策略f是一致的，如果存在正实数$T_1 \leq T_2 \leq \cdots \leq T_n$,使得$f(y_1, \cdots y_n) \geq t$ 当且仅当$\forall 0 < j < t, x_j = r(y_1, \cdots, y_t) \geq T_j$成立。这里$r(\cdot)$是收益函数，在时间平均情况中，$r(y_1, \cdots, y_j) = \frac{1}{j}\sum_{i=1}^{j} y_i$,在递减累计的情况中$r(y_1, \cdots, y_j) = \sum_{i=1}^{j} \alpha^{j-i} Y_i$.

对于时间平均和递减累计这两种问题，我们可以证明最优策略一定是某个一致截断策略，即我们不加证明的写出下述定理（事实上我们可以给出证明，但是非常冗长，因为这个证明需要从概率论方面做出最基础的严格定义，为保证论文可读性，我们略去这个证明）。

定理 3.2： 在时间平均目标函数和递减累计目标函数两种情形中，最优决策f一定是只依赖于$p(\cdot)$的一致截断策略。

3.3 对于时间平均情况截断值的分析

我们在证明目标函数是时间平均情况下，最优策略一定是一个一致截断策略，在本节中我们对于这些截断值做出分析，从而可以更好的刻画最佳策略的形态。本节中我们采用动态规划的方法。

假设共有n个独立同分布的随机变量，我们用$OPT(k, r)$来表示操作者在查看了前k个值，得到的当前收益为r的情况下，在后续选用最佳的策略能够获得的期望收益，这里$1 \leq k \leq n, r = \sum_{i=1}^{k} y_i$(为了下文叙述方便我们这里暂时采用和而不是平均)，我们有：

$$OPT(k, r) = \max\{\frac{r}{k}, \int_0^1 OPT(k+1, r+y)p(y)dy\}$$

我们考虑第二项：

$$\int_0^1 OPT(k+1, r+y)p(y)dy \geq \int_0^1 \frac{r+y}{k+1}p(y)dy = \frac{r+E}{k+1}$$

比较这个下界与$\frac{r}{k}$，我们知道：当$\frac{r}{k} \leq E$的时候，自然有$\int_0^1 OPT(k+1, r+y)p(y)dy \geq \frac{r+E}{k+1} \geq \frac{r}{k}$成立,也即操作者一定会接续查看下面的$y_{k+1}$值而不会停止，也就是说我们得到了截断值的一个下界：

$$T_k \geq k \cdot \mathbb{E}[Y]$$

通过另外一个角度的分析，我们可以得到一个关于截断值上界性质的结论，但并不是真正意义上的上界。

事实上，对于$0 < c < 1$,我们假设$p_c = \Pr[y \geq c]$，假设对于最优策略，当在在前k轮获得收益r，并且计划在下一轮看到不小于c 的值的时候它会选择停止，那么我们有：(令$E_c = \int_c^1 yp(y)dy$,我们称之为"高位期望"。)

$$\int_0^1 OPT(k+1, r+y)p(y)dy = \int_0^c OPT(k+1, r+y)p(y)dy + \int_c^1 OPT(k+1, r+y)p(y)dy \tag{3-1}$$

$$\leq (1-p_c)OPT(k+1, r+c) + p_c\frac{r}{k+1} + \frac{1}{k+1}\int_c^1 yp(y)dy \tag{3-2}$$

$$\leq (1-p_c)\frac{r+c}{k+1} + p_c\frac{r}{k+1} + \frac{1}{k+1}E_c \tag{3-3}$$

$$= \frac{r + c(1-p_c) + E_c}{k+1} \tag{3-4}$$

于是，如果$\frac{r+c(1-p_c)+E_c}{k+1} \leq \frac{k}{r}$，即$c(1-p_c) + E_c \leq \frac{r}{k}$的话，我们就会知道，最优策略一定不会在$k$轮拿到$r$的整体收益的情况下继续查看下一个随机变量的值。也就是说：如果$r + c > T_{k+1}$,并且$\frac{r}{k} \geq E_c + c(1-p_c)$,我们可以得到$r > T_k$。于是我们可以写出下面的不等式：

$$T_k \leq \min_c \max\{T_{k+1} - c, k(E_c + c(1-p_c))\}$$

这个不等式是非平凡的，虽然它并不是一个T_k的上界，但是它的一个分支结论给出了我们下述启示：取$c_k = T_{k+1} - T_k$,一般来说由于剩下的未探测的随机

