Задача о двуруком бандите. Краткий обзор моделей и подходов

А.В.Колногоров[1]

[1]Новгородский государственный университет им. Ярослава Мудрого
Alexander.Kolnogorov@novsu.ru

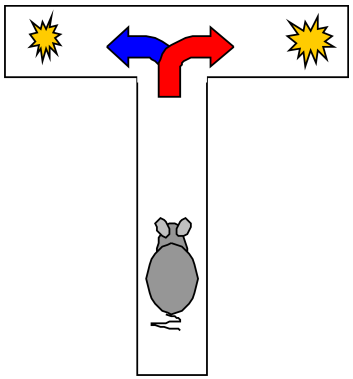
Семинар «Стохастический анализ в задачах»

г. Москва

21 сентября 2013

Целесообразное поведение в случайной среде

1. Цетлин М.Л. Исследования по теории автоматов и моделированию биологических систем. М.: Наука, 1969
2. Tsetlin, M.L. Automation Theory and Modeling of Biological Systems. Academic Press, New York. 1973.



Животное (обычно, крыса) должно выбрать одно из 2-х направлений в T-образном лабиринте. В конце лабиринта его ожидает удар тока с вероятностями q_1 и q_2.
Вероятности q_1, q_2 были фиксированы и крыса демонстрировала способность к обучению выбирать направление, которому соответствовала меньшая вероятность.

Формальное определение целесообразного поведения

Можно рассматривать случайный процесс реакций среды $\xi_1, \xi_2, \ldots, \xi_n$, зависящих только от текущих выбираемых действий (направлений) $y_1, y_2, \ldots, y_n$ следующим образом:

$$\Pr(\xi_n = 1 | y_n = \ell) = p_\ell, \quad \Pr(\xi_n = 0 | y_n = \ell) = q_\ell, \quad \ell = 1, 2.$$

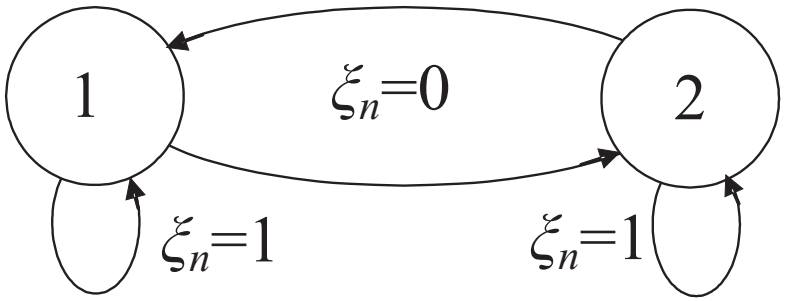
- $\xi_n = 1$ - нет удара током
- $\xi_n = 0$ - есть удар током
- $y_n = 1$ - поворот налево
- $y_n = 2$ - поворот направо

Вероятности p_1, p_2 фиксированы в процессе управления, но неизвестны. Значения процесса интерпретируются как доходы и цель состоит в максимизации среднего ожидаемого дохода.

Простейший автомат

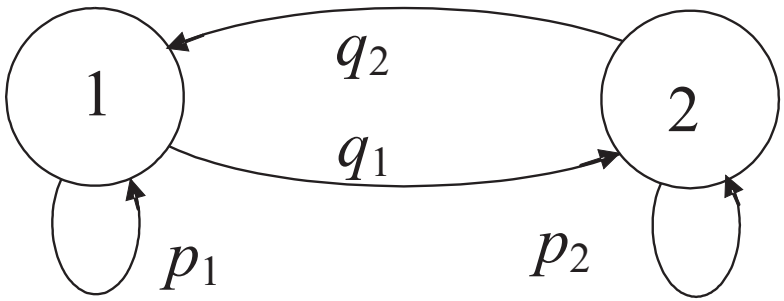
Возможная стратегия крысы в лабиринте:

- Изменить направление, если получен удар тока
- Не менять направление, если удар тока не получен



Стратегия может быть представлена симметричным автоматом с 2 состояниями.

