
Probability

Leo Breiman

C · L · A · S · S · I · C · S

In Applied Mathematics

siam

7



Probability



Leo Breiman
University of California, Berkeley

siam

Society for Industrial and Applied Mathematics
Philadelphia

Copyright ©1992 by the Society for Industrial and Applied Mathematics.

This SIAM edition is an unabridged, corrected republication of the work first published by Addison-Wesley Publishing Company, Inc., Reading, Massachusetts, 1968.

10 9 8 7 6

All rights reserved. Printed in the United States of America. No part of this book may be reproduced, stored, or transmitted in any manner without the written permission of the publisher. For information, write to the Society for Industrial and Applied Mathematics, 3600 University City Science Center, Philadelphia, PA 19104-2688.

Library of Congress Cataloging-in-Publication Data

Breiman, Leo

Probability / Leo Breiman.

p. cm. -- (Classics in applied mathematics ; 7)

Originally published: Reading, Mass. : Addison-Wesley Pub. Co.,
1968. (Addison-Wesley series in statistics)

Includes bibliographical references and index.

ISBN 0-89871-296-3

I. Probabilities. I. Title. II. Series.

QA273.B864 1992

519.2--dc20

92-1381

SIAM is a registered trademark.

Classics in Applied Mathematics (continued)

Cornelius Lanczos, *Linear Differential Operators*

Richard Bellman, *Introduction to Matrix Analysis, Second Edition*

Beresford N. Parlett, *The Symmetric Eigenvalue Problem*

Richard Haberman, *Mathematical Models: Mechanical Vibrations, Population Dynamics, and Traffic Flow*

Peter W. M. John, *Statistical Design and Analysis of Experiments*

Tamer Başar and Geert Jan Olsder, *Dynamic Noncooperative Game Theory, Second Edition*

Emanuel Parzen, *Stochastic Processes*

Petar Kokotović, Hassan K. Khalil, and John O'Reilly, *Singular Perturbation Methods in Control: Analysis and Design*

Jean Dickinson Gibbons, Ingram Olkin, and Milton Sobel, *Selecting and Ordering Populations: A New Statistical Methodology*

James A. Murdock, *Perturbations: Theory and Methods*

Ivar Ekeland and Roger Témam, *Convex Analysis and Variational Problems*

Ivar Stakgold, *Boundary Value Problems of Mathematical Physics, Volumes I and II*

J. M. Ortega and W. C. Rheinboldt, *Iterative Solution of Nonlinear Equations in Several Variables*

David Kinderlehrer and Guido Stampacchia, *An Introduction to Variational Inequalities and Their Applications*

F. Natterer, *The Mathematics of Computerized Tomography*

Avinash C. Kak and Malcolm Slaney, *Principles of Computerized Tomographic Imaging*

R. Wong, *Asymptotic Approximations of Integrals*

O. Axelsson and V. A. Barker, *Finite Element Solution of Boundary Value Problems: Theory and Computation*

David R. Brillinger, *Time Series: Data Analysis and Theory*

Joel N. Franklin, *Methods of Mathematical Economics: Linear and Nonlinear Programming, Fixed-Point Theorems*

Philip Hartman, *Ordinary Differential Equations, Second Edition*

Michael D. Intriligator, *Mathematical Optimization and Economic Theory*

Philippe G. Ciarlet, *The Finite Element Method for Elliptic Problems*

Jane K. Cullum and Ralph A. Willoughby, *Lanczos Algorithms for Large Symmetric Eigenvalue Computations, Vol. I: Theory*

M. Vidyasagar, *Nonlinear Systems Analysis, Second Edition*

Robert Mattheij and Jaap Molenaar, *Ordinary Differential Equations in Theory and Practice*

Shanti S. Gupta and S. Panchapakesan, *Multiple Decision Procedures: Theory and Methodology of Selecting and Ranking Populations*

Eugene L. Allgower and Kurt Georg, *Introduction to Numerical Continuation Methods*

Preface to the Classic Edition

This is the first of four books I have written; the one I worked the hardest on; and the one I am fondest of. It marked my goodbye to mathematics and probability theory. About the time the book was written, I left UCLA to go into the world of applied statistics and computing as a full-time freelance consultant.