变量变少，我们倾向于相信截断值的增量会减小，即$T_{k+1} - T_k$随k的增加而递减，也就是说上述不等式告诉我们

$$\frac{T_k}{k} \leq (E_{c_k} + c_k(1 - p_{c_k}))\}$$

通过简单的分析我们可以看到不等号右边是一个递减的序列，也即，我们得到一个$\frac{T_k}{k}$的随k增长越来越小的上界。

我们给出了T_k的上界与下界性质的结论，在本界的末尾，我们稍用定性分析来寻找T_k之间更深层次的关系，我们虽然到现在还没有办法证明下面的很强的结论，但是我们有理由相信它是对的：

猜想 3.1： （递减截断值）

$$T_1 \geq \frac{T_2}{2} \geq \frac{T_3}{3} \geq \cdots \geq \frac{T_{n-1}}{n-1} = \mathbb{E}[Y]$$

事实上我们可以这样分析，由于一开始，未知值的随机变量个数较多，因此我们停止的要求就会升高（不是足够满意的收益值是不能够阻止我们继续查看下面的值的），但是随着我们的查看不断进行，剩下的随机变量个数越来越少，我们能够满足的截断值就会不断下降（即，$\frac{T_k}{k}$不断减小），这就是上面的猜想所陈述的性质。

第 4 章　时间平均情况下对特殊概率分布下的分析

在本节中，我们在一些特殊的分布上，对原来的随机优化问题作出分析。事实证明，即便是在特殊的分布下，完全求解出T_k的值也是很难的，但是我们可以得到一些比一般情况下更强的结论。另一方面，我们虽然无法直接写出最优策略截断值表达式，但是我们可以给出算法计算它们并其分析复杂性。具体来说，我们将分别考虑下面几个情形：（1）概率分布是$B(p)$，即以概率p取值1,以概率$1-p$取值0;（2）$[0,1]$上的等距离散分布；（3）$[0,1]$上均匀分布。

4.1　$B(p)$分布

在这种特殊分布中，随机变量直取两种可能值$0,1$。我们的分析得到最优策略截断值以及预期收益的一些性质，但是并没有能够完全的解决整个问题。具体来说，我们令$OPT(k,r)$来表示操作者在查看了前k个值，得到的当前收益为r的情况下，在后续选用最佳的策略能够获得的期望收益，这里$1 \le k \le n$, $r = \sum_{i=1}^{k} y_i$(为了下文叙述方便我们这里暂时采用和而不是平均，在这里r的可能值是不超过k的整数)，我们有下面的结论：

定理 4.1：　如果$Y_1, \cdots Y_n$是在$\{0,1\}$中取值的独立同分布随机变量，令$X_j = \frac{1}{j}\sum_{i=1}^{j} Y_i$,我们有：

（1）当$t > (n-1)p$时，$OPT(k,t) = \frac{t}{k}$

（2）当$t > (n-2)p - (n-k+1)$时，$OPT(k,t) = \frac{t+(n-k)p}{n}$

证明 事实上我们需要用反向证明法来证明这个结论，我们首先考虑$OPT(n,t)$,毫无疑问这个时候操作者已经没有选择，故$\forall t, OPT(n,t) = \frac{t}{n}$.

考虑$OPT(n-1,t)$，这个时候操作者有两种选择，一种是继续查看下一个值，即以概率p得到最终受益$\frac{t+1}{n}$,以概率$1-p$ 得到最终受益$\frac{t}{n}$,这个期望是$\frac{t+p}{n}$;另一种选择是直接停止，这种选择的收益是$\frac{t}{n-1}$，比较两种选择，我们发现：当$\frac{t}{n-1} \ge p$时，操作者应该选择停止从而拿到$\frac{t}{n-1}$的收益，当$\frac{t}{n-1} \le p$ 时，操作者应该选择继续查看从而拿到$\frac{t+p}{n}$的收益,这一切满足定理中的描述。

假设对于$k > 1$以及任意的$t \le k$,$OPT(k,t)$满足定理的叙述，那么我们考虑$OPT(k-1,t)$的情况。假设$t > (n-1)p$成立，这个时候操作者有两种选择，

