浙江大学
ZheJiang University
运筹与决策

科技创新类通识课程

运筹与决策

浙江大学数学科学学院　谈之奕

数学规划

赛程编制问题

浙江大学 ZheJiang University

运筹与决策

- 2018世界杯南美赛区预选赛
  - 10个成员国，4.5个决赛阶段名额
  - 双循环主客场制，9阶段18轮。两轮为一个阶段，每阶段跨时一周，不同阶段相隔一月或数月
- 2002-2014世界杯南美赛区预选赛赛程

	1	2	3	4	5	6	7	8	9
ARG	CHI	VEN	BOL	COL	ECU	BRA	PAR	PER	URU
BOL	URU	COL	ARG	VEN	CHI	PAR	ECU	BRA	PER
BRA	COL	ECU	PER	URU	PAR	ARG	CHI	BOL	VEN
CHI	ARG	PER	URU	PAR	BOL	VEN	BRA	COL	ECU
COL	BRA	BOL	VEN	ARG	PER	ECU	URU	CHI	PAR
ECU	VEN	BRA	PAR	PER	ARG	COL	BOL	URU	CHI
PAR	PER	URU	ECU	CHI	BRA	BOL	ARG	VEN	COL
PER	PAR	CHI	BRA	ECU	COL	URU	VEN	ARG	BOL
URU	BOL	PAR	CHI	BRA	VEN	PER	COL	ECU	ARG
VEN	ECU	ARG	COL	BOL	URU	CHI	PER	PAR	BRA

- CONMEBOL -

FIFA WORLD CUP RUSSIA 2018

	Argentina 阿根廷
	Bolivia 玻利维亚
	Brazil 巴西
	Chile 智利
	Colombia 哥伦比亚
	Ecuador 厄瓜多尔
	Paraguay 巴拉圭
	Peru 秘鲁
	Uruguay 乌拉圭
	Venezuela 委内瑞拉

赛程特点

- **2002-2014世界杯南美赛区预选赛赛程特点**
  - 任意两队在前后两个半程各交手一次，两场比赛的主客场互换
    - 镜像双循环 1~10, 2~11, ……， 9~18
  - 不存在多于两场的连续主场与客场
  - 任一队不连续对阵巴西与阿根廷
- 赛程缺点
  - 存在同一阶段内两场比赛均为主场或客场的情况，且各队出现上述情况的次数不均衡
  - 同一阶段内各队先主后客和先客后主的次数不均衡
  - 赛程编制原理不透明，关键比赛存在争议

最后一轮：阿根廷——乌拉圭

	2002-2014		
	主主,客客	主客	客主
ARG	0	9	0
BOL	4	2	3
BRA	0	0	9
CHI	2	1	6
COL	2	6	1
ECU	2	4	3
PAR	2	3	4
PER	2	6	1
URU	2	4	3
VEN	2	1	6

赛程编制新举措

浙江大学 ZheJiang University
运筹与决策

- **2018世界杯新举措**
  - 各成员国提交候选方案，南美洲足联投票决定最终赛程模板
  - 赛程模板中各队用编号代替，抽签决定编号与球队对应关系（种子队与非种子队分别抽签）
  - **Durán**团队为智利足联编制赛程已逾十年，他们设计的方案为智利足联所采纳，并最终在投票中胜出

Guillermo Durán
Professor of Department of Mathematics and Calculus Institute Faculty of Exact and Natural Sciences University of Buenos Aires

Alarcón F, Durán G, Guajardo M. Referee assignment in the Chilean football league using integer programming and patterns. *International Transactions in Operational Research*, 21: 415-438, 2014.

Bonomo F, Cardemil A, Durán G, et al. An application of the traveling tournament problem: The Argentine volleyball league. *Interfaces*, 42: 245-259, 2012.

Durán G, Guajardo M, Wolf-Yadlin R. Operations research techniques for scheduling Chile's second division soccer league. *Interfaces*, 42: 273-285, 2012.