The book went out of print well over ten years ago, but before it did a generation of statisticians, engineers, and mathematicians learned graduate probability theory from its pages. Since the book became unavailable, I have received many calls asking where it could be bought and then for permission to copy part or all of it for use in graduate probability courses.

These reminders that the book was not forgotten saddened me and I was delighted when SIAM offered to republish it in their Classics Series. The present edition is the same as the original except for the correction of a few misprints and errors, mainly minor.

After the book was out for a few years it became commonplace for a younger participant at some professional meeting to lean over toward me and confide that he or she had studied probability out of my book. Lately, this has become rarer and the confiders older. With republication, I hope that the age and frequency trends will reverse direction.

Leo Breiman
University of California, Berkeley
January, 1992

Preface

A few years ago I started a book by first writing a very extensive preface. I never finished that book and resolved that in the future I would write first the book and then the preface. Having followed this resolution I note that the result is a desire to be as brief as possible.

This text developed from an introductory graduate course and seminar in probability theory at UCLA. A prerequisite is some knowledge of real variable theory, such as the ideas of measure, measurable functions, and so on. Roughly, the first seven chapters of *Measure Theory* by Paul Halmos [64] is sufficient background. There is an appendix which lists the essential definitions and theorems. This should be taken as a rapid review or outline for study rather than as an exposition. No prior knowledge of probability is assumed, but browsing through an elementary book such as the one by William Feller [59, Vol. I], with its diverse and vivid examples, gives an excellent feeling for the subject.

Probability theory has a right and a left hand. On the right is the rigorous foundational work using the tools of measure theory. The left hand “thinks probabilistically,” reduces problems to gambling situations, coin-tossing, motions of a physical particle. I am grateful to Michel Loève for teaching me the first side, and to David Blackwell, who gave me the flavor of the other.

David Freedman read through the entire manuscript. His suggestions resulted in many substantial revisions, and the book has been considerably improved by his efforts. Charles Stone worked hard to convince me of the importance of analytic methods in probability. The presence of Chapter 10 is largely due to his influence, and I am further in his debt for reading parts of the manuscript and for some illuminating conversations on diffusion theory.

Of course, in preparing my lectures, I borrowed heavily from the existing books in the field and the finished product reflects this. In particular, the books by M. Loève [108], J. L. Doob [39], E. B. Dynkin [43], and K. Ito and H. P. McKean [76] were significant contributors.

Two students, Carl Maltz and Frank Kontrovich, read parts of the manuscript and provided lists of mistakes and unreadable portions. Also, I was blessed by having two fine typists, Louise Gaines and Ruth Goldstein, who rose above mere patience when faced with my numerous revisions of the “final draft.” Finally, I am grateful to my many nonmathematician friends who continually asked when I was going to finish “that thing,” in voices that could not be interminably denied.

Leo Breiman
Topanga, California
January, 1968

*To my mother and father
and Tuesday's children*

Contents

Chapter 1 Introduction

- 1 n independent tosses of a fair coin 1
- 2 The “law of averages” 1
- 3 The bell-shaped curve enters (fluctuation theory) 7
- 4 Strong form of the “law of averages” 11
- 5 An analytic model for coin-tossing 15
- 6 Conclusions 17

Chapter 2 Mathematical Framework

- 1 Introduction 19
- 2 Random vectors 20
- 3 The distribution of processes 21
- 4 Extension in sequence space 23
- 5 Distribution functions 25
- 6 Random variables 29
- 7 Expectations of random variables 31
- 8 Convergence of random variables 33

Chapter 3 Independence

- 1 Basic definitions and results 36
- 2 Tail events and the Kolmogorov zero-one law 40
- 3 The Borel-Cantelli lemma 41
- 4 The random signs problem 45
- 5 The law of pure types 49
- 6 The law of large numbers for independent random variables 51
- 7 Recurrence of sums 53
- 8 Stopping times and equidistribution of sums 58
- 9 Hewitt-Savage zero-one law 63

Chapter 4 Conditional Probability and Conditional Expectation

- 1 Introduction 67
- 2 A more general conditional expectation 73
- 3 Regular conditional probabilities and distributions 77

Chapter 5 Martingales

- 1 Gambling and gambling systems 82
- 2 Definitions of martingales and submartingales 83
- 3 The optional sampling theorem 84
- 4 The martingale convergence theorem 89
- 5 Further martingale theorems 91
- 6 Stopping times 95
- 7 Stopping rules 98
- 8 Back to gambling 101

