

SEQUENTIAL CONTROL WITH INCOMPLETE INFORMATION

THE BAYESIAN APPROACH TO
MULTI-ARMED BANDIT PROBLEMS

E. L. PRESMAN

I. N. SONIN

Academy of Sciences of the USSR
Central Economic Mathematical Institute
Moscow, USSR

Translated and Edited by

E. A. MEDOVA-DEMPSTER

Department of Applied
Mathematics
Technical University of
Nova Scotia
Halifax, Nova Scotia, Canada

M. A. H. DEMPSTER

Department of Mathematics
Statistics and
Computing Science
and
School of Business
Administration
Dalhousie University
Halifax, Nova Scotia, Canada
and
Balliol College
Oxford, England

This is a volume in
ECONOMIC THEORY, ECONOMETRICS,
AND MATHEMATICAL ECONOMICS

A Series of Monographs and Textbooks

Consulting Editor: KARL SHELL

A complete list of titles in this series appears at the end of this volume.




ACADEMIC PRESS

Harcourt Brace Jovanovich, Publishers
London San Diego New York
Boston Sydney Tokyo Toronto

ACADEMIC PRESS LIMITED
24/28 Oval Road,
LONDON NW1 7DX

United States Edition published by
ACADEMIC PRESS INC.
San Diego, CA92101

Copyright © 1990
ACADEMIC PRESS LIMITED

This book is printed on acid-free paper 

All Rights Reserved

No part of this book may be reproduced in any form by photostat, microfilm, or
any other means without written permission from the publishers

British Library Cataloguing in Publication Data
is available

ISBN 0-12-564435-3

Printed by St Edmundsbury Press Limited,
Bury St Edmunds, Suffolk

CONTENTS

Foreword	ix
Preface to the English Edition	xi
Translators' Preface	xv
Introduction	xix
Some Notation	xxv
1. Basic Schemes in Discrete and Continuous Time	1
1.1 Formulation of the sequential control problem with incomplete information	1
1.2 Basic schemes in discrete time	6
1.3 Relation of the basic scheme to the general theory of sequential control with incomplete information	8
1.4 Evolution of <i>a posteriori</i> probabilities in discrete time	10
1.5 Finite Bayesian problem	12
1.6 Transition to the continuous time case	13
1.7 Basic scheme in continuous time	16
1.8 Local optimality equation	20
1.9 Reduction to the Pontryagin problem	22
1.10 Linear control problems with Poisson jumps	25
1.11 Results obtained	27

2. Problem Formulation and Solution Methods in Discrete Time	35
2.1 Formulation of the basic scheme as a control problem	35
2.2 General problems of sequential control with incomplete information	48
2.3 Existence of an optimal action rule, coincidence of $F_\infty(\xi)$ and $F(\xi)$ and convexity of $F_\nu(\xi)$	57
2.4 Optimality equation and optimal strategies	66
2.5 Evolution of <i>a posteriori</i> probabilities	86
3. Solutions of Some Problems in the Basic Discrete Time Scheme	95
3.1 F -matrices and B -matrices	96
3.2 Loss for an F -matrix on an infinite time interval	98
3.3 Loss for a B -matrix on an infinite time interval	107
3.4 Optimal strategies for the case $m = N = 2$	109
3.5 Scheme with sharable resources	117
4. Problem Formulation and Solution Methods in Continuous Time	125
4.1 Reduction of continuous to discrete time	125
4.2 Basic scheme problem statement	126
4.3 Existence of an optimal action rule	135
4.4 Reduction to a discrete time problem and existence of a Markov uniformly optimal strategy in the Markov case	145
4.5 Local optimality equation and optimal synthesis	152
5. Solutions of Some Problems in the Basic Continuous Time Scheme	157
5.1 F -matrices and B -matrices	157
5.2 Minimization of loss over an infinite horizon for the case $m = N = 2$	158
5.3 Minimization of loss over a finite horizon for the case $m = N = 2$	170