一种是继续查看下一个值，即以概率p得到最终受益$OPT(k,t+1)$,以概率$1-p$得到最终受益$OPT(k,t)$,这个期望是$p\cdot OPT(k,t+1)+(1-p)\cdot OPT(k,t)$; 另一种选择是直接停止，这种选择的收益是$\frac{t}{k-1}$，根据归纳假设，$t>(n-1)p$的时候，$t+1>(n-1)p$也成立，所以$OPT(k,t)=\frac{t}{k}$,$OPT(k,t+1)=\frac{t+1}{k}$,从而操作者选择继续的收益将是$\frac{t+p}{k}$,而这个值小于$\frac{t}{k-1}$，所以操作者此时应该选择停止。

另一方面，假设$t>(n-2)p-(n-k)$成立，这个时候操作者有两种选择，一种是继续查看下一个值，即以概率p得到最终受益$OPT(k,t+1)$,以概率$1-p$ 得到最终受益$OPT(k,t)$,这个期望是$p\cdot OPT(k,t+1)+(1-p)\cdot OPT(k,t)$; 另一种选择是直接停止，这种选择的收益是$\frac{t}{k-1}$，根据归纳假设，$t>(n-2)p-(n-k+2)$的时候，$t+1>(n-2)p-(n-k+1)$也成立，所以$OPT(k,t)=\frac{t+(n-k)p}{n}$,$OPT(k,t+1)=\frac{t+1+(n-k)p}{n}$,从而操作者选择继续的收益将是$\frac{t+(n-k)p}{n}$, 而这个值大于$\frac{t}{k-1}$，所以操作者此时应该选择继续查看从而得到高的期望收益值。

□

注意到，这个结论并不是很强，他只陈述了t比较大的时候和比较小的时候我们可以清晰地计算出操作者的最优策略以及期望收益，但是在t取中间值的时候并没有给出讨论，事实上，经过计算t取中间值的时候很难计算出具体的T_i值，或者说它的具体解析形式非常之复杂。

4.2 [0, 1]上的等距离散分布

上一届的$B(p)$分布事实上是本节希望讨论的等距离散分布的一种特殊情况。具体来说等距离散分布可以定义为：$\exists q\in\mathbb{N}$, 概率分布仅在$\{\frac{i}{q}\}_{0\le i\le q}$这些点处有离散的概率值并且加起来等于1，在其他地方没有概率密度。事实上，对于这类的概率分布，我们可以设计算法来求出它的期望收益和最优策略截断值。我们采用动态规划的算法。我们的结论可以被概括为下面的定理。

定理 4.2： 当n个独立同分布变量满足等距离散分布（间距为$\frac{1}{q}$）时,存在算法计算最优策略的期望收益值，算法复杂性为$O(n^2q^2)$.

证明 假设给定的分布为$p(Y=\frac{i}{q})=p_i$

令$OPT(k,r)$为操作者在查看了前k个值，得到的当前收益为r的情况下，在后续选用最佳的策略能够获得的期望收益（这里$r=\frac{t}{q}$），则我们可以得到下面的等式：

$$OPT(k,r) = \max\{\frac{r}{k}, \sum_{i=0}^{q} OPT(k+1, r+\frac{i}{q})\}$$

于是我们想要计算$OPT(0,0) = \sum_{i=0}^{q} OPT(1, \frac{i}{q})$,我们只需要计算出来所有的$OPT(k,r)$。事实上$k$总共有$n$种可能的取值，$r$总共有$nq$种可能的取值。我们从$k = n$的$OPT(k,r)$开始计算，之后逐渐减小$k$的值，我们可以看到，计算每一个值我们需要$O(q)$次计算，因此计算的总量不超过$O(n^2q^2)$.

□

事实上，这个定理给予我们近似处理一般分布的方法，那就是用等距离散分布来近似表示连续分布，比如，将区间$[0,1]$等分为若干段然后将原来的一般分布中每一段概率密度求积分之后转化到小段端点处的离散概率。这样的调整不会对于计算期望收益值有很大的变化，因而可以作为很好的近似算法,我们通过下面的分析给出这样的结论。

定理 4.3： 对于任意连续分布随机变量Y，随机变量个数n以及实现给定的$\epsilon > 0$,存在一个算法可以近似计算出最佳策略的期望收益，误差在ϵ之内.算法的复杂性是$O(\frac{n^2}{\epsilon^2})$.