镜像赛程

- 根据世界杯南美赛区预选赛的特点，不必考虑连续两场比赛之间的**break**，只需考虑同一阶段两场比赛之间的**double-round break**

如何证明？

- **10**支队的镜像赛程的**double-round break**数至少为**16**
- **10**支队的法制赛程的**double-round break**数可以为**0**

	1	2	3	…	9	10	11	12	…	17	18	
镜像（mirror）	1	2	3	…	9	1	2	3	…	8	9	意、德
法制（French）	1	2	3	…	9	2	3	4	…	9	1	法、俄
英制（English）	1	2	3	…	9	9	1	2	…	7	8	奥
逆向（Inverted）	1	2	3	…	9	9	8	7	…	2	1	瑞士

整数规划

- 决策变量 $x_{ijk}=\begin{cases}1 & 第k轮队i在主场与队j比赛,\\ 0 & 其他,\end{cases}$ $i,j=1,\cdots,10,k=1,\cdots 18$
- 约束条件（部分）
  - 每轮各队恰有一场比赛
    - 对任意队 j，在第 k 轮，与除本队外的某一队进行比赛，比赛可能是主场，也可能是客场
    - 若比赛是主场，则存在唯一的 i_0，使得 $x_{ji_0k}=1$，对 $i\neq i_0$，$x_{jik}=0$；若比赛是客场，则存在唯一的 i_0，使得 $x_{i_0jk}=1$，对 $i\neq i_0$，$x_{ijk}=0$

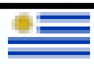
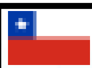
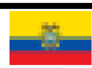
$$\sum_{i=1}^{10}\left(x_{ijk}+x_{jik}\right)=1,\quad j=1,\cdots,10,k=1,\cdots,18$$

  - 任意两队在前后半程各交手一次，两场比赛中各有一个主场

$$\sum_{k=1}^{9}\left(x_{ijk}+x_{jik}\right)=1,\ i,j=1,\cdots,10\qquad \sum_{k=10}^{18}\left(x_{ijk}+x_{jik}\right)=1,\ i,j=1,\cdots,10\qquad \sum_{k=1}^{18}x_{ijk}=1,\quad i,j=1,\cdots,10,i\neq j$$

最终赛程

	1	2	3	4	5	6	7	8	9
ARG	ECU	PAR	BRA	COL	CHI	BOL	URU	VEN	PER
BOL	URU	ECU	VEN	PAR	COL	ARG	PER	CHI	BRA
BRA	CHI	VEN	ARG	PER	URU	PAR	ECU	COL	BOL
CHI	BRA	PER	COL	URU	ARG	VEN	PAR	BOL	ECU
COL	PER	URU	CHI	ARG	BOL	ECU	VEN	BRA	PAR
ECU	ARG	BOL	URU	VEN	PAR	COL	BRA	PER	CHI
PAR	VEN	ARG	PER	BOL	ECU	BRA	CHI	URU	COL
PER	COL	CHI	PAR	BRA	VEN	URU	BOL	ECU	ARG
URU	BOL	COL	ECU	CHI	BRA	PER	ARG	PAR	VEN
VEN	PAR	BRA	BOL	ECU	PER	CHI	COL	ARG	URU

排名										
积分	41	31	28	27	26	26	24	20	14	12
净胜球	30	12	3	2	1	-1	-6	-3	-22	-16

	2018		
	主主, 客客	主客	客主
ARG	0	5	4
BOL	0	5	4
BRA	0	4	5
CHI	0	5	4
COL	0	5	4
ECU	0	4	5
PAR	0	4	5
PER	0	4	5
URU	0	4	5
VEN	0	5	4

Durán G, Guajardo M, Sauré D. Scheduling the South American Qualifiers to the 2018 FIFA World Cup by integer programming. *European Journal of Operational Research*, 262, 1109-1115, 2017.

随机变量

- 证券市场有若干种股票，如何投资可使得收益最大而风险最小
  - 股票的收益率具有不确定性，可视作一随机变量
- 随机变量的数字特征
  - 期望：$\mu = \mathrm{E}r$
  - 方差：$\sigma^2 = \mathrm{Var}(r) = \mathrm{E}(r-\mathrm{E}r)^2$
  - 协方差：$\sigma_{ij}=\mathrm{Cov}(r_i,r_j) = \mathrm{E}(r_i-\mathrm{E}r_i)(r_j-\mathrm{E}r_j)$
- 随机变量线性函数的数字特征
  - 随机变量向量 $\mathbf{r}=(r_1,r_2,\cdots,r_n)^{\mathrm{T}}$，$\mathrm{E}r_j=\mu_j$，$\mathrm{Cov}(r_i,r_j)=\sigma_{ij}$，期望向量 $\boldsymbol{\mu}=(\mu_1,\mu_2,\cdots,\mu_n)^{\mathrm{T}}$，协方差矩阵为 $\mathbf{V}=(\sigma_{ij})_{n\times n}$
  - 线性函数 $X=\sum_{i=1}^{n}x_i r_i$，记 $\mathbf{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$，则 $X=\mathbf{x}^{\mathrm{T}}\mathbf{r}$
  - 期望 $\mathrm{E}\left(\mathbf{x}^{\mathrm{T}}\mathbf{r}\right)=\mathrm{E}\left(\sum_{i=1}^{n}x_i r_i\right)=\sum_{i=1}^{n}x_i\mu_i=\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}$，方差