6. Application of Pontryagin's Maximum Principle to Control Problems with Random Jumps	189
6.1 The linear control problem with Poisson jumps	189
6.2 Reduction of the initial formulation to a Pontryagin type problem and description of results	192
6.3 The problem $B(q)$	195
6.4 Formula for the derivative of the Hamiltonian along trajectories	200
6.5 Optimal syntheses for the case of symmetric hypotheses	204
6.6 Problems with an infinite number of jumps	213
7. Some Other Problems	215
7.1 Description of main results	215
7.2 Discrimination of hypotheses	217
7.3 Maximization of first jump probability	223
7.4 Minimax formulation of the multi-armed bandit problem	230
7.5 A price control model	235
7.6 Other formulations of sequential control problems with incomplete information: Unsolved and new problems	236
Appendix	243
References	249
Further References	257
Index	263

FOREWORD

This book is devoted to a specific problem in the general theory of optimal control—sequential control under conditions of incomplete information. The main results concern the case in which at each moment of (continuous or discrete) time only a finite number of controls are admissible and the results of control action conducted in a Bayesian framework are represented by realizations of random variables whose distributions for a given control correspond to one of several alternative hypotheses.

This situation is related to problems in the sequential distribution of resources with incomplete information, problems in the sequential setting of prices in the face of random demand, search problems, etc. Similar problems are found in the general theory of statistical decisions and in the theory of planning experiments—under the name of *multi-armed bandit* problems—and in the theory of automatic control—as problems of *dual control*.

The book is suitable for specialists in applied mathematics, probability theory, statistics, mathematical economics and control theory.

Editors of Nauka, Moscow, 1982 edition:
V. I. Arkin, Yu. M. Kabanov.

Preface to the English Edition

The Russian edition of this book was published in 1982, but its title does not contain the words “multi-armed bandit” simply because there were at that time no one-armed bandit machines in our country. Nevertheless, we use this term in our earlier papers.

With the publication of the English edition of our book (PS), there exist three books devoted to sequential allocation of resources which contain in their titles the term “bandit”. The two others are: D.A. Berry & Ø. Fristedt (1985), *Bandit Problems: Sequential Allocation of Experiments* (BF) and J.C. Gittins (1989), *Multi-armed Bandit Allocation Indices* (G).

B. (Bert)

Let us try to explain briefly the fact—at first glance peculiar—that while all three books have equal right to use the term “bandit”, the intersection of the contents of (PS) and (G) is almost empty and the intersection of (PS) and (BF) is rather small, as by the way is the intersection of (BF) and (G).

According to the introduction to (BF), “A bandit problem in statistical decision theory involves sequential selections from $k \geq 2$ stochastic processes (or ‘arms’, machines, treatments, etc.). Time may be discrete or continuous and the processes themselves may be discrete or continuous. The processes are characterized by parameters which are typically unknown. The process selected for observation at any time depends on the previous selections and results...” (p. 1).

This quotation sets the general framework for all three books. The approaches to these problems in (PS) and (BF) have much in common in that they both use mainly the Bayesian approach and for the discrete time case consider Bernoulli processes (principally in (BF), and

almost exclusively, in (PS)). Both books also consider the continuous time case, but in different ways. Wiener and Lévy processes are considered, following Chernoff, in one chapter of (BF). In (PS), sequences of Bernoulli trials in discrete time are replaced by Poisson processes and a substantial part of this book is devoted to the continuous time case. Much attention is paid in (BF) to different ways of discounting and their approach to this subject is more general than that of other works.

Now let us turn to the main differences between the three books without detailing the contents of (BF) and (G).

According to (BF), "Most of [their] monograph treats *independent arms*" (p. 6). For our purposes here, independent arms can be taken to mean simply that at any time information is obtained only about the parameters of the arm in use. Quite the opposite situation is treated in (PS). Most of the concrete results (e.g. finding optimal strategies) in the present book deal with *dependent arms*, when this dependence is essential. The most typical case of dependent arms is the symmetric "Feldman case" in which there are two Bernoulli arms with parameters (θ_1, θ_2) equal to either (a, b) or (b, a) .

More generally, this book considers the case when the distribution on the space of parameters is concentrated on a finite number of points, termed hypotheses, so that the whole situation may be described by a matrix (λ_i^j) , where λ_i^j is the probability of a success on the j^{th} arm when hypothesis i is true. One of the main problems investigated is the problem of the asymptotic behaviour of the (optimal) value function for different matrices. It is worth mentioning here that in the Feldman case "losses" (the difference between ideal and achieved results) are finite. By contrast, in the situation first considered by Bellman (1956), in which the parameters of one arm are known, losses are infinite. The notion of loss allows the treatment of the undiscounted infinite horizon problem which, unlike (BF) and (G), is one of central interest in (PS). Much attention is also paid to finding explicitly optimal strategies and value functions and to the asymptotic behaviour of *a posteriori* probabilities.