Ассоциированная цепь Маркова



Система уравнений для предельных вероятностей имеет вид

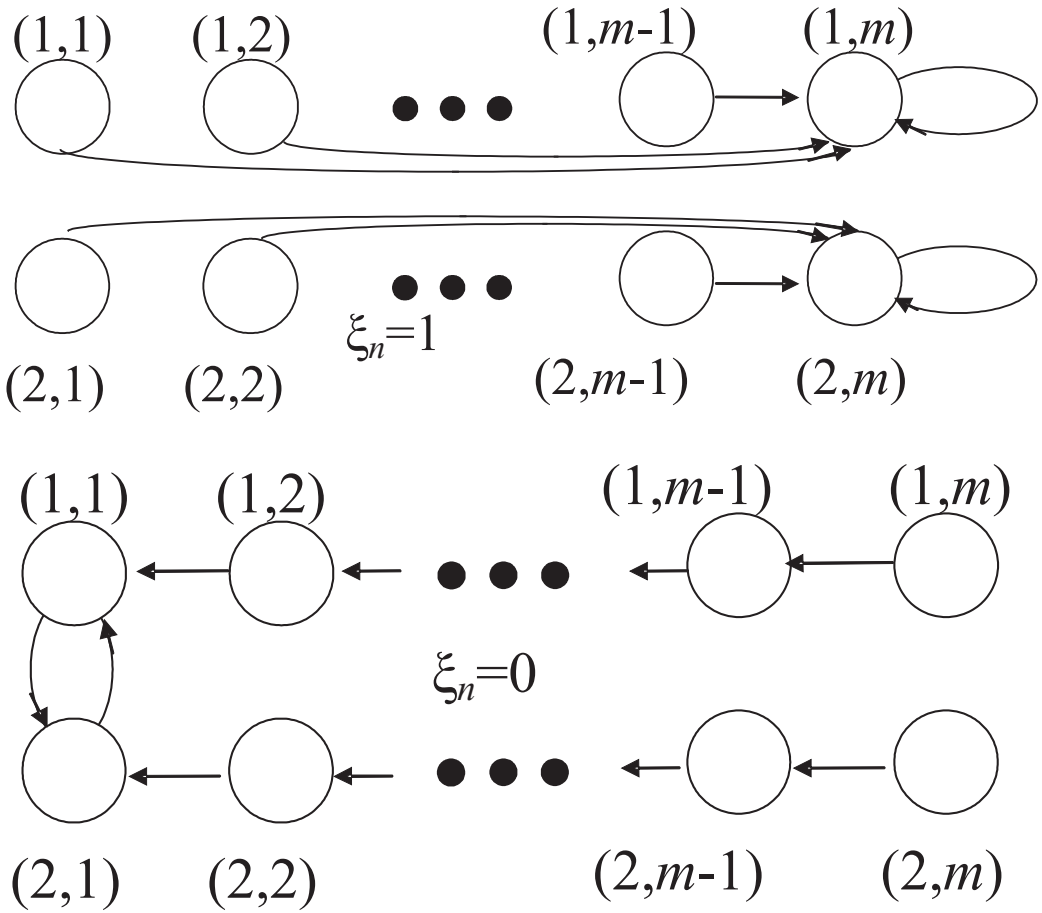
$$\pi_1 = \pi_1 p_1 + \pi_2 q_2,$$
$$\pi_1 + \pi_2 = 1,$$

где второе уравнение дает условие нормировки. Решение этой системы имеет вид

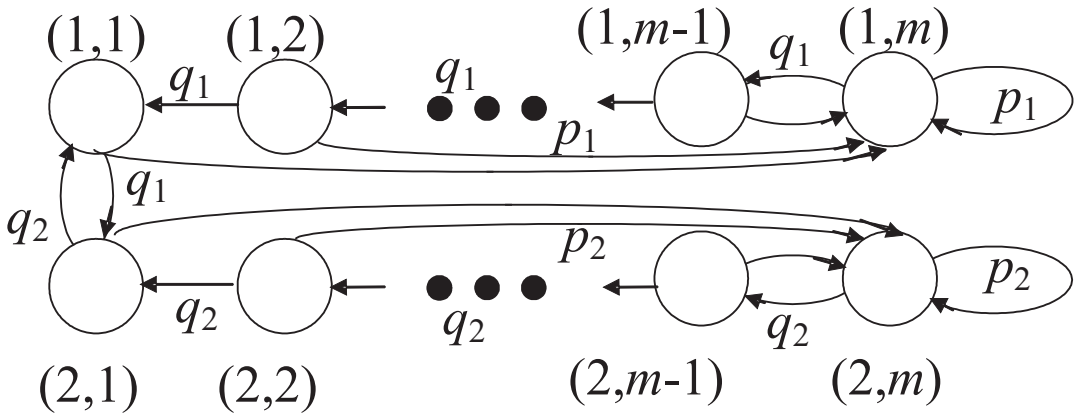
$$\frac{\pi_1}{\pi_2} = \frac{q_2}{q_1} \to \pi_1 > \pi_2 \Leftrightarrow p_1 > p_2, \quad \pi_1 < \pi_2 \Leftrightarrow p_1 < p_2.$$

Доверчивый автомат (Роббинса - Кринского)

- Перейти в самое глубокое состояние, если удар тока не получен
- Перейти в менее глубокое состояние или поменять действие в наименее глубоком состоянии, если удар тока получен



Ассоциированная цепь Маркова



Предельные вероятности удовлетворяют системе уравнений

$$\begin{aligned}
&\mu_{1,1} = \mu_{1,2}q_1 + \mu_{2,1}q_2, && \mu_{2,1} = \mu_{2,2}q_2 + \mu_{1,1}q_1,\\
&\mu_{1,2} = \mu_{1,3}q_1, && \mu_{2,2} = \mu_{2,3}q_2,\\
&\dots && \dots\\
&\mu_{1,m-1} = \mu_{1,m}q_1, && \mu_{2,m-1} = \mu_{2,m}q_2,\\
&\mu_{1,m} = \pi_1 p_1, && \mu_{2,m} = \pi_2 p_2,
\end{aligned}$$