$$\mathrm{Var}\left(\mathbf{x}^{\mathrm{T}}\mathbf{r}\right)=\mathrm{Var}\left(\sum_{i=1}^{n}x_i r_i\right)=\sum_{i=1}^{n}\sum_{j=1}^{n}x_i x_j\mathrm{E}(r_i-\mathrm{E}r_i)(r_j-\mathrm{E}r_j)=\sum_{i=1}^{n}\sum_{j=1}^{n}\sigma_{ij}x_i x_j=\mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{x}$$

资产投资组合

- 证券市场有 n 种股票，股票 j 的收益率为 r_j
  - 随机变量向量 $\mathbf{r}=(r_1,r_2,\cdots,r_n)^{\mathrm{T}}$ 的期望向量为 $\boldsymbol{\mu}$，协方差矩阵为 $\mathbf{V}$
- 风险（**risk**）：可能发生的危险
  - 股票的风险为其收益率的标准差，反映了收益率围绕其均值波动的幅度
- 将总投资额单位化，投资于股票 j 的份额为 $x_j, j=1,\cdots,n$，该投资组合（**portfolio**）用向量 $\mathbf{x}=(x_1,x_2,\cdots,x_n)^{\mathrm{T}}$表示
  - 收益为 $\mathrm{E}(\mathbf{x}^{\mathrm{T}}\mathbf{r})=\mathbf{x}^{\mathrm{T}}\boldsymbol{\mu}$，风险的平方为 $\mathrm{Var}(\mathbf{x}^{\mathrm{T}}\mathbf{r})=\mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{x}$

Harry Markowitz
(1927-　)
美国经济学家
1990年**Nobel**经济学奖得主

浙江大学 ZheJiang University

运筹与决策

Markowitz模型

- 选择投资组合 $\mathbf{x}^*(\mu)$，在收益达到给定值 μ的前提下，组合的风险最小

$$\begin{aligned} \min \quad & \mathbf{x}^{\mathrm{T}}\mathbf{V}\mathbf{x} \\ \text{s.t.} \quad & \mathbf{x}^{\mathrm{T}}\boldsymbol{\mu} = \mu \\ & \mathbf{x}^{\mathrm{T}}\mathbf{e} = 1 \end{aligned}$$

变双目标为单目标

条件极值问题

- 对应于μ的极小风险组合

$$\mathbf{x}^*(\mu) = \frac{1}{ac-b^2}\mathbf{V}^{-1}(\boldsymbol{\mu} \quad \mathbf{e})\begin{pmatrix} c & -b \\ -b & a \end{pmatrix}\begin{pmatrix} \mu \\ 1 \end{pmatrix}$$

其中

$$a = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{V}^{-1}\boldsymbol{\mu}, b = \boldsymbol{\mu}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{e}, c = \mathbf{e}^{\mathrm{T}}\mathbf{V}^{-1}\mathbf{e}$$

- 最优组合下风险与收益的关系

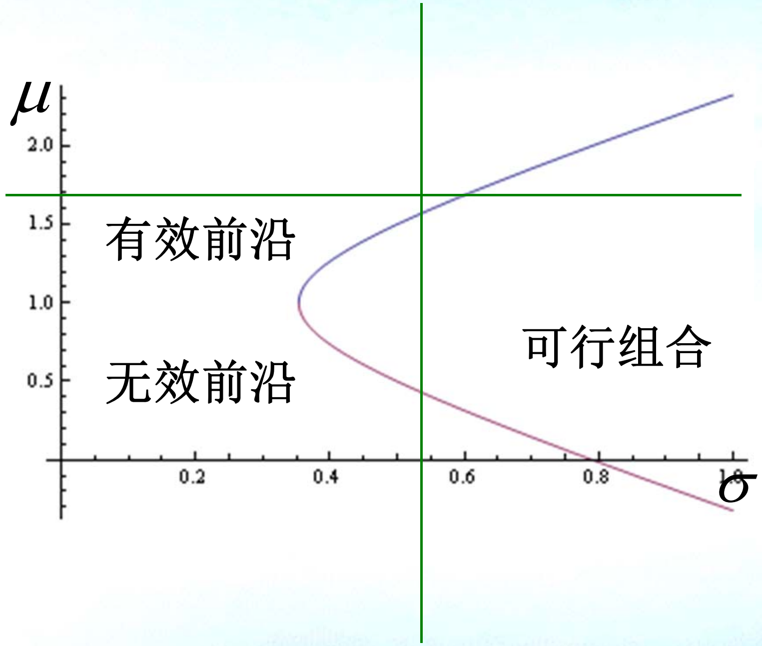
$$\sigma^*(\mu)^2 = \mathbf{x}^*(\mu)^{\mathrm{T}}\mathbf{V}\mathbf{x}^*(\mu) = \frac{a-2b\mu+c\mu^2}{ac-b^2}$$

