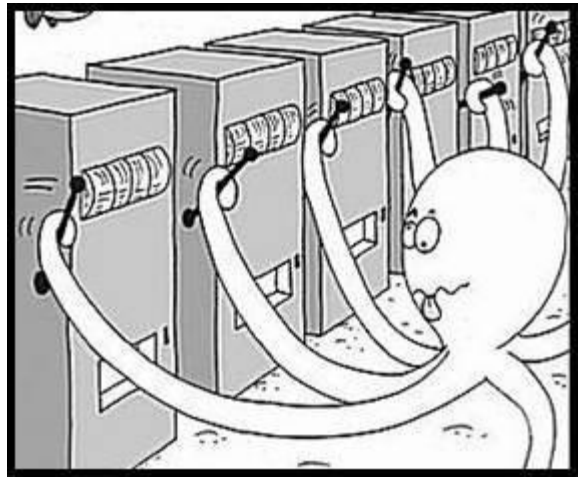

Пуассоновский двурукий бандит

Костюков А.В.

Постановка задачи

- Имеются два разных действия, обычно это представляется как использование ручек у двурукого игрового бандита
- Применение любого действия сопровождается случайным доходом, вероятности получения дохода неизвестны, но фиксированы
- Необходимо выбирать действия так, чтобы доход был максимальным

Пуассоновский двурукий бандит

- Отличается тем, что рассматривается не дискретное время, а непрерывное
- Распределение Пуассона описывает вероятности наступления событий в заданном промежутке времени, если они порождаются простым потоком событий.
- Простой поток событий характеризуется свойствами стационарности, ординарности и отсутствия последействия

$$p(i, t; \lambda) = \frac{(\lambda t)^i}{i!} e^{-\lambda t}$$

Стратегия выбора действия

- Байесовская стратегия состоит в минимизации функции потерь на всём множестве допустимых значений параметра $\Theta = \{\theta_1, \ldots, \theta_K\}$. Функцией потерей является разность между максимальным возможным и реальным полным ожидаемым доходом

$$L_T(\sigma, \theta) = T \max(\lambda_1, \lambda_2) - \mathbb{E}_{\sigma,\theta}(X_1(T) + X_2(T))$$

- Байесовский риск вычисляется по следующей формуле

$$R_T(\overline{\mu}) = \min_{\sigma} \left(\sum_{i=1}^{K} L_T(\sigma, \theta_i) \mu_i \right)$$

Байесовская стратегия

- Апостериорное распределение вероятностей вычисляется при наличии предыстории

$$\mu_i(X_1, t_1, X_2, t_2) = \frac{p(X_1, t_1; \lambda_{1i})p(X_2, t_2; \lambda_{2i})\mu_i}{\mu(X_1, t_1, X_2, t_2)}$$

$$\mu(X_1, t_1, X_2, t_2) = \sum_{i=1}^{K} p(X_1, t_1; \lambda_{1i})p(X_2, t_2; \lambda_{2i})\mu_i$$

- Это классическая Байесовская формула
- Априорное распределение обычно выбирается в ходе экспертной оценки

Байесовская стратегия

- Необходимо на каждом шаге выбирать действие с наименьшим Байесовским риском. Для первой ручки формула риска будет выглядеть так

$$R^{(1)}(X_1,t_1,X_2,t_2) = R(X_1,t_1+\Delta,X_2,t_2)+$$
$$+\sum_{i=1}^{K}\mu_i(X_1,t_1,X_2,t_2)\times((\lambda_{2i}-\lambda_{1i})^{+}+$$
$$+\lambda_{1i}D^{(1)}R(X_1,t_1+\Delta,X_2,t_2))\Delta+o(\Delta),$$

$$D^{(1)}R(X_1,t_1,X_2,t_2) = R(X_1+1,t_1,X_2,t_2)-R(X_1,t_1,X_2,t_2)$$

Спасибо за внимание