где $\pi_\ell = \mu_{\ell,1} + \dots + \mu_{\ell,m}$, $\ell = 1, 2$; $\pi_1 + \pi_2 = 1$. При этом

$$\frac{\pi_1}{\pi_2} = \left(\frac{q_2}{q_1}\right)^m.$$

Стохастические автоматы с переменной структурой

Варшавский В.И. Коллективное поведение автоматов. М.: Наука, 1973.
Состояние автомата в момент времени n – вектор $\boldsymbol{\pi}(n) = (\pi_1(n), \pi_2(n))$ вероятностей выбора действий $\ell = 1, 2$. Такой автомат имеет бесконечную память. Переходы описываются правилом

$$\pi_1(n+1) = \mathcal{P}_{[0,1]}\big(\pi_1(n) + \Delta\pi_1(\boldsymbol{\pi}(n), \ell_n, \xi_n)\big),$$
$$\pi_2(n+1) = \mathcal{P}_{[0,1]}\big(\pi_2(n) + \Delta\pi_2(\boldsymbol{\pi}(n), \ell_n, \xi_n)\big),$$
$$\mathcal{P}_{[0,1]}(x) = \begin{cases} x, & \text{если } x \in [0,1], \\ 0, & \text{если } x < 0, \\ 1, & \text{если } x > 1. \end{cases}$$

Если за выбор текущего действия ℓ_n получено поощрение ($\xi_n = 1$), вероятность выбора его на следующем шаге $\pi_\ell(n+1)$ растет, если штраф ($\xi_n = 0$) — то уменьшается.

Пример автомата с переменной структурой

$$\Delta\pi_1(\boldsymbol{\pi}(n),\ell_n,\xi_n)=\begin{cases}+\alpha\pi_1^{\beta}(n)\pi_2^{1+\beta}(n), & \text{если } \ell_n=1,\xi_n=1,\\ -\alpha\pi_1^{\beta}(n)\pi_2^{1+\beta}(n), & \text{если } \ell_n=1,\xi_n=0,\\ -\alpha\pi_1^{1+\beta}(n)\pi_2^{\beta}(n), & \text{если } \ell_n=2,\xi_n=1,\\ +\alpha\pi_1^{1+\beta}(n)\pi_2^{\beta}(n), & \text{если } \ell_n=2,\xi_n=0,\end{cases}$$

где $\alpha>0$, $\beta>0$. При этом $\Delta\pi_2(\boldsymbol{\pi}(n),\ell_n,\xi_n)=-\Delta\pi_1(\boldsymbol{\pi}(n),\ell_n,\xi_n)$. Можно проверить, что

$$\mathbb{E}(\pi_1(n+1)|\boldsymbol{\pi}(n))=\pi_1(n)+2\alpha(p_1-p_2)\pi_1^{1+\beta}(n)\pi_2^{1+\beta}(n).$$

При $\pi_1(n)>0$, $\pi_2(n)>0$ и $p_1-p_2>0$ вероятность $\pi_1(n)$ растет, а при $p_1-p_2<0$ уменьшается.

Коллективное поведение автоматов

1. Варшавский В. И., Поспелов Д. А. Оркестр играет без дирижера: размышления об эволюции некоторых технических систем и управлении ими.-М.: Наука, 1984.
2. Варшавский В.И., Мелешина М. В., Цетлин М. Л. Организация дисциплины ожидания в системах массового обслуживания с использованием модели коллективного поведения автоматов // Пробл. передачи информ., 4:1 (1968), 73-76.
3. Стефанюк В. Л., Цетлин М. Л. О регулировке мощности в коллективе радиостанций// Пробл. передачи информ., 3:4 (1967), 49-57.
4. Гинзбург С. Л., Цетлин М. Л. О некоторых примерах моделирования коллективного поведения автоматов// Пробл. передачи информ., 1:2 (1965), 54-62.
5. Варшавский В. И., Мелешина М. В., Цетлин М. Л. Поведение автоматов в периодических случайных средах и задача синхронизации при наличии помех// Пробл. передачи информ., 1:1 (1965), 65-71.

Идентификационный подход

1. Срагович В.Г. Теория адаптивных систем. М.: Наука. 1976
2. Срагович В.Г. Адаптивное управление. М.: Наука, 1981.
3. Sragovich, V.G. Mathematical Theory of Adaptive Control // Interdisciplinary Mathematical Sciences – Vol. 4. World Scientific. New Jersey, London, . . . 2006.

В процессе управления делаются оценки параметров среды $\hat{\theta}_n = (\hat{p}_{1n}, \hat{p}_{2n})$. Пересчет вектора состояния осуществляется по правилу:

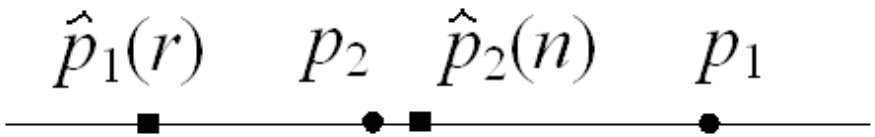
$$\pi(n+1) = Q_n(\pi(n), \hat{\theta}_n).$$

Возможная стратегия. Применить действия по очереди и найти $\hat{\theta}_2$. Затем при $n = 3, 4, \ldots$ повторять процедуру: применять действие, соответствующее текущей большей оценке $\hat{p}_{\ell n}$ с вероятностью $1 - \delta_n$, текущей меньшей оценке $\hat{p}_{\ell n}$ – с вероятностью δ_n, после чего пересчитывать $\hat{\theta}_{n+1}$. Для оптимальности нужно, чтобы

$$\lim_{n\to\infty} \delta_n = 0, \quad \sum_{n=1}^{\infty} \delta_n = \infty.$$

Правило “play-the-leader”(«играй лучшего»)

Состоит в том, что сначала для набора статистики каждый из вариантов применяют одинаковое число r раз (в том числе возможно, что $r = 1$), а затем каждый раз применяют тот вариант, которому соответствует бóльшая текущая оценка из $\hat{p}_1(n)$, $\hat{p}_2(n)$, пересчитывая после этого оценку, соответствующую примененному варианту. Правило не гарантирует оптимального управления.