- 在(σ,μ)平面上，极小风险组合的收益 μ与风险 $\sigma^*(\mu)$的轨迹为一条双曲线的右支

$$\frac{\sigma^*(\mu)^2}{\frac{1}{c}} - \frac{\left(\mu-\frac{b}{c}\right)^2}{\frac{ac-b^2}{c^2}} = 1$$

有效前沿

- 双曲线上半部称为有效前沿（**efficient frontier**）。其上每一点对应的组合为有效组合，即收益固定时风险最小的组合或风险固定时收益最大的组合
- 双曲线下半部为无效前沿（**inefficient frontier**）
- 双曲线顶点为总体最小风险组合



高收益对应高风险

Markowitz, H., Portfolio Selection, *The Journal of Finance*, 7, 77-91, 1952

投资组合理论

Finally, I would like to add a comment concerning portfolio theory as a part of the microeconomics of action under uncertainty. It has not always been considered so. For example, when I defended my dissertation as a student in the Economics Department of the University of Chicago, Professor Milton Friedman argued that portfolio theory was not Economics, and that they could not award me a Ph.D. degree in Economics for a dissertation which was not in Economics. I assume that he was only half serious, since they did award me the degree without long debate. As to the merits of his arguments, at this point I am quite willing to concede: at the time I defended my dissertation, portfolio theory was not part of Economics. But now it is.

Foundations of Portfolio Theory

——Nobel Prize Lecture by Harry Markowitz

支持向量机

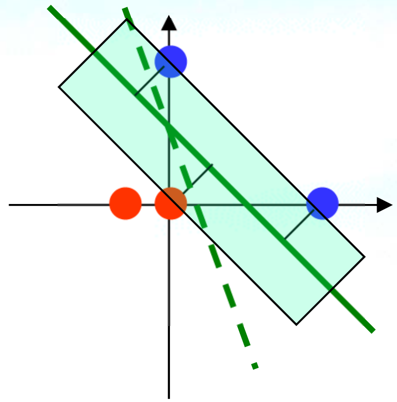
- 支持向量机（**Support Vector Machine**）
  - 拟将一数据集分为 C_1, C_2 两类。每个数据有 n 个特征，用 n 维实向量表示数据
  - 训练集 $S=\{\mathbf{x}_1,\mathbf{x}_2,\cdots,\mathbf{x}_m\}$，其分类已知，记 $y_i=\begin{cases}1 & \mathbf{x}_i\in C_1\\ -1 & \mathbf{x}_i\in C_2\end{cases}$
  - 训练集可线性分离（**linearly separable**），即存在超平面 $\mathbf{w}\cdot\mathbf{x}+b=0$，使得 $\begin{cases}\mathbf{w}\cdot\mathbf{x}_i+b>0 & \mathbf{x}_i\in C_1\\ \mathbf{w}\cdot\mathbf{x}_i+b<0 & \mathbf{x}_i\in C_2\end{cases}$，或 $y_i(\mathbf{w}\cdot\mathbf{x}_i+b)>0$
- 超平面
  - 设 $\mathbf{w}$为 n 维实向量，b 为实数，称 $\mathbf{w}\cdot\mathbf{x}+b=0$为 $\mathbb{R}^n$ 中的超平面（**hyperplane**）
  - $\mathbb{R}^n$中点 $\mathbf{x}$ 到超平面 $\mathbf{w}\cdot\mathbf{x}+b=0$的距离为 $\dfrac{|\mathbf{w}\cdot\mathbf{x}+b|}{\sqrt{\mathbf{w}\cdot\mathbf{w}}}$

不妨要求 $\mathbf{w}\cdot\mathbf{w}=1$

Cortes C, Vapnik V. Support-vector networks. *Machine Learning*, *20*(3), 273-297, 1995.