Chapter 6 Stationary Processes and the Ergodic Theorem

- 1 Introduction and definitions 104
- 2 Measure-preserving transformations 106
- 3 Invariant sets and ergodicity 108
- 4 Invariant random variables 112
- 5 The ergodic theorem 113
- 6 Converses and corollaries 116
- 7 Back to stationary processes 118
- 8 An application 120
- 9 Recurrence times 122
- 10 Stationary point processes 125

Chapter 7 Markov Chains

- 1 Definitions 129
- 2 Asymptotic stationarity 133
- 3 Closed sets, indecomposability, ergodicity 135
- 4 The countable case 137
- 5 The renewal process of a state 138
- 6 Group properties of states 141
- 7 Stationary initial distributions 143
- 8 Some examples 145
- 9 The convergence theorem 150
- 10 The backward method 153

Chapter 8 Convergence in Distribution and the Tools Thereof

- 1 Introduction 159
- 2 The compactness of distribution functions 160
- 3 Integrals and \mathcal{D} -convergence 163
- 4 Classes of functions that separate 165
- 5 Translation into random-variable terms 166
- 6 An application of the foregoing 167
- 7 Characteristic functions and the continuity theorem 170

- 8 The convergence of types theorem 174
- 9 Characteristic functions and independence 175
- 10 Fourier inversion formulas 177
- 11 More on characteristic functions 179
- 12 Method of moments 181
- 13 Other separating function classes 182

Chapter 9 The One-Dimensional Central Limit Problem

- 1 Introduction 185
- 2 Why normal? 185
- 3 The nonidentically distributed case 186
- 4 The Poisson convergence 188
- 5 The infinitely divisible laws 190
- 6 The generalized limit problem 195
- 7 Uniqueness of representation and convergence 196
- 8 The stable laws 199
- 9 The form of the stable laws 200
- 10 The computation of the stable characteristic functions 204
- 11 The domain of attraction of a stable law 207
- 12 A coin-tossing example 213
- 13 The domain of attraction of the normal law 214

Chapter 10 The Renewal Theorem and Local Limit Theorem

- 1 Introduction 216
- 2 The tools 216
- 3 The renewal theorem 218
- 4 A local central limit theorem 224
- 5 Applying a Tauberian theorem 227
- 6 Occupation times 229

Chapter 11 Multidimensional Central Limit Theorem and Gaussian Processes

- 1 Introduction 233
- 2 Properties of \mathcal{N}_k 234
- 3 The multidimensional central limit theorem 237
- 4 The joint normal distribution 238
- 5 Stationary Gaussian process 241
- 6 Spectral representation of stationary Gaussian processes 242
- 7 Other problems 246

Chapter 12 Stochastic Processes and Brownian Motion

- 1 Introduction 248
- 2 Brownian motion as the limit of random walks 251
- 3 Definitions and existence 251
- 4 Beyond the Kolmogorov extension 254
- 5 Extension by continuity 255
- 6 Continuity of Brownian motion 257
- 7 An alternative definition 259
- 8 Variation and differentiability 261
- 9 Law of the iterated logarithm 263
- 10 Behavior at $t = \infty$ 265
- 11 The zeros of $X(t)$ 267
- 12 The strong Markov property 268

Chapter 13 Invariance Theorems

- 1 Introduction 272
- 2 The first-exit distribution 273
- 3 Representation of sums 276
- 4 Convergence of sample paths of sums to Brownian motion paths 278
- 5 An invariance principle 281
- 6 The Kolmogorov-Smirnov statistics 283
- 7 More on first-exit distributions 287
- 8 The law of the iterated logarithm 291
- 9 A more general invariance theorem 293

Chapter 14 Martingales and Processes with Stationary, Independent Increments

- 1 Introduction 298
- 2 The extension to smooth versions 298
- 3 Continuous parameter martingales 300
- 4 Processes with stationary, independent increments 303
- 5 Path properties 306
- 6 The Poisson process 308
- 7 Jump processes 310
- 8 Limits of jump processes 312
- 9 Examples 316
- 10 A remark on a general decomposition 318