Предположим, что $p_1 > p_2$. С некоторой вероятностью может выполниться неравенство $\hat{p}_1(r) \ll p_2$. Кроме того, с некоторой вероятностью, оценка $\hat{p}_2(n)$ может незначительно отклоняться от p_2 при всех n и, следовательно, будет выполнено неравенство $\hat{p}_1(n) < \hat{p}_2(n)$ при всех $n > 2r$. В этом случае правило “play-the-leader” приведет к тому, что при всех $n > 2r$ будет выбираться второй вариант, и вероятность этого события ненулевая.

Цели управления

Должны выполняться для всех сред из некоторого класса:

$$\lim_{n\to\infty} E\xi_n \geq (p_1 \vee p_2) - \varepsilon, \quad \lim_{n\to\infty} N^{-1} \sum_{n=1}^{N} E\xi_n \geq (p_1 \vee p_2) - \varepsilon,$$

$$\Pr\left(\lim_{n\to\infty} N^{-1} \sum_{n=1}^{N} \xi_n \geq (p_1 \vee p_2) - \varepsilon \right) = 1$$

– ε-оптимальность в слабом и сильном смыслах,

$$\lim_{n\to\infty} E\xi_n = p_1 \vee p_2, \quad \lim_{n\to\infty} N^{-1} \sum_{n=1}^{N} E\xi_n = p_1 \vee p_2,$$

$$\Pr\left(\lim_{n\to\infty} N^{-1} \sum_{n=1}^{N} \xi_n = p_1 \vee p_2 \right) = 1$$

– асимптотическая оптимальность в слабом и сильном смыслах.

Рекуррентные алгоритмы

1. Назин А.В., Позняк А.С. Адаптивный выбор вариантов. М.: Наука, 1986.
2. Juditsky A., Nazin A. V., Tsybakov A.B., Vayatis N. Gap-free Bounds for Stochastic Multi-Armed Bandit // Proceedings of the 17th World Congress The International Federation of Automatic Control Seoul, Korea, July 6-11, 2008.
3. Poznyak, A.S. and Najim, K. Learning Automata and Stochastic Optimization. Lecture Notes in Control and Information Sciences 225. Springer-Verlag. Berlin, Heidelberg, New York. 1997.

Проанализированы известные и предложены новые алгоритмы типа САПС. Для анализа использовались метод стохастической аппроксимации, алгоритм зеркального спуска и др. Наряду с асимптотической оптимальностью алгоритмов исследовалась гарантированная скорость сходимости в среднеквадратическом

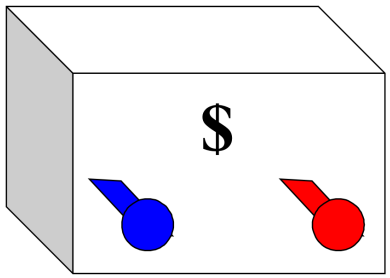
$$\sup_{p_1,p_2} \mathbb{E}\left((p_1 \vee p_2) - N^{-1}\sum_{n=1}^{N}\xi_n\right)^2.$$

Последовательное управление по неполным данным

1. Пресман Э. Л., Сонин И.М. Последовательное управление по неполным данным. – М.: Наука, 1982.
2. Presman, E.L. and Sonin, I.M. Sequential Control with Incomplete Information. Academic Press. New York. 1990.

Рассмотрена байесовская постановка задачи управления с конечным множеством параметров в дискретном и непрерывном времени. В дискретном времени управляемый процесс бинарный, в непрерывном – пуассоновский. Целью является максимизация ожидаемого полного дохода.
В непрерывном времени предложена схема одновременного применения действий (с разделением ресурса). Решена задача синтеза оптимального управления на конечном и бесконечном времени. Установлен пороговый характер оптимальной стратегии.

Задача о двуруком бандите



Это игральный автомат с двумя рукоятками. При нажатии ℓ-ой рукоятки доход игрока равен 1 с вероятностью p_ℓ и 0 с вероятностью q_ℓ $(p_\ell + q_\ell = 1,\ \ell = 1, 2)$.

Игрок может нажать рукоятки в общей сложности N раз. Его целью является максимизация математического ожидания полного дохода. Вероятности p_1, p_2 фиксированы в процессе управления, но неизвестны игроку.

Дилемма «Информация или управление»

Для игрока оптимальной стратегией было бы всегда выбирать ту рукоятку, которой соответствует максимальное значение вероятностей p_1, p_2. Но чтобы определить эту рукоятку, он должен протестировать их обе, и это ведет к уменьшению его полного выигрыша.