支持向量机

- 若（I）有解，（I）与（II）等价
  - （I）的可行域包含在（II）的可行域内
  - （II）的最优解在（I）的可行域内
    - 由于（I）有解，存在 $\mathbf{w},b$，满足 $\mathbf{w}\cdot\mathbf{w}=1$ 与 $y_i(\mathbf{w}\cdot\mathbf{x}_i+b)>0,\ i=1,\cdots,m$ 。这也是（II）的一组可行解，故（II）的最优值非负。因此（II）的最优解 $\mathbf{w}^*,b^*$ 总满足 $y_i(\mathbf{w}^*\cdot\mathbf{x}_i+b^*)>0,\ i=1,\cdots,m$
  - 在（I）的可行域内，对相同决策变量，（I）的目标值与（II）的目标值相等
    - 由于 $y_i=\pm1$，若 $y_i(\mathbf{w}\cdot\mathbf{x}_i+b)>0$，则 $y_i(\mathbf{w}\cdot\mathbf{x}_i+b)=|\mathbf{w}\cdot\mathbf{x}_i+b|$



所有点至超平面距离的最小值尽可能大

（I）
$$\begin{aligned}\max\quad & \min_{i=1,\cdots,m}|\mathbf{w}\cdot\mathbf{x}_i+b|\\ \text{s.t.}\quad & y_i(\mathbf{w}\cdot\mathbf{x}_i+b)>0,\ i=1,\cdots,m\\ & \mathbf{w}\cdot\mathbf{w}=1\end{aligned}$$

（II）
$$\begin{aligned}\max\quad & \min_{i=1,\cdots,m} y_i(\mathbf{w}\cdot\mathbf{x}_i+b)\\ \text{s.t.}\quad & \mathbf{w}\cdot\mathbf{w}=1\end{aligned}$$

支持向量机

- 若 $\mathbf{w}_0, b_0$ 是（III）的最优解，则 $\dfrac{\mathbf{w}_0}{\sqrt{\mathbf{w}_0\cdot\mathbf{w}_0}}, \dfrac{b_0}{\sqrt{\mathbf{w}_0\cdot\mathbf{w}_0}}$ 是（II）的最优解
  - 设 $\mathbf{w}^*, b^*$ 是（II）的最优解，$\mathbf{w}^*\cdot\mathbf{w}^*=1$，最优值为 $\gamma^* = \min\limits_{i=1,\cdots,m} y_i(\mathbf{w}^*\cdot\mathbf{x}_i + b^*)$
  - $y_i(\mathbf{w}^*\cdot\mathbf{x}_i + b^*) \ge \gamma^*$，$i=1,\cdots,m$，即 $y_i\left(\dfrac{\mathbf{w}^*}{\gamma^*}\cdot\mathbf{x}_i + \dfrac{b^*}{\gamma^*}\right) \ge 1,\ i=1,\cdots,m$，
    故 $\dfrac{\mathbf{w}^*}{\gamma^*}, \dfrac{b^*}{\gamma^*}$ 是（III）的可行解
  - 由于 $\mathbf{w}_0, b_0$ 是（III）的最优解，$\sqrt{\mathbf{w}_0\cdot\mathbf{w}_0} \le \sqrt{\dfrac{\mathbf{w}^*}{\gamma^*}\cdot\dfrac{\mathbf{w}^*}{\gamma^*}} = \dfrac{1}{\gamma^*}$
  - $y_i\left(\dfrac{\mathbf{w}_0}{\sqrt{\mathbf{w}_0\cdot\mathbf{w}_0}}\cdot\mathbf{x}_i + \dfrac{b_0}{\sqrt{\mathbf{w}_0\cdot\mathbf{w}_0}}\right) \ge y_i\left(\gamma^*\mathbf{w}_0\cdot\mathbf{x}_i + \gamma^* b_0\right) = \gamma^* y_i\left(\mathbf{w}_0\cdot\mathbf{x}_i + b_0\right) \ge \gamma^*,\ i=1,\cdots,m$ ，
    故 $\dfrac{\mathbf{w}_0}{\sqrt{\mathbf{w}_0\cdot\mathbf{w}_0}}, \dfrac{b_0}{\sqrt{\mathbf{w}_0\cdot\mathbf{w}_0}}$ 的目标值不小于 $\mathbf{w}^*, b^*$ 的目标值，也是（II）的最优解

（II）
$$\begin{aligned} \max \quad & \min_{i=1,\cdots,m} y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \\ \text{s.t.} \quad & \mathbf{w}\cdot\mathbf{w} = 1 \end{aligned}$$

带不等式约束的二次规划

（III）
$$\begin{aligned} \min \quad & \mathbf{w}\cdot\mathbf{w} \\ \text{s.t.} \quad & y_i(\mathbf{w}\cdot\mathbf{x}_i + b) \ge 1,\ i=1,\cdots,m \end{aligned}$$


浙江大学
ZheJiang University
运筹与决策
谢 谢