Chapter 15 Markov Processes, Introduction and Pure Jump Case

- 1 Introduction and definitions 319
- 2 Regular transition probabilities 320
- 3 Stationary transition probabilities 322
- 4 Infinitesimal conditions 324
- 5 Pure jump processes 328
- 6 Construction of jump processes 332
- 7 Explosions 336
- 8 Nonuniqueness and boundary conditions 339
- 9 Resolvent and uniqueness 340
- 10 Asymptotic stationarity 344

Chapter 16 Diffusions

- 1 The Ornstein-Uhlenbeck process 347
- 2 Processes that are locally Brownian 351
- 3 Brownian motion with boundaries 352
- 4 Feller processes 356
- 5 The natural scale 358
- 6 Speed measure 362
- 7 Boundaries 365
- 8 Construction of Feller processes 370
- 9 The characteristic operator 375
- 10 Uniqueness 379
- 11 $\varphi_+(x)$ and $\varphi_-(x)$ 383
- 12 Diffusions 385

Appendix: On Measure and Function Theory 391

Bibliography 405

Index 412

CHAPTER 1

INTRODUCTION

A good deal of probability theory consists of the study of limit theorems. These limit theorems come in two categories which we call strong and weak. To illustrate and also to dip into history we begin with a study of coin-tossing and a discussion of the two most famous prototypes of weak and strong limit theorems.

1. n INDEPENDENT TOSSES OF A FAIR COIN

These words put us immediately into difficulty. What meaning can be assigned to the words, *coin, fair, independent*? Take a pragmatic attitude—all computations involving n tosses of a fair coin are based on two givens:

- a) There are 2^n possible outcomes, namely, all sequences n -long of the two letters H and T (Heads and Tails).
- b) Each sequence has probability 2^{-n} .

Nothing else is given. All computations regarding odds, and so forth, in fair coin-tossing are based on (a) and (b) above. Hence we take (a) and (b) as being the complete definition of n independent tosses of a fair coin.

2. THE “LAW OF AVERAGES”

Vaguely, almost everyone believes that for large n , the number of heads is *about the same* as the number of tails. That is, if you toss a fair coin a large number of times, then about half the tosses result in heads.

How to make this mathematics? All we have at our disposal to mathematize the “law of averages” are (a) and (b) above. So if there is anything at all corresponding to the law of averages, it must come out of (a) and (b) with no extra added ingredients.

Analyze the 2^n sequences of H and T . In how many of these sequences do exactly k heads appear? This is a combinatorial problem which clearly can be rephrased as: Given n squares, in how many different ways can we distribute k crosses on them? (See Fig. 1.1.) For example, if $n = 3$, $k = 2$, then we have the result shown in Fig. 1.2, and the answer is 3.

To get the answer in general, take the k crosses and subscript them so they become different from each other, that is, $+_1, +_2, \dots, +_k$. Now we

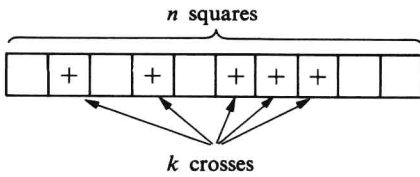


Figure 1.1

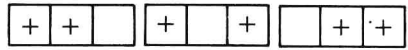


Figure 1.2

may place these latter crosses in n squares in $n(n - 1) \cdots (n - k + 1)$ ways [$+_1$ may be put down in n ways, then $+_2$ in $(n - 1)$ ways, and so forth]. But any permutation of the k subscripted crosses among the boxes they occupy gives rise to exactly the same distribution of unsubscripted crosses. There are $k!$ permutations. Hence

Proposition 1.1. *There are exactly*

$${}_n C_k = \frac{n!}{k!(n - k)!}$$

sequences of H, T, n -long in which k heads appear.

Simple computations show that if n is even, ${}_n C_k$ is a maximum for $k = n/2$ and if n is odd, ${}_n C_k$ has its maximum value at $k = (n - 1)/2$ and $k = (n + 1)/2$.

Stirling's Approximation [59, Vol. I, pp. 50 ff.]

$$(1.2) \quad n! = e^{-n} n^n \sqrt{2\pi n} (1 + \epsilon_n),$$

where $\epsilon_n \rightarrow 0$ as $n \rightarrow \infty$.