Байесовский подход - 1

Формально доходы рассматриваются как случайный процесс $\xi_1, \xi_2, \ldots, \xi_n$, зависящий только от текущих выбираемых действий $y_1, y_2, \ldots, y_n$, именно $\Pr(\xi_n = 1 | y_n = \ell) = p_\ell$, $\Pr(\xi_n = 0 | y_n = \ell) = q_\ell$, $\ell = 1, 2$. Стратегия σ определяет выбор действий y_n, $n = 1, \ldots, N$ и может использовать всю текущую информацию о процессе: n_1, n_2 – количества нажатий и m_1, m_2 – полные выигрыши на обеих рукоятках. Функция потерь такова

$$L_N(\sigma, \theta) = N(p_1 \vee p_2) - E_{\sigma,\theta} \left(\sum_{n=1}^{N} \xi_n \right),$$

где $\theta = (p_1, p_2)$ -– параметр процесса. Пусть $\Lambda(d\theta)$ есть априорное распределение на множестве параметров Θ. Байесовский риск равен

$$R_N^B(\Lambda) = \inf_{\{\sigma\}} \int_\Theta L_N(\sigma, \theta) \Lambda(d\theta),$$

соответствующая оптимальная стратегия σ^B называется байесовской.

Байесовский подход - 2

Berry, D.A. and Fristedt, B. Bandit Problems: Sequential Allocation of Experiments. Chapman and Hall. London, New York.,1985.

Известен простой рекуррентный алгоритм определения байесовских стратегии и риска. Как пишут Берри и Фристедт: "...it is not that researchers in bandit problems tend to "Bayesians"; rather Bayes's theorem provides a convenient mathematical formalism that allows for adaptive learning and so is an ideal tool in sequential decision problems". Для вычисления риска надо последовательно решать уравнение

$$R_n^B(\Lambda) = \min_{\ell=1,2} \mathbb{E}_\Lambda \left((p_{\overline{\ell}} - p_\ell)^+ + \mathbb{E}_x^{(\ell)} R_{n-1}^B(\Lambda(y_1 = \ell, \xi_1 = x)) \right),$$

при этом байесовская стратегия – та, которая обеспечивает решение этого уравнения.Цветом выделены минимальные полные потери при условии выбора на 1-ом шаге ℓ-ого действия.
Здесь $x^+ = x \vee 0$, $\overline{\ell} = 3 - \ell$, $\mathbb{E}_x^{(\ell)} R(\ell, x) = q_\ell R(\ell, 0) + p_\ell R(\ell, 1)$.

Адаптивное обучение и байесовский формализм

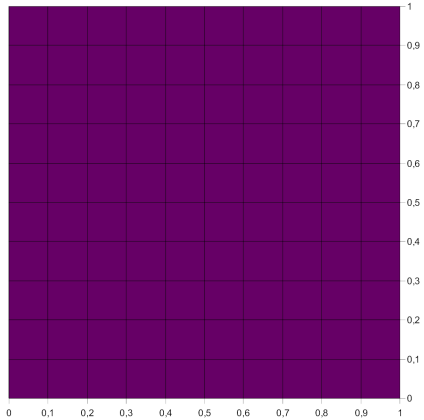
Адаптация = Идентификация + Управление

$$R_n^B(\Lambda) = \min_{\ell=1,2} \mathbb{E}_\Lambda \left((p_{\overline{\ell}} - p_\ell)^+ + \mathbb{E}_x^{(\ell)} R_{n-1}^B(\Lambda(y_1 = \ell, \xi_1 = x)) \right).$$

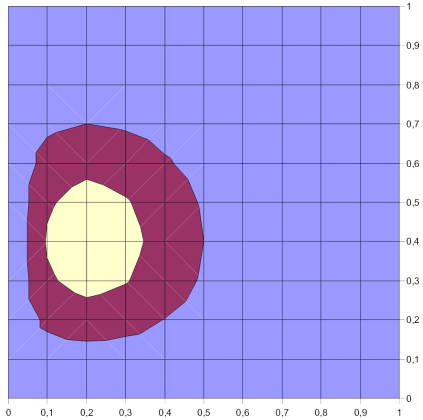
Цветом выделены части уравнения, обеспечивающие идентификацию и управление.