证明 对于一个连续分布$p(y)$，另$q = \frac{1}{\epsilon}$我们将它转换成另外一个离散分布$\{p_i\}_{i=1}^{q}$，满足

$$p_i = \int_{\frac{i-1}{q}}^{\frac{i}{q}} p(y)dy$$

即，对于原来的分布进行了离散化等距分割，把每一段的概率已到了一个端点上。我们的算法就是之前定理中所叙述的动态规划算法，根据之前的定理我们知道它的复杂性是$O(n^2q^2) = O(\frac{n^2}{\epsilon^2})$.

下面我们证明这个算法的近似性。

我们依照之前的定义，令$OPT(k,r)$为在Y遵从原来的连续分布时，操作者在查看了前k个值，得到的当前收益为r的情况下(这里我们依然令r为所有随机变量值之和而不是平均)，在后续选用最佳的策略能够获得的期望收益，令$OPT'(k,r)$为在Y遵从新的等距离散分布时，操作者在查看了前k 个值，得到的当前收益为r 的情况下，在后续选用最佳的策略能够获得的期望收益。

希望证明我们算法计算最优策略下的期望收益和真实的最优策略期望收益相差不超过ϵ，我们相当于想要证明：

$$\int_0^1 OPT(1,y)p(y)dy+\epsilon \le \sum_{i=1}^{q} OPT'(1,\frac{1}{q})p_i \le \int_0^1 OPT(1,y)p(y)dy+\epsilon$$

事实上，左边的不等号是显然的，因为我们相当于构建了一个完全占优的新分布，因而得到的最优策略期望收益一定大于之前的最优策略期望收益。

对于右边的不等号，我们用数学归纳法证明$OPT'(k,r) \le OPT(k,r)+\frac{\epsilon}{k}$.

首先考虑$k=n$，我们有$OPT'(n,r)=OPT'(n,r)=\frac{r}{n}$，结论成立。

其次考虑$k=n-1$，我们有$OPT(n-1,r)=\max\{\int_0^1 \frac{r+y}{n}p(y)dy,\frac{r}{n-1}\}$,而 $OPT'(n-1,r)=\max\{\sum_{i=1}^{q}\frac{r+\frac{i}{q}}{q}p_i,\frac{r}{n-1}\}$，我们只需要证明：

$$\sum_{i=1}^{q}\frac{r+\frac{i}{n}}{q}p_i \le \int_0^1 \frac{r+y}{n}p(y)dy+\frac{\epsilon}{n-1}$$

而这可以由下式推出

$$\frac{r+\frac{i}{n}}{q}p_i \le \int_{\frac{i-1}{q}}^{\frac{i}{q}} \frac{r+y}{n}p(y)dy+\frac{\epsilon}{n}$$

此式是显然的。

我们假设命题在$k=m$的时候成立，考虑$k=m-1$的情况，我们仍然只需要证明

$$\sum_{i=1}^{q} OPT'(m,r+\frac{1}{q})p_i \le \int_0^1 OPT(m,r+y)p(y)dy+\frac{\epsilon}{m}$$

由于最优策略是阶段策略，我们知道存在一个c（事实上这个c是可能大于1或者小于0的，我们得证明中，在这两种情况下我们直接把它定为1 和0）使得当$y \ge c$ 的时候有$OPT(m,r+y)=\frac{r+y}{n}$ 成立，$y<c$ 的时候有$OPT(m,r+y)=\int_0^1 OPT(m+1,r+y)p(y)dy$ 成立,所以我们把上式左边也拆成两半，我们证明

$$\sum_{i=1}^{t} OPT'(m,r+\frac{1}{q})p_i+\sum_{i=1}^{t}\frac{r+\frac{i}{q}}{m} \le \int_0^{\frac{t}{q}} OPT(m,r+y)p(y)dy+\int_{\frac{t}{q}}^1 \frac{r+y}{m}p(y)dy+\frac{\epsilon}{m}$$

一方面,经简单计算我们有

$$\sum_{i=1}^{t}\frac{r+\frac{i}{q}}{m} \le \int_{\frac{t}{q}}^1 \frac{r+y}{m}p(y)dy+\frac{\epsilon}{m}(\sum_{i=t+1}^{q}p_i)$$

另一方面，通过归纳假设我们有

$$\sum_{i=1}^{t} OPT'(m, r+\frac{1}{q})p_i \le \int_0^{\frac{t}{q}} OPT(m, r+y)p(y)dy + \frac{\epsilon}{m+1}(\sum_{i=1}^{t} p_i)$$