We use this to get

$$(1.3) \quad {}_{2n} C_n = \frac{(2n)!}{n! n!} = \frac{e^{-2n} (2n)^{2n} \sqrt{4\pi n}}{e^{-2n} n^{2n} (2\pi n)} (1 + \delta_n) \\ = 2^{2n} \cdot \frac{1}{\sqrt{\pi n}} (1 + \delta_n),$$

where $\delta_n \rightarrow 0$ as $n \rightarrow \infty$.

In $2n$ trials there are 2^{2n} possible sequences of outcomes H, T . Thus (1.3) implies that $k = n$ for only a fraction of about $1/\sqrt{\pi n}$ of the sequences. Equivalently, the probability that the number of heads equals the number of tails is about $1/\sqrt{\pi n}$ for n large (see Fig. 1.3).

Conclusion. As n becomes large, the proportion of sequences such that heads comes up exactly $n/2$ times goes to zero (see Fig. 1.3).

Whatever the "law of averages" may say, it is certainly not reasonable in a thousand tosses of a fair coin to expect exactly 500 heads. It is not

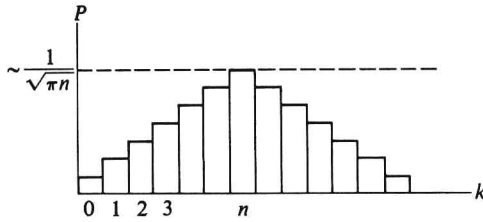


Figure 1.3 Probability of exactly k heads in $2n$ tosses.

possible to fix a number M such that for n large most of the sequences have the property that the number of heads in the sequence is within M of $n/2$. For $2n$ tosses this fraction of the sequences is easily seen to be less than $2M/\sqrt{\pi n}$ (forgetting δ_n) and so becomes smaller and smaller.

To be more reasonable, perhaps the best we can get is that usually the *proportion* of heads in n tosses is close to $\frac{1}{2}$. More precisely—

Question. Given any $\epsilon > 0$, for how many sequences does the proportion of heads differ from $\frac{1}{2}$ by less than ϵ ?

The answer to this question is one of the earliest and most famous of the limit theorems of probability. Let $N(n, \epsilon)$ be the number of sequences n -long satisfying the condition of the above question.

Theorem 1.4. $\lim_n 2^{-n} N(n, \epsilon) = 1$.

In other words, the fraction of sequences such that the proportion of heads differs from $\frac{1}{2}$ by less than ϵ goes to one as n increases for any $\epsilon > 0$.

This theorem is called *the weak law of large numbers for fair coin tossing*. To prove this theorem we need to show that

$$(1.5) \quad \lim_n \left[\frac{1}{2^n} \sum_{k; |k/n - 1/2| < \epsilon} {}_n C_k \right] = 1.$$

Theorem 1.4 states that most of the time, if you toss a coin n times, the proportion of heads will be close to $\frac{1}{2}$. Is this what is intuitively meant by *the law of averages*? Not quite—the abiding faith seems to be that no matter how badly you have done on the first n tosses, eventually things will settle down and smooth out *if you keep tossing the coin*.

Ignore this faith for the moment. Let us go back and establish some notation and machinery so we can give Theorem 1.4 an interesting proof. One proof is simply to establish (1.5) by direct computation. It was done this way originally, but the following proof is simpler.

Definition 1.6

- a) Let Ω_n be the space consisting of all sequences n -long of H, T . Denote these sequences by $\omega_1, \omega_2, \dots, \omega_N, N = 2^n$.
- b) Let A, B, C , and so forth, denote subsets of Ω_n . The probability $P(A)$ of any subset A is defined as the sum of the probabilities of all sequences in A , that is,

$$P(A) = 2^{-n} (\text{number of sequences in } A),$$

equivalently, $P(A)$ is the fraction of the total number of sequences that are in A .

For example, one interesting subset of Ω_n is the set A_1 of all sequences such that the first member is H . This set can be described as "the first toss results in heads." We should certainly have, if (b) above makes sense, $P(A_1) = \frac{1}{2}$. This is so, because there are exactly 2^{n-1} members of Ω_n whose first member is H .