Априорное и апостериорное распределения

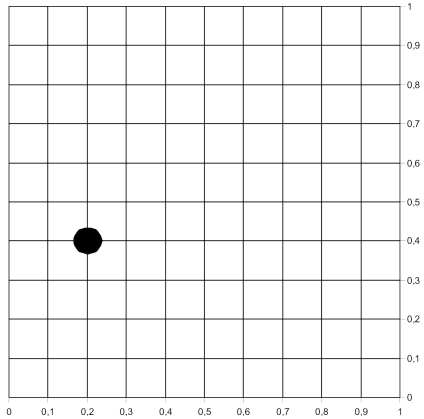
Равномерное априорное распределение



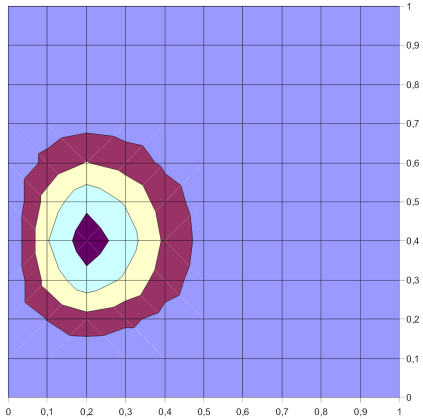
Апостериорные распределения



Фактическое значение параметра



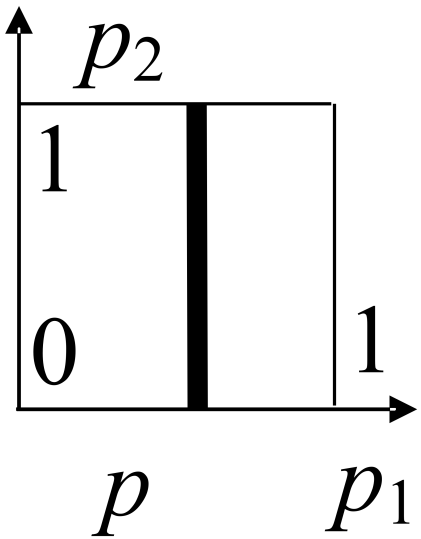
$m_1 = 1,\ n_1 = 5,\ m_2 = 2,\ n_2 = 5$



$m_1 = 2,\ n_1 = 10,\ m_2 = 4,\ n_2 = 10$

Одна известная вероятность

Bradt, R.N., Johnson, S.M., and Karlin, S. (1956). On sequential designs for maximizing the sum of *n* observations. Ann. Math. Stat., V. 27, 1060-1074.



В этом случае $\Theta = \{\theta : \theta = (p, x); x \in [0, 1]\}$.

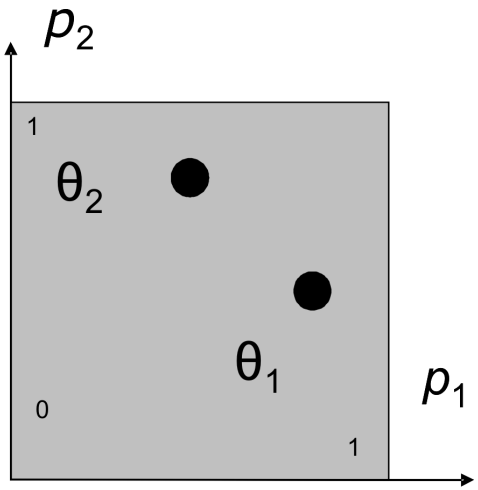
Основная идея

Выбор первой рукоятки не меняет информацию, известную игроку. Поэтому если первая рукоятка однажды будет выбрана, то она будет выбираться до конца управления.

Следовательно, оптимальная стратегия предписывает выбирать вторую рукоятку на некоторой начальной стадии управления. а затем переключается на выбор первой рукоятки до конца управления.

Близорукая стратегия Фельдмана

Feldman, D. Contributions to the "Two–Armed Bandit" Problem. Ann. Math. Stat. 1962. V. 33. P. 847–856.

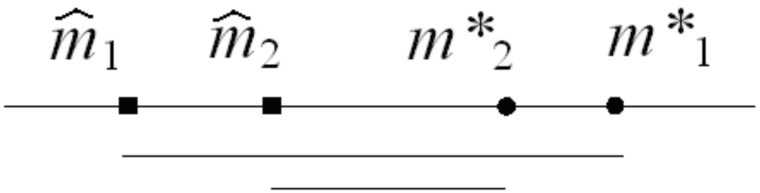


В этом случае $\Theta = \{\theta_1, \theta_2\}$, $\theta_1 = (p_1, p_2)$, $\theta_2 = (p_2, p_1)$, $p_1 > p_2$, $\lambda(\theta_1) = \lambda_1$, $\lambda(\theta_2) = \lambda_2$, $\lambda_1 + \lambda_2 = 1$.

- Выбрать 1-ое действие, если больше текущая апостериорная вероятность θ_1
- Выбрать 2-ое действие, если больше текущая апостериорная вероятность θ_2

Асимптотическая байесовская теорема - 1

1. Lai, T.L. and Robbins, H. Asymptotically Efficient Adaptive Allocation Rules. Advances in Applied Mathematics, 1985, V. 6, P. 4-22.
2. Lai, T.L. Adaptive treatment allocation and the multi-armed bandit problem. The Annals of Statist., 1987, V. 25, P.1091-1114.



Предложены стратегии, выбирающие на каждом шаге действие, которому соответствует бóльшее значение верхней границы доверительного интервала оценки параметра. Ширина интервала тем больше, чем меньше применялось данное действие. В случае нормально распределенных доходов с единичными дисперсиями и мат. ожиданиями m_1, m_2

$$m^*_\ell = \hat{m}_\ell + \left(\frac{\ln(N/n_\ell)}{2n_\ell}\right)^{1/2}, \quad \hat{m}_\ell = \frac{\sum_{y_{n_i}=\ell} \xi_{n_i}}{n_\ell},$$

n_ℓ – число применений ℓ-ого действия.