综合这两个结论我们得到$k = m-1$的时候结论依然成立。故命题得证。

□

4.3 [0, 1]上均匀分布

在本节中，我们考虑Y的分布是$[0,1]$上的均匀分布。

我们沿用前面定义的策略函数$f : [0,1]^n \to [n]$。当随机变量$Y_1, \cdots, Y_n$分别取值$y_1 \cdots y_n$的时候(我们记$y = (y_1, \cdots, y_n)$)，我们用$r_f(y)$来代表策略在此观察之下的收益，即

$$r_f(y) = \frac{\sum_{i=1}^{f(y)} y_i}{f(y)}$$

下面我们计算 $\max_f \mathbb{E}_{Y_1,\cdots,Y_n} r_f(y)$。

在前面的讨论中我们知道，最优策略是一个一致截断策略，由截断值$T_1, \cdots, T_n$完全决定，当观察到第i个变量的取值的时候，我们只关注部分和$\sum_{j=1}^{i} y_i$是否超过了T_i,于是我们定义函数$OPT(k, r)$为操作者在查看了前k个值，得到的当前收益为r的情况下，在后续选用最佳的策略能够获得的期望收益，我们有：

$$OPT(k, r) = \begin{cases} \frac{r}{k}, & r \ge T_k\ ; \\ \mathbb{E}_{Y_{k+1}} OPT(k+1, r+Y_{k+1}), & r < T_i. \end{cases}$$

鉴于具体的截断值难于计算，我们在这里仅详细计算出$n = 3, 4$的情况。

当$n = 3$时,

$$OPT(3, r) = \frac{r}{3}$$

$$OPT(2, r) = \begin{cases} \frac{r}{2}, & r \ge 1\ ; \\ \int_r^{r+1} OPT(3, y)dy = \int_r^{r+1} \frac{y}{3}dy = \frac{2r+1}{6}, & r < 1. \end{cases}$$

于是

$$OPT(1,r)=\begin{cases} r, & r\geq T_1\ ;\\ \int_r^{r+1} OPT(2,y)dy, & r<T_1. \end{cases}$$

$$OPT(1,r)=\begin{cases} r, & r\geq T_1\ ;\\ \int_r^{1} \frac{2y+1}{6}dy+\int_1^{1+r} \frac{y}{2}dy=\frac{1}{12}r^2+\frac{1}{3}r+\frac{1}{3}, & r<T_1. \end{cases}$$

我们得到：T_1是方程 $r=\frac{1}{12}r^2+\frac{1}{3}r+\frac{1}{3}$的根，近似计算知道：$T_1=4-2\sqrt{3}\sim 0.5358$。

当$n=4$时,

$$OPT(4,r)=\frac{r}{4}$$

$$OPT(3,r)=\begin{cases} \frac{r}{3}, & r\geq 1.5\ ;\\ \int_r^{r+1} OPT(4,y)dy=\frac{2r+1}{8}, & r<1.5. \end{cases}$$

$$OPT(2,r)=\begin{cases} \frac{r}{2}, & r\geq T_2\ ;\\ \int_r^{1.5} \frac{2y+1}{8}dy+\int_{1.5}^{r+1} \frac{y}{3}dy=\frac{y^2}{24}+\frac{5y}{24}+\frac{25}{96}, & 0.5\leq r<T_2.\\ \int_r^{r+1} \frac{2y+1}{8}dy=\frac{y+1}{4}, & r<0.5. \end{cases}$$

经过计算我们得到$T_2=3.5-\sqrt{6}$。于是我们继续计算T_1

$$OPT(1,r)=\begin{cases} r, & r\geq T_1\ ;\\ \int_r^{r+1} OPT(2,y)dy=F_{U_2}(r+1)-F_{U_2}(r), & r<T_1. \end{cases}$$

其中我们令 $F_{U_2}(y)=\int_0^y OPT(2,z)dz$，为原函数,那么

$$F_{U_2}=\begin{cases} 0.125y^2+0.25y & 0.<y\leq 0.5\\ 0.0138889y^3+0.104167y^2+0.260417y-0.00173611 & 0.5<y\leq 1.05051\\ 0.25y^2+0.126998 & 1.05051<y\leq 2. \end{cases}$$

