

MINIMAX APPROACH TO THE GAUSSIAN MULTI-ARMED BANDIT

A. V. Kolnogorov

We consider the multi-armed bandit problem in an application to batch processing of big data if there are more than two alternative processing methods with different a priori unknown efficiencies. During processing, it is necessary to determine a more effective method and ensure its preferential use. Control is performed on the basis of cumulative incomes in batches, which, by virtue of the central limit theorem, have approximately Gaussian distributions. An important feature of batch processing is that it almost does not lead to an increase in maximum losses of the total expected income, i.e., to an increase in minimax risk if the numbers of processed data and batches into which the data is divided are large enough. This means that Gaussian multi-armed bandit provides universal approach to optimal control of big data when one-step incomes satisfy the central limit theorem. According to this approach, minimax strategy and risk are sought using the main theorem of game theory as Bayesian ones calculated relative to the worst-case prior distribution at which the Bayesian risk is maximal. For this purpose, the characterization of the worst-case prior distribution is given and recursive equations in a usual and invariant form with a control horizon equal to one are obtained. In the limiting case when the number of processed batches grows infinitely, a second-order partial differential equation is obtained. A numerical example of finding minimax risk and strategy for a 3-armed bandit is given.

Keywords: Gaussian multi-armed bandit, Game with nature, Bayesian and minimax approaches, Main theorem of game theory, Batch processing.

MSC: 91A35, 62L05, 62C10

<https://doi.org/10.21538/0134-4889-2025-31-3-fon-06>

REFERENCES

1. Berry D.A., Fristedt B. *Bandit problems: sequential allocation of experiments*. London, NY, Springer Dordrecht, 1985, 275 p. <https://doi.org/10.1007/978-94-015-3711-7>
2. Lattimore T., Szepesvari C. *Bandit algorithms*. Cambridge, Cambridge Univ. Press, 2020, 518 p. <https://doi.org/10.1017/9781108571401>
3. Presman E.L., Sonin I.M. *Sequential control with incomplete information*. NY, Acad. Press, 1990, 290 p. ISBN-13: 9780125644358. Original Russian text published in Presman E.L., Sonin I.M. *Posledovatel'noye upravleniye po nepolnym dannym*, Moscow, Nauka Publ., 1982, 256 p.
4. Tsetlin M.L. *Automation theory and modeling of biological systems*, vol. 102. NY, Acad. Press, 1973. doi: 10.1016/s0076-5392(08)x6049-7. Original Russian text published in Tsetlin M. L. *Issledovaniya po teorii avtomatov i modelirovaniyu biologicheskikh sistem*, Moscow, Nauka Publ., 1969, 316 p.
5. Sragovich V.G. *Mathematical theory of adaptive control*, vol. 4. NJ, Interdisc. Math. Sci., London, World Sci., 2006, 473 p. <https://doi.org/10.1142/5857>. Original Russian text published in Sragovich V. G. *Adaptivnoye upravleniye*, Moscow, Nauka Publ., 1981, 381 p.
6. Perchet V., Rigollet P., Chassang S., Snowberg E. Batched bandit problems. *Ann. Stat.*, 2016, vol. 44, no. 2, pp. 660–681. <https://doi.org/10.1214/15-aos1381>
7. Slivkins A. *Introduction to multi-armed bandits*. 2019, 188 p. <https://arxiv.org/pdf/1904.07272>
8. Kolnogorov A.V. Determination of the minimax risk for the normal two-armed bandit. *IFAC Proc. Vol.*, 2010, vol. 43, iss. 10, pp. 231–236. <https://doi.org/10.3182/20100826-3-TR-4015.00044>
9. Feldman D. Contributions to the “Two-armed bandit” problem. *Ann. Math. Statist.*, 1962, vol. 33, pp. 847–856. <https://doi.org/10.1214/aoms/1177704454>
10. Zaborskis A.A. A sequential Bayesian plan for choosing the best method of curing. *Autom. Remote Control*, 1976, vol. 37, no. 11, pp. 1750–1757.
11. Rodman L. On the many-armed bandit problem. *Ann. Probab.*, 1978, vol. 6, no. 3, pp. 491–498.
12. Fabius J., van Zwet W.R. Some remarks on the two-armed bandit. *Ann. Math. Statist.*, 1970, vol. 41, no. 6, pp. 1906–1916. <https://doi.org/10.1214/aoms/1177696692>

13. Kolnogorov A.V. On a limiting description of robust parallel control in a random environment. *Autom. Remote Control*, 2015, vol. 76, no. 7, pp. 1229–1241. <https://doi.org/10.1134/S0005117915070085>
14. Lai T.L., Robbins H. Asymptotically efficient adaptive allocation rules. *Adv. Appl. Math.*, 1985, vol. 6, iss. 1, pp. 4–22. [https://doi.org/10.1016/0196-8858\(85\)90002-8](https://doi.org/10.1016/0196-8858(85)90002-8)
15. Vogel W. An asymptotic minimax theorem for the two-armed bandit problem. *Ann. Math. Statist.*, 1960, vol. 31, no. 2, pp. 444–451. <https://doi.org/10.1214/aoms/1177705907>
16. Juditsky A., Nazin A.V., Tsybakov A.B., Vayatis N. Gap-free bounds for stochastic multi-armed bandit. In: *Proc. 17th World Congress IFAC*, Seoul, 2008, vol. 41, iss. 2, pp. 11560–11563. <https://doi.org/10.3182/20080706-5-KR-1001.01959>
17. Ershov M.A., Voroshilov A.S. UCB strategy for Gaussian and Bernoulli multi-armed bandits. In: Khachay M., Kochetov Yu., Ereemeev A., Khamisov O., Mazalov V., Pardalos P. (eds.) *Proc. 22nd Inter. Conf. “Mathematical optimization theory and operations research: recent trends” (MOTOR 2023)*, vol. 1881. Cham, Springer, 2023, pp. 67–78. https://doi.org/10.1007/978-3-031-43257-6_6
18. Russo D., Van Roy B. Learning to optimize via posterior sampling. *Math. Operat. Res.*, 2014, vol. 39, no. 4, pp. 1221–1243. <https://doi.org/10.1287/moor.2014.0650>
19. Kolnogorov A.V. Gaussian two-armed bandit and optimization of batch data processing. *Probl. Inf. Transm.*, 2018, vol. 54, no. 1, pp. 84–100. <https://doi.org/10.1134/S0032946018010076>

Received May 12, 2025

Revised June 16, 2025

Accepted June 23, 2025

Published online June 26, 2025

Funding Agency. 1. The research was conducted as part of the scientific research project “Mathematical Modeling of Natural Processes”, implemented under the state assignment in the field of scientific activity. 2. The study was carried out using the infrastructure of the Shared Use Center “High-Performance Computing and Big Data” (SUC “Informatics”) of the Federal Research Center for Information Technology and Control of the Russian Academy of Sciences (Moscow).

Alexander Valerianovich Kolnogorov, Dr. Phys.-Math. Sci., Prof., Yaroslav-the-Wise Novgorod State University, Veliky Novgorod, 173003 Russia, e-mail: kolnogorov53@mail.ru .

Cite this article as: A. V. Kolnogorov. Minimax approach to the Gaussian multi-armed bandit. *Trudy Instituta Matematiki i Mekhaniki UrO RAN*, 2025, vol. 31, no. 3, pp. 150–166 .