Асимптотическая байесовская теорема - 2

При широких предположениях при $N \to \infty$ установлены оценки

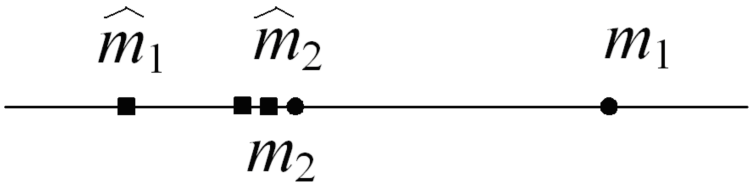
$$L_N(\sigma^B, \theta) \sim \frac{|m_1 - m_2| \ln N}{I(m_1, m_2)} = \frac{2 \ln N}{|m_1 - m_2|}, \quad R_N^B(\Lambda) \propto \ln^2 N.$$

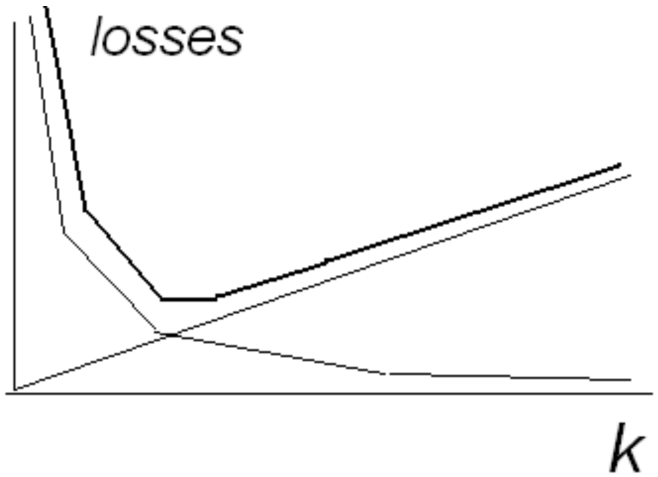
Здесь $\theta = (m_1, m_2)$, $f(x|m) = (2\pi)^{-1/2} \exp(-(x-m)^2/2))$,

$$I(m_1, m_2) = \int_{-\infty}^{\infty} \ln \left(\frac{f(x|m_1)}{f(x|m_2)} \right) f(x|m_1) dx = \frac{(m_1 - m_2)^2}{2}$$

– информационное число Куллбака-Лайблера (Kullback-Leibler).

Поясним первую оценку - 1





Пусть известны m_1, m_2, но неизвестно их соответствие действиям.

Если оба действия сначала применялись по $k \ll N$ раз и за них были получены доходы X_1, X_2, а затем применялось только то действие, которому соответствовал больший начальный доход, то

$$L_N(\sigma^B, \theta) \sim (m_1 - m_2)\{k + P_{err} \times (N - 2k)\} \qquad (1),$$

где P_{err} – вероятность ошибки, равная $(\mathbb{E}(X_1 - X_2) = k(m_1 - m_2))$

$$\Pr[X_1 - X_2 < 0] = F_N\left(-\frac{k(m_1 - m_2)}{(2k)^{1/2}}\right) < \exp\left(-\frac{k(m_1 - m_2)^2}{4}\right)$$

Поясним первую оценку - 2

Поэтому второе слагаемое в (1) при $k = 4\alpha \ln N/(m_1 - m_2)^2$ приблизительно равно $(m_1 - m_2)N \cdot N^{-\alpha}$ и $\ll \ln N$ при $\alpha \approx 1$. Тогда первое слагаемое в (1) дает

$$(m_1 - m_2)k = 4 \ln N/(m_1 - m_2),$$

т.е. порядок тот же, что и в оценке Lai-Robbins.

A problem of two populations

Robbins, H. Some aspects of the sequential design of experiments. Bulletin of Amer. Math. Soc., 1952, V. 58, P.527-535.

"... In what follows we shall discuss a few simple problems in sequential design which are now under investigation and which are different from those usually met with in statistical literature. Optimum solutions to these problems are not known. Still, it is often better to have reasonably good solutions of the proper problems than optimum solutions of the wrong problems. In the present state of statistical theory this principle applies with particular force to problems in sequential design."

"... It would be interesting to know the value of $\phi(N)$[1] and the explicit description of any "minimax" rule R for which the value $\phi(N)$ is attained."

[1] $\phi(N) = N^{-1} R_N^M(\Theta)$, $R_N^M(\Theta)$ – minimax risk.

Минимаксные риск и стратегия

1. Vogel, W. An asymptotic minimax theorem for the two-armed bandit problem. Ann. Math. Stat., 1961, V. 31, P.444-451
2. Fabius, J., and van Zwet, W.R. Some remarks on the two-armed bandit. Ann. Math. Stat., 1970, V. 41, 1906 -1916.

Минимаксный риск равен:

$$R_N^M(\Theta) = \inf_{\{\sigma\}} \sup_{\Theta} L_N(\sigma, \theta),$$

соответствующая оптимальная стратегия σ^M называется минимаксной[2]. Прямое нахождение минимаксных стратегии и риска практически невозможны. Как пишут Фабиус и Ван Цвет: “the algebra involved becomes progressively more complicated with increasing N and seems to remain prohibitive already for N as small as 5”. Однако известна асимптотическая минимаксная теорема Фогеля, гласящая, что при $N \to \infty$, $D = 0,25$:

$$0,530 \leq (DN)^{-1/2} R_N^M(\Theta) \leq 0,752.$$

[2] $\phi(N) = N^{-1} R_N^M(\Theta)$

Некоторые свойства минимаксного подхода

Робастность

При применении стратегии σ^M выполнено неравенство

$$L_N(\sigma^M, \theta) \leq R_N^M(\Theta), \quad \forall \theta \in \Theta.$$

Невозможность прямого определения

Как пишут Fabius и van Zwet по поводу бернуллиевского двурукого бандита: “the algebra involved becomes progressively more complicated with increasing N and seems to remain prohibitive already for N as small as 5”.