- c) Let $X(\omega)$ be any real-valued function on Ω_n . Define the expected value of X as

$$EX = \sum_{\omega \in \Omega_n} X(\omega) \cdot \frac{1}{2^n}.$$

Note that the expected value of X is just its average weighted by the probability. Suppose $X(\omega)$ takes the value x_1 on the set of sequences A_1 , x_2 on A_2 , and so forth; then, of course,

$$EX = \sum_i x_i P(A_i).$$

And also note that EX is an integral, that is,

$$E(\alpha X + \beta Y) = \alpha EX + \beta EY,$$

where α, β are real numbers, and $EX \geq 0$, for $X \geq 0$. Also, in the future we will denote by $\{\omega; \dots\}$ the subset of Ω_n satisfying the conditions following the semicolon.

The proof of 1.4 will be based on the important Chebyshev inequality.

Proposition 1.7. For $X(\omega)$ any function on Ω_n and any $\epsilon > 0$,

$$P(\omega; |X(\omega)| \geq \epsilon) \leq \frac{1}{\epsilon^2} EX^2.$$

Proof

$$\begin{aligned} P(\omega; |X| \geq \epsilon) &= \frac{1}{2^n} (\text{number of } \omega; |X(\omega)| \geq \epsilon) = \sum_{\{\omega; |X| \geq \epsilon\}} \frac{1}{2^n} \\ &\leq \sum_{\{\omega; |X| \geq \epsilon\}} \frac{X^2(\omega)}{\epsilon^2} \frac{1}{2^n} \leq \frac{1}{\epsilon^2} \sum_{\omega \in \Omega_n} X^2(\omega) \cdot \frac{1}{2^n} = \frac{1}{\epsilon^2} EX^2. \end{aligned}$$

Define functions $X_1(\omega), \dots, X_n(\omega), S_n(\omega)$ on Ω_n by

$$(1.8) \quad X_j(\omega) = \begin{cases} 1 & \text{if } j\text{th member of } \omega \text{ is } H, \\ 0 & \text{if } j\text{th member of } \omega \text{ is } T, \end{cases}$$

$$S_n(\omega) = X_1(\omega) + \dots + X_n(\omega),$$

so that $S_n(\omega)$ is exactly the number of heads in the sequence ω . For practice, note that

$$EX_1 = 0 \cdot P(\omega; \text{first toss} = T) + 1 \cdot P(\omega; \text{first toss} = H) = \frac{1}{2},$$

$$EX_1X_2 = 0 \cdot P(\omega; \text{either first toss or second toss} = T) \\ + 1 \cdot P(\omega; \text{both first toss and second toss} = H) = \frac{1}{4}$$

(since there are 2^{n-2} sequences beginning with HH). Similarly, check that if $i \neq j$, then

$$(X_i - \frac{1}{2})(X_j - \frac{1}{2}) = \begin{cases} \frac{1}{4} & \text{on } 2^{n-1} \text{ sequences,} \\ -\frac{1}{4} & \text{on } 2^{n-1} \text{ sequences,} \end{cases}$$

so that

$$E(X_i - \frac{1}{2})(X_j - \frac{1}{2}) = 0, \quad i \neq j.$$

Also,

$$(X_i - \frac{1}{2})^2 \equiv \frac{1}{4}, \quad \Rightarrow E(X_i - \frac{1}{2})^2 = \frac{1}{4}.$$

Finally, write

$$S_n - \frac{n}{2} = \sum_{j=1}^n (X_j - \frac{1}{2})$$

so that

$$(1.9) \quad E\left(\frac{S_n}{n} - \frac{1}{2}\right)^2 = \frac{1}{n^2} E\left(\sum_{i,j} (X_i - \frac{1}{2})(X_j - \frac{1}{2})\right) \\ = \frac{1}{n} E(X_1 - \frac{1}{2})^2 = \frac{1}{4n}.$$

Proof of Theorem 1.4. By Chebyshev's inequality,

$$P\left(\omega; \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{E(S_n/n - \frac{1}{2})^2}{\epsilon^2}.$$

Use (1.9) now to get

$$(1.10) \quad P\left(\omega; \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \geq \epsilon\right) \leq \frac{1}{4n\epsilon^2},$$

implying

$$\lim_n P\left(\omega; \left|\frac{S_n(\omega)}{n} - \frac{1}{2}\right| \geq \epsilon\right) = 0.$$

Since $P(\Omega_n) = 1$, this completes the proof.