于是T_1 是方程 $r_1=F_{U_2}(r+1)-F_{U_2}(r)$的根，我们通过近似计算得到$T_1=0.553772$。

第 5 章　结语

本文研究了先知不等式，最优停时以及最优停时的近似计算方法。首先对于先知不等式的特殊目标函数进行了研究，换最大值目标函数为时间平均与递减累积两种情况，分别完整解出了$n = 2$的最优停时以及先知优势因子。对于$n \geq 3$的情形，我们看到了应用基本方法计算出优势因子的难度，因此我们转而探寻最优停时的策略模式。我们证明了最优策略一定是一个截断策略，并对于一般的n我们得到了这些截断值的一些性质。最后，我们对于一些特殊的分布计算了一些截断值。同时对于一个特殊的等距离散分布，我们设计了动态规划算法来计算其期望最优收益值，并且根据这个结论设计出了对于一般的概率分布近似计算最优收益值的方法。对于未来的研究方向，先知不等式更换为其他形态下的目标函数的情形仍然值得研究，并且其他随机优化问题的最优策略形态，截断策略的一般使用性范围也值得探寻。

公式索引

参考文献

[1] P.Auer et al. *Finite-time Analysis of the Multiarmed Bandit Problem*. In *Machine Learning*, 47, 235-256, 2002.

[2] Theodore P. Hill and Robert P. Kertz. *A survey of prophet inequalities in optimal stopping theory*. In *Contemporary Mathematics*, 125:191-207, 1992.

[3] P.R Freeman. *The secretary problem and its extensions: A review*. In *International Statistical Review*, 51(2), 189-206, 1983.

[4] C.Wei et al. *Combinatorial multi-armed bandit: general framework, results and applications*. In *Proceedings of the 30th International Conference on Machine Learning (ICML)*, 2013.

[5] S.M. Kakade et al. *Playing Games with Approximation Algorithms*. In *SIAM Journal of Computing*, Vol. 39, No. 3, pp. 1088-1106, 2009.

[6] A.Badanidiyuru et al. *Bandits with Knapsacks*. In *54th IEEE Symposium on Foundations of Computer Science (FOCS)*, 2013.

[7] T.M. Cover et al. *On Determining the Irrationality of the Mean of a Random Variable*. In *The Annals of Statistics*, Volume 1, No. 5, 862-871, 1973.

致　谢

我首先希望感谢清华大学数学系王振波老师，他的耐心的指导以及对于论文方向把控的能力，让我在写论文和做理论研究的过程中都有眼界和手法上的提升。其次，我希望感谢现在明尼苏达大学的助理教授，也是清华校友王子卓老师，论文中曾经遇到选择分支研究问题的阻碍，得到了老师的尽心的指点从而明确了后续发展方向。同时，我希望感谢在我本文研究中和我进行过讨论，对我有很大帮助的北京大学曾力玮同学和清华李孚同学，两位同学在我做研究遇到细节问题与我进行过深入讨论，帮助我完成一些计算以及特殊情况的分析，这对于研究的顺利进展也起到至关重要的作用。

声　明

本人郑重声明：所呈交的学位论文，是本人在导师指导下，独立进行研究工作所取得的成果。尽我所知，除文中已经注明引用的内容外，本学位论文的研究成果不包含任何他人享有著作权的内容。对本论文所涉及的研究工作做出贡献的其他个人和集体，均已在文中以明确方式标明。

签　名：_____ 日　期：_____

附录 A　英文文献调研报告

In this literature review, two fundamental series of optimal stopping problems are reviewed: Secretary Problem[3] and Prophet Inequality[2]. Both of them has the flavor of finding the optimal stopping strategy. However, they have different objectives. The secretary problem assumes that the secretaries come in a completely random order, aiming to find the optimal strategy and compute the competitive ratio if possible. The prophet inequality problem assumes that the sequence of coming item are fixed but the knowledge of us are only their distribution, aiming to compute the advantage that god (with complete foresight) can have against us. In short, Secretary Problems cares for the strategy and the Prophet Inequalities focuses more on the exact ratio.

The standard secretary problem assumes that n secretaries come in a uniform random order, each reveals her value when she comes to you and you must make an irrevocable decision between stopping and take the current one and continuing to see the next one. The objective is to design a strategy so that the probability for you to stop at the best secretary is maximized. The standard problem can be solved standardly using dynamic programming, a celebrated method in control theory as well as stochastic optimization. The key point here is that the assumption "coming in a random order" seems to make the decision harder, but actually turns the whole problem into a probability-computing one since with this assumption, the probability for a specific strategy to win can be clearly computed, and therefore a clean recurrence equation can be set up. Along with dynamic programming, we often figure out the solution by observing and prove it by induction. And it turns out that the best stopping strategy is simply ignoring the first $\frac{1}{e}$ fraction of coming secretary and stop one you see a secretary that is better than all previous ones.