Асимптотическая минимаксная теорема (W.Vogel)

Следующие неравенства выполняются асимптотически при $N \to \infty$

$$0.530 \leq (DN)^{-1/2} R_N^M(\Theta) \leq 0.752.$$

Некоторые свойства байесовского подхода

Простой рекуррентный алгоритм нахождения

Как пишут Berry и Fristedt: "...it is not that researchers in bandit problems tend to "Bayesians"; rather Bayes's theorem provides a convenient mathematical formalism that allows for adaptive learning and so is an ideal tool in sequential decision problems".

Основная теорема теории игр

При нежестких ограничениях минимаксный риск совпадает с байесовским, соответствующим наихудшему априорному распределению, т.е.

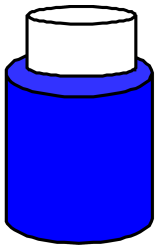
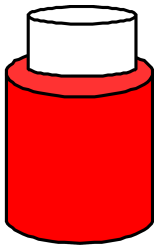
$$R_N^M(\Theta) = \sup_{\{\Lambda\}} R_N^B(\Lambda) = R_N^B(\Lambda^0).$$

Метод нахождения минимаксного риска

В дальнейшем минимаксный риск ищется как байесовский, вычисленный относительно наихудшего априорного распределения.

Двухэтапный подход - 1

1. Cheng, Y. (1994). Multistage decision problems. Sequential Analysis. V. 13, 329-350.
2. Witmer, J.A. (1986). Bayesian multistage decision problems. Ann. Stat. V. 14, 283-297.

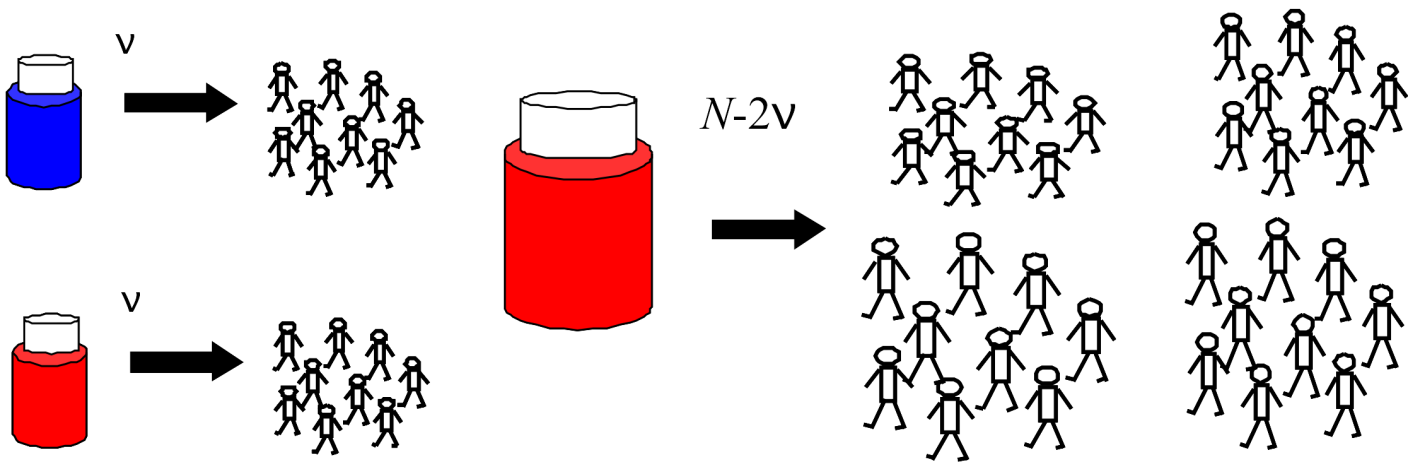
Пусть имеется очень большая группа N пациентов и 2 альтернативных лекарства с различными и неизвестными эффективностями.

Эти лекарства могут рассматриваться как действия, причем p_ℓ, q_ℓ -– вероятности успешного и неуспешного лечения, $\ell = 1, 2$. Значения процесса определяются так: $\xi_n = 1$, если пациент номер n поправился и $\xi_n = 0$, если нет.

Действия в данном случае нельзя применять последовательно один за другим, так как лечение пациента требует значительного времени. В этом случае требуется использовать *параллельную обработку*!

Двухэтапный подход - 2

На 1-ом этапе оба лекарства даются равным достаточно большим группам ν пациентов. В конце 1-ого этапа подсчитываются количества выздоровевших пациентов в обеих группах. Затем более эффективное лекарство дается оставшимся $N - 2\nu$ пациентам на 2-ом этапе. При $N \to \infty$ оптимально выбирать $\nu \propto N^{2/3}$.



Спасибо за внимание

Спасибо за внимание