

**Колногоров А. В.** (НовГУ им. Ярослава Мудрого, Великий Новгород, Россия). **Задача о гауссовском одноруком бандите с двумя неизвестными параметрами в приложении к пакетной обработке данных.**

Гауссовский двурукий бандит – это управляемый случайный процесс  $\xi_n$ ,  $n = 1, 2, \dots, N$ , значения которого интерпретируются как доходы, зависят только от примененных в текущие моменты времени действий  $y_n$  и имеют нормальные (гауссовские) распределения ( $y_n \in \{1, 2\}$ ). Нормальные распределения доходов являются следствием того, что поступающие данные обрабатываются пакетами и суммарные доходы в пакетах используются для управления. Однорукий бандит – это двурукий бандит, у которого математическое ожидание дохода за применение первого действия  $m_1$  известно и без ограничения общности равно нулю (иначе можно рассмотреть процесс  $\xi_n - m_1$ ,  $n = 1, 2, \dots, N$ ). В случае пакетов больших объемов дисперсии доходов можно считать известными, так как их можно оценить во время обработки начальных пакетов и затем использовать эти оценки для управления. В случае же пакетов умеренных объемов оценивать неизвестные дисперсии следует на протяжении всего процесса управления. Эта ситуация аналогична тому, что в статистике при получении интервальных оценок средних в выборках большого объема используют нормальное распределение, в которое подставляют точечную оценку дисперсии, но в выборках умеренных и малых объемов используют распределение Стьюдента. Математически задача описывается гауссовским одноруким бандитом, характеризуемым параметром  $\theta = (m, D)$ , где  $m, D$  – математическое ожидание и дисперсия дохода за применение второго действия.

В [1] получено рекуррентное уравнение типа уравнения Беллмана для вычисления байесовского риска для однорукого бандита в зависимости от текущей статистики  $(n_1, X, S, n_2)$ , где  $n_1, n_2$  – текущие полные времена применения первого и второго действий,  $X = \sum_{i:i \leq n, y_i=2} \xi_i$ ,  $S = \left( \sum_{i:i \leq n, y_i=2} \xi_i^2 \right) - X^2/n_2$  – текущие полный доход и  $s^2$ -статистика за применение второго действия. При этом для пересчета  $X, S$  используются текущие доходы за применение действия  $y_2$ . Однако для пересчета  $S$  можно использовать и  $s^2$ -статистику, вычисленную по внутренним данным пакетов.

**Теорема 1.** *Рекуррентное уравнение для вычисления байесовского риска в задаче о гауссовском одноруком бандите с двумя неизвестными параметрами допускает обобщение на случай, если для пересчета текущей статистики  $(n_1, X, S, n_2)$  наряду с текущим доходом за применение второго действия используется  $s^2$ -статистика, вычисленная по внутренним данным этого пакета. Уравнение может быть представлено в инвариантной форме на единичном горизонте управления в том смысле, что оно зависит не от полного числа данных и реальных размеров пакетов, а только от относительных размеров пакетов.*

#### СПИСОК ЛИТЕРАТУРЫ

- [1] A.V. Kolnogorov, “Gaussian one-armed bandit with both unknown parameters”, *Сибирские электронные математические известия* **19** (2022), 639–650.