The secretary problem has many variations, where the modifications of the original problem are mainly three-fold: objective function, number of items and number of choices. For objective function, we can maximize the expected rank of the selected secretary of a strategy rather than the probability for it to select the best one. As a generalization, we can assign general utility function to secretaries (instead of rank), and

the objective function is naturally the expected utility. Such modification can usually be solved by dynamic programming, but they are not lucky enough to bring about an elegant form of the best strategy as the standard problem. For number of items, we can change n to a random variable with known distribution, or consider the following setting: The secretaries come according to a poisson process and you are required to select one by some time horizon T. Such modifications divert the main concern of the standard problem: "structural probability computing" into "special analysis", i.e. to focus on characteristics of special distribution (like poisson distribution) and its centralized computing. Often the remaining problem becomes a function analysis work, namely computing the zeroes or maximal points of some weird-looking objective functions. Another aspect is to change the number of choices (originally we are only allowed to select one secretary, this rule could be replaced by selecting k secretaries, or more generally, the selected secretaries must satisfy some combinatorial constraints, etc). These modifications have various application in real life, but unsurprisingly, to solve most of them it suffices to analyze the combinatorial structure into details and create a rigorous proof of a solution inheriting the solution of original problem.

Prophet inequality also considers a "optimal stopping" problem, but focus on how much advantage the god (the strategy designer with complete foresight) has against us. The standard problem is that given a class C of random variables $\mathbf{X} = (X_1, X_2, \cdots)$, we are required to find the universal inequalities valid for all $\mathbf{X}$ in C which compare the expected supremum of the sequence with the optimal stopping value of the sequence. To be specific, if M denotes the expected supremum:

$$M = M(\mathbf{X}) = \mathbb{E}[\sup_n X_n]$$

and V denoted the optimal stopping value (over the set $\mathcal{T} = \mathcal{T}(\mathbf{X})$ of stopping rules for $\mathbf{X}$)

$$V = V(\mathbf{X}) = \sup_{t \in \mathcal{T}} \mathbb{E}[X_t]$$

If the random variables are independent and only take non-negative values, there have been a celebrated result:

$$V \leq M \leq 2V$$

And it is shown that this bound is tight.

Also, variations make the prophet inequality interesting in many other cases. As the most representative example, the objective function can be changed into time-average payoff:

$$Y_j = \frac{1}{j}\sum_{i=1}^{j} X_i$$

or the time-discount payoff

$$Y_j = \sum_{i=1}^{j} \alpha^{j-i} X_i$$

Unlike the secretary problem, the techniques used in different cases are also not alike. But the following lemma mentioned in[2] (called the dilation lemma) remains core in solving the problem. This is because we are maximizing over all distribution, but if we can adjust the distribution (into special forms) without decreasing the value of objective function, then it suffices to consider this small class of distributions and the problem often becomes much easier.

引理 A.1： (Dilation Lemma)

Let X be any integrable r.v and $-\infty < a < b < +\infty$, $(X)_a^b$ is a r.v. satisfying $(X)_a^b = X$ if $X \notin [a,b]$, $(X)_a^b = a$ with probability $(b-a)^{-1}\int_{X\in[a,b]}(b-X)dP(x)$, and $(X)_a^b = b$ with probability $(b-a)^{-1}\int_{X\in[a,b]}(X-a)dP(x)$, this $(X)_a^b$ is called the dilation of X on interval $[a,b]$, and the following two properties hold.

(1)$[X] = [(X)_a^b]$.

(2)If Y is any r.v independent of both X and $(X)_a^b$, then $\mathbb{E}[\max\{X,Y\}] \leq \mathbb{E}[\max\{(X)_a^b, Y\}]$.

综合论文训练记录表

学生姓名		学号		班级	
论文题目					
主要内容以及进度安排	指导教师签字： 考核组组长签字： 年　　月　　日				
中期考核意见	考核组组长签字： 年　　月　　日				

指导教师评语	指导教师签字：_____ 年　　月　　日
评阅教师评语	评阅教师签字：_____ 年　　月　　日
答辩小组评语	答辩小组组长签字：_____ 年　　月　　日

总成绩：_____

教学负责人签字：_____

年　　月　　日