The whole book (G) is in some sense the generalization of independent Bernoulli processes to much more general objects—semi-Markov decision processes. But, as in the main part of (BF), according to

the introduction of (G), "problems are characterized by alternative *independent* ways in which time or effort may be consumed." For this case, by the efforts of Gittins and others, the wonderful result is proved that for every arm there exists an index depending on its past history, but not on that of any other project, and at each decision time it is optimal to allocate effort only to the project with the highest current index value. The characterization of these indices (Gittins' indices—dynamic allocation indices) is also given.

We could conclude our brief overview of the contents of the three books here if it were not for the following quotation from Professor Whittle's foreword to (G).

The multi-armed bandit is a prototype of [a] class of problems propounded during the Second World War and soon recognized as so difficult that it quickly became a classic, and a byword for intransigence. In fact, John Gittins had solved the problem by the late sixties, although the fact that he had done so was not generally recognized until the early eighties. I can illustrate the mode of propagation of this news, when it began to propagate, by telling of an American friend of mine, a colleague of high repute, who asked an equally well-known colleague, "What would you say if you were told that the multi-armed bandit problem had been solved?" The reply was somewhat in the Johnsonian form: "Sir, the multi-armed bandit problem is not of such a nature that it can be solved." My friend then undertook to convince the doubter in a quarter of an hour. This is indeed a feature of John's solution: that, once explained, it carries conviction even before it is proved.

We find it difficult to completely agree with this citation. The case of two symmetric arms, solved in the profound paper of Feldman (1962), is a classical problem of the bandit type—as it was called for instance in the pioneering paper of Brandt, Johnson & Karlin (1956)—involving *dependent* processes. The solution, though simple, is not in any way trivial. Moreover, the statement seems obvious only at first glance, when one is not acquainted with results concerning the best discrimination of hypotheses for this case. The Feldman case is referred to as the "two-armed bandit" problem in numerous papers, but it does not belong to the case which Whittle calls the "many-armed bandit".

So in accordance with the above remark, but with due regard for the brilliant results of Gittins and other scientists working in this area, we would like to suggest that—at least for us—the initial position of the second well-known American colleague is the more appropriate.

Indeed, we consider our book, and the whole area, far from completeness, for there are many open problems and new formulations and applications.

In the English edition of this book we have taken the opportunity to eliminate some misprints, improve some proofs, and delete some material we felt inappropriate to the present edition. We retain, of course, responsibility for any remaining errors.

The reference list has also been marginally revised, but it is still far from complete. Many references may be found in (G) and especially in (BF), which contains a very comprehensive annotated bibliography with more than 200 entries and thorough comments.

Finally, we would like to note that by our reasoning the translation of our book into English was not an easy task. We take this opportunity to express our appreciation to the translators for their painstaking work.

Moscow
August 1989

E.L.P.
I.M.S.

Translators' Preface

The English edition of this book has been prepared from the Russian edition (published by Nauka, Moscow, 1982) in collaboration with the authors. Some deletions and many corrections, alterations and additions have been made to the Russian text. In particular, recent work on the discounted case is described in Chapter 7 and the references are augmented accordingly. No attempt has been made to translate the Russian text literally, but on the other hand, no serious attempt has been made to change or standardize the notation. Nevertheless, it is hoped that the result will be a monograph useful to graduate students and researchers in a number of fields. Notionally, the prerequisite to reading this book is basic probability theory (e.g. Shiryaev, 1984), but there is little doubt that familiarity with modern stochastic process theory (to which references are given in the Appendix) will aid the reader's comprehension. However, "the extensive literature showing that the multi-armed bandit and its variants can be used to model the decision problems in job scheduling, resource allocation, sequential random sampling, clinical trials, investment in new products, random search, etc." (Varaiya *et al.* 1985) indicates that careful study of this work might be expected to repay the effort.

By agreement with the authors it has been left to us to provide a collection of further references—mainly appearing in the literature since the publication of the Russian edition—which relate to extensions of multi-armed bandit problems, the wider literature on stochastic control with complete and incomplete information, and applications. Briefly, these may be classified as follows.

Following the seminal work of Gittins & Jones (1974) on multi-